MALLEABILITY OF ATTITUDES OR MALLEABILITY OF THE IMPLICIT ASSOCIATION TEST?

DISSERTATION
Presented in Partial Fulfillment of the Requirements for
The Degree Doctor of Philosophy in
the Graduate School of The Ohio State University

By

HYO-JUNG ANNA HAN, B.A., B.S., M.A.
Graduate Program in Psychology

The Ohio State University
2009

Doctoral Examination Committee:

Dr. Russell H. Fazio, Advisor
Dr. Richard E. Petty
Dr. Kentaro Fujita
ABSTRACT

The present research sought to explore the inherent ambiguity characteristic of the Implicit Association Test (IAT) as a function of the vague response labels that are typically employed. The goals were three-fold: 1) demonstrate the malleability of the traditionally employed IAT as a function of ambiguous valence labels; 2) investigate the consequences of the malleability by reexamining prior research interpreting the malleability of the IAT as an evidence of attitude change when no such change seemed likely; and 3) explore how the ambiguity present in the evaluative labels may also impact the construal of the category labels (ambiguity in the category labels).

Experiments 1 and 2 demonstrate the malleability of the IAT, as traditionally implemented. IAT scores are shown to be influenced by conversational norms and by perspective mindsets induced by an unrelated preceding task. Then, Experiments 3 and 4 illustrate how the malleability of the IAT can lead to the mistaken inference that attitude change has occurred even when there is very good reason to believe it has not. Experiments 5 and 6 demonstrate the ambiguity of the category labels (vis-à-vis) the ambiguous valence labels and how atypical stimulus exemplars can lead to redefinition of the category labels. Lastly, Experiment 7 illustrates the value of focused and unambiguous evaluative labels for deterring category redefinition. Together, the experiments suggest methodological improvements to enhance the validity of IAT measures.
Dedicated to my family, friends, and colleagues who have supported me in all my endeavors.
ACKNOWLEDGMENTS

I wish to thank my advisor, Russ Fazio, for his guidance, support, and endless patience on this, and many other projects. Russ has been instrumental in the development of both this project and my graduate career more generally. The education and knowledge he has provided has changed me forever and will leave an indelible mark on my professional life.

I would also like to thank my secondary advisors Ken Fujita and Gifford Weary. I have enjoyed working with them, and the exchange of ideas and active debate with both of them has shaped the ideas presented here, as well as my thinking more generally.

I would also like to thank the members of my dissertation committee, Richard E. Petty and Kentaro Fujita, for all the time and feedback they have devoted to both this project and my prior research endeavors.

Finally, I would like to thank the many research assistants who helped collect this data and the members of the Ohio State University’s Group for Attitudes and Persuasion and Social Cognition Research Group for their fruitful comments on this work as it has developed.
VITA

February 10, 1979 ........................................... Born – Seoul, South Korea

2002 ........................................................................ B.A., B.S. Psychology, Ohio State University

2004 ........................................................................ M.A. Psychology, Ohio State University

2002 – present .................................................. Graduate Associate, Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Psychology

Specialization: Social Psychology
Minor Field: Quantitative Psychology

Secondary Minor: Political Psychology
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>v</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Goals and Overview of the Studies</td>
<td>8</td>
</tr>
<tr>
<td>2. Demonstration of the Malleability of the IAT</td>
<td>10</td>
</tr>
<tr>
<td>Experiment 1</td>
<td>10</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>17</td>
</tr>
<tr>
<td>General Discussion</td>
<td>23</td>
</tr>
<tr>
<td>3. Consequences of Malleability of the IAT</td>
<td>25</td>
</tr>
<tr>
<td>Experiment 3</td>
<td>29</td>
</tr>
<tr>
<td>Experiment 4</td>
<td>34</td>
</tr>
<tr>
<td>General Discussion</td>
<td>39</td>
</tr>
<tr>
<td>4. Stimulus Exemplar Effects</td>
<td>42</td>
</tr>
<tr>
<td>Experiment 5</td>
<td>45</td>
</tr>
<tr>
<td>Experiment 6</td>
<td>48</td>
</tr>
<tr>
<td>Discussion</td>
<td>50</td>
</tr>
<tr>
<td>5. On the Value of Focusing the IAT</td>
<td>52</td>
</tr>
<tr>
<td>Experiment 7</td>
<td>53</td>
</tr>
<tr>
<td>6. General Discussion</td>
<td>58</td>
</tr>
<tr>
<td>Conclusion</td>
<td>66</td>
</tr>
</tbody>
</table>
List of References ................................................................................................................................. 68
Appendix A - Figures ............................................................................................................................. 84
Appendix B – Experiment 1 instructions .............................................................................................. 84
Appendix C – Experiment 2 Screening Measure ..................................................................................... 86
Appendix D – Experiment 4 Knowledges Task Exemplar ...................................................................... 88
Appendix E – Experiment 5 Stimuli ....................................................................................................... 103
Appendix F – Experiments 6 & 7 Elderly Stimuli .................................................................................. 109
Appendix G – Experiment 7 Asian Stimuli ............................................................................................ 113
Appendix H – Experiment 7 Screening Measure ................................................................................... 115
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Experiment 3 Results (original Scoring)</td>
<td>74</td>
</tr>
<tr>
<td>2.</td>
<td>Experiment 3 Results (D-Scores)</td>
<td>75</td>
</tr>
<tr>
<td>3.</td>
<td>Experiment 4 Results (original Scoring)</td>
<td>76</td>
</tr>
<tr>
<td>4.</td>
<td>Experiment 4 Results (D-Scores)</td>
<td>77</td>
</tr>
<tr>
<td>5.</td>
<td>Experiment 5 Results (original Scoring)</td>
<td>78</td>
</tr>
<tr>
<td>6.</td>
<td>Experiment 5 Results (D-Scores)</td>
<td>79</td>
</tr>
<tr>
<td>7.</td>
<td>Experiment 6 Results (original Scoring)</td>
<td>80</td>
</tr>
<tr>
<td>8.</td>
<td>Experiment 6 Results (D-Scores)</td>
<td>81</td>
</tr>
<tr>
<td>9.</td>
<td>Experiment 7 Results (original Scoring)</td>
<td>82</td>
</tr>
<tr>
<td>10.</td>
<td>Experiment 7 Results (D-Scores)</td>
<td>83</td>
</tr>
</tbody>
</table>
Most behavioral scientists who employ questionnaire measures have experienced the occasional, troubling realization that participants interpreted a specific question differently than they had intended. Whether the cause was poor wording, the implications of a preceding set of questions, or even some unexpected natural event that cast a different light on the matter at hand, it became apparent that the participants’ responses were less than correspondent to the query that had been posed. Indeed, years of research have been devoted to the study of survey responding as an exercise in communication between the questioner and the respondent; “questions shape answers” (Schwarz, 1999, p.93). It has been shown repeatedly that when survey questions are ambiguous, participants guess or use contextual information to disambiguate and respond as best they can (Bickart, 1992; Billiet, Waterplas, & Loosveldt, 1992; Krosnick & Alwin, 1987; Krosnick, 1992). Although they may stem from a desire to be cooperative, such efforts to disambiguate have the potential to produce less than meaningful data and can lead the researcher to draw inappropriate inferences. Hence, knowledgeable and experienced survey researchers consistently strive to construct questions that are free of ambiguities.

Perhaps because of the basic assumptions underlying implicit measures (e.g., Greenwald & Banaji, 1995), the possible impact of ambiguity on these attitude measurement tools has not yet been as carefully delineated. Researchers have devoted considerable effort to determining
exactly what measures like the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) are assessing, as well as the mechanisms that underlie the measurement procedures (e.g., De Houwer, 2009). Although it has become apparent that the IAT is characterized by ambiguity of various forms (e.g., Fazio & Olson, 2003; Govan & Williams, 2004; Karpinski & Hilton, 2001), the issue needs to be addressed more directly so as to inform researchers about the appropriateness of inferences that might be drawn from IAT findings. Most importantly, the implications of any such ambiguity for inferences regarding the measurement procedures versus underlying representation require careful analysis. When change is observed on the IAT, does that change reflect the malleability of attitudes or the malleability of the measure? I argue that the ambiguity present in the IAT labels can lead to biased attitude estimates and erroneous conclusions regarding the malleability of attitudes.

The Implicit Association Test

The IAT (Greenwald, McGhee, & Schwartz, 1998) measures the degree to which two target concepts (e.g. Black/White) are associated in memory with positive and negative evaluations (e.g. Pleasant/Unpleasant). The task is to quickly categorize the stimuli by using one of two key assignments (left/right). The assumption is that when two highly associated concepts are assigned to the same response key (e.g. compatible block; White-pleasant vs. Black-unpleasant), the participant will be faster to categorize stimulus examples than when two non-associated concepts (e.g. incompatible block, White- unpleasant vs. Black- pleasant) are represented by the same key (Greenwald & Nosek, 2001). The differences in latencies between the incompatible and compatible blocks reveal the evaluation of the two target concepts, and hence, provide an attitude estimate (e.g. racial prejudice).
Since its inception, the IAT has become the preferred implicit measure for many psychological variables. Because implicit measures are presumed to be relatively immune from many of the concerns that plague self-report measures, the IAT has been especially useful in domains in which social desirability is a concern. This has led to its extensive use in areas such as stereotyping and prejudice (e.g., Greenwald et al., 1998; Nosek, Greenwald, & Banaji, 2005; Blair, 2002), health behaviors (e.g., Sherman et al., 2002), and self-esteem (e.g., Greenwald & Farnham, 2000; Jordan et al., 2003). Despite the wide popularity and usage, there remains a debate about what the IAT is “tapping into.” Emerging evidence seems to indicates that the IAT as traditionally implemented may not be as robust to momentary, irrelevant contextual considerations as once believed (e.g., Karpinski & Hilton, 2001; Olson & Fazio, 2004).

The IAT, Its Ambiguity, and the Role of Extrapersonal Associations

Olson and Fazio (2004) argued that the features of the traditionally implemented IAT, which utilizes the category labels “pleasant/unpleasant” or “good/bad,” are ambiguous enough that any information useful for coping with the demands of the categorization task can be activated – sometimes in the form of normative, culturally-shared associations and/or other specific individual’s associations. Most importantly, the perspective intended by the most frequently used category labels is inherently ambiguous: “pleasant”/“unpleasant” or “good”/“bad” to whom? Should the stimuli be categorized from the perspective of the self, the culture, or the researcher, to consider just a few possibilities? An important consequence of this category label ambiguity is that the attitude estimates provided by the traditional IAT versions can be influenced by extrapersonal associations (e.g., Han, Olson, & Fazio, 2006; Karpinski & Hilton, 2001; Olson & Fazio, 2004) – attitude-irrelevant knowledge that does not form the basis of the individual’s attitude toward to the object. Such extrapersonal knowledge sometimes
arises from the recognition that others have attitudes that differ from one’s own, whether those others be a specific individual, a class of people, or even more general cultural norms. Despite the prevalence of such knowledge, extrapersonal associations do not consistently impact our own judgments or non-public behaviors (e.g., the knowledge that my neighbor loves cats do not impact my preference for cats). Nevertheless, such knowledge can facilitate a respondent’s efforts to cope with the dual meaning of the response keys during the IAT (Olson & Fazio, 2004).

Thinking about Uncle Joe or my neighbor’s preferences, the researcher’s presumed intent, or the cultural perspective may promote faster responding for a given response mapping on the IAT.

So exactly what is an extrapersonal association? Simply put, extrapersonal associations are items of information that, although associated with the attitude object, have not contributed to one’s summary evaluation of that object. This may be because the information was deemed inconsequential or irrelevant to one’s personal tastes, or because it was rejected as untrue (Petty, Briñol, & DeMarree, 2007). Or, it may simply be because the information played no role as the attitude developed. An individual with an allergy to cats in all likelihood developed a negative attitude as a consequence of the first causally identifiable allergic reaction, and this aversion will have been reinforced by the individual’s regular surveillance of the environment for signs of cats. A sibling’s or a friend’s love for cats played no role in the development of this negative attitude, nor did other knowledge related to cats, such as the cuteness of Hello Kitty or the funniness of LOL cats (http://icanhascheezburger.com/).

Information of this sort is certainly available in memory, but it did not contribute to the person’s negative attitude toward cats. In contrast, personal associations are one’s attitudes – summary evaluations that are automatically activated upon encountering an attitude object and that then
guide construal of the object in the immediate situation and ultimately approach/avoidance
decisions (“Ekkk! A cat! I might want to die!”) (See Fazio, 2007).

Some researchers appear to view extrapersonal associations as directly corresponding
to cultural associations (Nosek & Hansen, 2008a). However, there is no necessary equivalence
between the two. Indeed, cultural knowledge, i.e., how the attitude object is generally
portrayed within the culture, may serve as the very basis for a given individual’s attitude. It is
when personal attitudes deviate from the more cultural, normative view that such cultural
knowledge can be considered extrapersonal for a given person. In such cases, the individual’s
knowledge of the more culturally predominant negativity assumes the status of an
extrapersonal association.

Indeed, extrapersonal associations can influence performance on the IAT. Han et al.
(2006) demonstrated such effects by experimentally creating both attitudes and extrapersonal
associations towards novel stimuli. In their experiments, participants first formed an attitude
towards two novel Pokémon characters, one of which was objectively superior to the other, and
hence, clearly preferred. A questionnaire administered at this point in the procedure induced
participants to rehearse and express their attitudes multiple times. Extrapersonal associations
then were introduced via a video of two 10-years-old boys discussing the same characters.
Depending on the condition, the boys either agreed with participants as to which character was
superior (consistent condition) or disagreed, expressing an unjustified preference for the
objectively inferior character (inconsistent condition). Across two experiments, results revealed
that participants’ IAT scores differed as a function of the extrapersonal associations to which
they had been exposed. When exposed to inconsistent extrapersonal associations (i.e., when
the boys’ opinions contrasted with participants’ attitudes), participants’ IAT scores reflected a
significantly reduced preference for the objectively superior Pokémon character compared to those who had been given consistent extrapersonal associations (when the boys agreed). It appeared that participants in the extrapersonal inconsistent condition were able to recruit the boys’ opinion to solve the mapping problem posed in the incompatible block, in which the inferior Pokémon character was assigned to the same key as “pleasant” and the superior Pokémon with “unpleasant.” This change in IAT scores occurred despite the fact that participants rated the boys as irrational and foolish and chose the superior Pokémon card over the other as a parting bonus at the end of the study. Moreover, in contrast to the traditional IAT, a subliminal priming measure of attitudes (in Experiment 1; Han et al., 2006), and a personalized version of the IAT (in Experiment 2), where the “pleasant/unpleasant” labels were changed to “I like/I don’t like,” remained unaffected by the introduction of the inconsistent extrapersonal video. Both measures revealed only a preference for the objectively superior Pokémon.

Personalized IAT

The personalized IAT was developed by Olson and Fazio (2004) specifically for the purpose of focusing the IAT on personal attitudes and limiting its susceptibility to extrapersonal information. The two most important changes with respect to the traditional IAT involve suppressing error feedback and changing the category labels from “pleasant” or “good” to “I like” and from “unpleasant” or “bad” to “I don’t like.” These changes are intended to focus thoughts on personal likes and dislikes. Error feedback encourages a normative focus by indicating that a correct answer, independently of the person’s own liking or disliking of the target objects, exists. The perspective intended by the labels “pleasant/unpleasant” or “good/bad” is ambiguous, enabling extrapersonal associations to intervene in solving the
mapping problem of the IAT. In contrast, the use of the labels “I like/I don’t like” precludes such construals because it focuses attention on one’s personal feelings regarding the object.

Across several attitude domains, evidence supports the idea that these modifications to the traditional IAT procedure indeed reduce the influence of extrapersonal associations on the IAT effect. Thus, Olson and Fazio’s (2004) Experiments 1 and 2 found evidence of less racial bias on the personalized IAT than on the traditional IAT, suggesting that negative extrapersonal associations about Black people reflecting negative culturally-shared stereotypes may enhance the bias observed on the traditional IAT. Relatedly, Goff, Williams, Eberhardt, & Jackson (2008) showed that cultural stereotypes associating Blacks with apes were not related to racial attitudes as measured with the personalized IAT. In their Experiment 3, Olson and Fazio (2004) found the traditional IAT to reflect greater preference for apples over candy bars than the personalized IAT, suggesting that positive extrapersonal associations about apples may inflate the traditional IAT effect. Finally, as noted earlier, Han et al. (2006) experimentally created both attitudes and extrapersonal associations about game characters and showed that both attitudes and extrapersonal associations affected the traditional IAT, whereas only attitudes affected the personalized IAT.

Additionally, compared to the traditional IAT, personalized IATs have yielded stronger relations with reports of past behavior and preferences regarding both foods and political candidates (Olson & Fazio, 2004). The measure has also served more successfully as a tool for discriminating cigarette smokers from nonsmokers (De Houwer, Custers, & De Clercq, 2006) and for predicting alcohol consumption among heavy drinking students than traditional versions of the IAT (Houben & Wiers, 2007).
Why does a personalized IAT improve correspondence with explicit measures and behaviors? The improvement may be due to the inherent ambiguity of the labels in the traditional version of the IAT. The more focused labels (“I like/I don’t like”) carry a clear meaning that disambiguates the valenced category labels “pleasant/unpleasant.” The focus becomes one’s own attitudes, diminishing tendencies for the activation of information that does not form a basis for those attitudes, such as someone else’s perspective or irrelevant cultural norms.

The role of error feedback

One might ask whether the usual absence of error feedback in the personalized IAT also plays a role. Past research on personalized IAT has not observed the absence of feedback to have a more direct effect on performance by lessening participants’ concerns about making errors. Indeed, none of the five experiments reported in Olson & Fazio (2004) or Han et al. (2006) revealed the personalized and traditional IATs to yield differential error rates (all t’s < 1). This null finding stands in contrast to the conclusion reached by Nosek and Hansen (2008b), who reported greater error rates for the personalized IAT than the traditional across a large number of web-based studies. One can only conjecture that the discrepancy stems from lesser attentiveness and concentration when individuals participate over the web as opposed to an experimenter-supervised laboratory context. Nosek and Hansen (2008b) report average error rates of 11.9% and 9.2% for the personalized and traditional IATs, respectively. Across the five experiments in Olson & Fazio (2004) and Han et al. (2006), the mean error rates for the personalized and traditional IATs were 5.6% and 5.3%, respectively, roughly half the rate observed in the web-based research. Thus, the differences that we have discussed with respect to the validity of the two IAT versions cannot be dismissed as due to a lesser understanding of
the task instructions or a lesser emphasis on accuracy when the IAT is personalized.

Furthermore, it should be noted that Experiment 2 of Han et al. (2006) included error feedback in both the personalized and the traditional versions of the IAT. That is, the two IATs differed only with respect to the category labels. Despite the equivalence with respect to error feedback, the personalized IAT showed less susceptibility to the influence of extrapersonal associations than did the traditional IAT.

Goals and Overview of Studies

Since the valence labels of the traditional IAT are potentially ambiguous (much like ambiguous survey questions), the current research explored whether the IAT is also affected by contextual factors known to bias responses on explicit survey measures. That is, we investigated the ambiguous nature of the IAT and how participants’ scores on this measure may reflect the influence of factors other than or in addition to personal evaluations of the target categories (Experiments 1 and 2). Specifically, Experiment 1 explores how an unrelated preceding task can change participants’ interpretation of the IAT category labels, and Experiment 2 explores the impact of conversational norms on the IAT. Experiments 3 and 4 explore the consequences of the IAT’s susceptibility to these effects by replicating and shedding light on two well-known studies concerning the malleability of automatically-activated attitudes. Specifically, these experiments examine whether the results outlined in these past experiments might simply reflect the malleability of the IAT itself rather than genuine changes in participants’ attitudes. Finally, Experiment 5, 6, and 7 investigated the potential ambiguity of the category labels (vs. the valence labels) and sought out ways to improve the IAT.
CHAPTER 2

DEMONSTRATION OF MALLEABILITY OF THE IAT

Since the valence labels of the traditional IAT are potentially ambiguous (much like ambiguous survey questions), it is important to investigate whether the IAT is also affected by contextual factors known to bias responses on explicit survey measures. Hence, Experiments 1 and 2 explore the ambiguous nature of the IAT and how participants’ scores on this measure may reflect the influence of factors other than personal evaluations of the target categories. Specifically, Experiment 1 investigates how an unrelated preceding task can change participants’ interpretation of the IAT category labels, and Experiment 2 explores the impact of conversational norms on the IAT.

EXPERIMENT 1

It is well known that previous tasks can influence how people interpret and answer subsequent questions. For example, after being asked to select two high quality TV shows from a list of 10, participants rated TV programming as significantly more satisfying than after selecting two low quality TV shows (Bless & Wänke, 2000). Similarly, Salancik and Conway (1975) demonstrated that having participants complete a behavioral inventory that pairs pro-religious behaviors with the adverb “occasionally” and anti-religious behaviors with the adverb “frequently” makes pro-religious behaviors relatively salient and enhances self-reports of religiosity on immediately subsequent survey questions. Shavitt and Fazio (1991) found participants to rate the upscale brand Perrier more favorably if the measure had been preceded
by a questionnaire in which the participants indicated the extent to which various actions would make a good impression on others as compared to a questionnaire involving how good various foods taste (even though Perrier was not included in either of the initial questionnaires).

Borrowing from these paradigms, we investigated whether the labels on the IAT as traditionally implemented are ambiguous enough to allow participants to shift their meaning as a function of the information made accessible by a previous, unrelated task. Participants in Experiment 1 first completed a 40-item questionnaire in which they rated either how much “people like/don’t like” or how much “I like/don’t like” various non-race related attitude objects. All participants then performed a traditional IAT assessing racial attitudes. It was expected that the ambiguous IAT labels would be interpreted differently depending on which version of the questionnaire had been completed. In particular, we expected that the rating task would prime participants to view the IAT labels in a manner consistent with the questionnaire labels, leading to different IAT scores. Consistent with arguments regarding the prevalence of negative cultural portrayals of Blacks (Goff et al., 2008; Olson & Fazio, 2004), we expected to observe greater racial bias in the normative focus prime condition (“People like/don’t like”) than the personal focus prime condition (“I like/don’t like”).

**Participants**

Participants were fifty-three introductory psychology students (11 males, 42 females) who participated in this experiment in partial fulfillment of their course requirements. Up to four individuals participated in any given session, with each occupying a private cubicle and each randomly assigned to the “People like” or “I like” condition. The data from one participant was excluded because she correctly guessed the purpose of the experiment. The final sample consisted of 52 participants (11 males, 41 females).
Materials and Procedure

Procedure. The current experiment was conducted in groups of 2 to 4 participants, and they were run simultaneously, albeit in individual cubicles. After answering some demographic questions (e.g., age and gender), participants learned that they would be completing two short separate studies, each of which was computer-administered. They were told the first study was a “norming” study on likes and dislikes and the second was about categorization abilities. In the first, participants were asked to make 40 evaluative judgments concerning a variety of non-race related attitude objects such as foods, sports, occupations, and social issues (e.g., apples, hockey, teachers, and global warming) on a seven-point scale. In the ostensibly unrelated second study, participants completed a traditional IAT assessing racial attitudes, with the usual category labels of pleasant/unpleasant.

The first part of the experiment was a manipulation specifically designed to elicit a normatively focused (“People like/don’t like”) or personally focused (“I like/I don’t like”) mindset (see Shavitt & Fazio, 1991). It was presented as “Study 1: Norming study,” and in both conditions, participants evaluated the same 40 non-race related items. However, in the normatively focused condition, participants were asked, “Please rate the extent to which you think PEOPLE IN GENERAL like or dislike the following item: _________” on a 7-point scale anchored by the labels People do not like very much (1) and People like very much (7). Participants were further instructed to ignore their own likes and dislikes and answer with what they thought people in general like/dislike. Conversely, in the personal focus condition, they were asked, “Please rate the extent to which YOU like or dislike the following item: _________” anchored by the labels I do not like very much (1) and I like very much (7), and were asked to
ignore what people in general like/dislike and focus on their own personal preferences (see Appendix B for the full instructions).

Upon completing the rating task, participants were informed that Study 1 was over and were thanked.

Participants were then asked to complete a second, ostensibly unrelated study for another researcher. To boost the credibility that this was an unrelated study, the computer started a new program once the participants clicked on a button to participate in the second study.

The IAT used in this experiment was designed following a procedure established by Greenwald et al. (1998). The IAT consisted of seven blocks with 20 trials in noncritical blocks and 20 trials in critical blocks. The first two blocks were practice blocks; Block 1 required categorizing black/white names (e.g. “Tyrone”, “Josh”) as Black or White, and Block 2 required the categorization of valenced words (e.g. “love”, “murder”) as pleasant or unpleasant. Then two critical combined blocks (blocks 3-4) were presented, with Black names being paired with the pleasant category while White names were paired with the unpleasant category (or visa versa, depending on counterbalancing). The next block (5) involved categorization of the Black and White names with the keys reversed relative to block 1. Two more critical combined blocks (6-7) were presented but involved the reverse categorization from blocks 3-4. Instructions on the meaning of the keys and type of items to categorize were presented at the beginning of each block. The order in which the participants performed the critical combined blocks was counterbalanced.

The experimenter described the IAT as a categorization task that required both speed and accuracy. Participants were told that they would be asked to categorize some of
stereotypically black and white names and words that generally have positive and negative connotations. Participants were then instructed to categorize the items on the screen by using the keyboard.

Following completion of all materials, participants were probed for suspicion and debriefed.

Results and Discussion

The IAT analyses in this and the other experiments to follow were conducted in two ways: 1) following the procedure originally established by Greenwald et al. (1998) and 2) via the D-score algorithm involving a 600ms penalty for each error response (Greenwald, Nosek, & Banaji, 2003). For all experiments, the result sections will report both analyses. Both sets of analyses are reported due to the controversies surrounding the computation of IAT scores. The D-score method, which has been derived from analyzing web-based data, imposes a 600ms penalty on the latency for each trial on which the participant has made an error. Some, including the current author, view this as an arbitrary means of weighting and combining error and latency data into a single composite and, hence, would prefer to follow the more standard practice in latency research whereby error trials are simply excluded from the analysis. However, D600 scores were included to assure skeptical readers that the two approaches yielded similar outcomes.

Original Scoring Method

The original IAT score was computed by dropping the response latencies for the first two trials of each block and recoding latencies under 300 and over 3,000 ms to 300 and 3,000 ms, respectively. The latencies were then log-transformed. Means of each critical block type were then computed (Blocks 3-4 and Blocks 6-7). For all participants, compatible blocks were
identified as White/+ vs. Black/-- and incompatible blocks as White/-- vs. Black/+.

The IAT score was computed by subtracting the latencies of compatible blocks from those for incompatible blocks, whereby higher IAT score indicated greater racial bias against Blacks. All analyses were done using the log transformation, but raw latencies will be presented for ease of interpretation.

As expected, scores on the IAT varied as a function of the scale labels presented on the previous task, \( t(50) = 2.20, p = .035 \). Compared to the “people like/people don’t like” condition, participant who completed “I like/don’t like” questionnaire showed a significantly reduced racial bias (\( M = 272, SD = 143 \) vs. \( M = 182, SD = 127 \), respectively).

The D-Score Algorithm

In accord with the D-score algorithm, the data were first searched for trials and participants meeting the specified criteria for exclusion. There were no trials on which latencies exceeded 10,000 ms. Nor did any participants display latencies less than 300ms on more than 10\% of the trials. Then mean correct latencies for each of the 4 critical blocks were computed, as well as pooled standard deviations for blocks 3 and 6 and for blocks 4 and 7. All error latencies were replaced by the block mean plus 600 ms penalty, and average resulting value of each block was then calculated. Mean latency differences were computed (i.e., first incompatible block – compatible block and second incompatible block – compatible block) and then divided by their corresponding pooled standard deviations. Finally, they were averaged to form a D-score.

---

\(^1\) The current definition of compatible and incompatible blocks applies only to the Race IATs in Experiments 1, 2, 4, 5. For other types of IATs, the result section will specify how the compatible/incompatible blocks were identified.
The D-score method mirrored the findings on the original scoring method. The D-scores varied as a function of the scale labels presented on the previous task, $t(50) = 2.08, p = .043$. Compared to the “people like/people don’t like” questionnaire condition, participant who completed the “I like/don’t like” questionnaire showed a significantly reduced racial bias ($M = .55, SD = .32$ vs. $M = .73, SD = .28$, respectively).

These findings suggest that the IAT, as traditionally implemented, can easily be affected by the context created by preceding tasks. It appears that the traditional IAT labels are indeed ambiguous and that ambiguity allows for multiple interpretations of the task. That is, an earlier experience in an unrelated task can affect how people view the attribute labels and whether they adopt a normative (i.e., “people like/don’t like”) or personal (i.e., “I like/don’t like”) focus while completing the IAT. The accessibility of these differential perspectives can influence performance, and hence, the attitude estimates that are obtained.

Interestingly, further evidence regarding the importance of such perspective effects is suggested by recent research in which the administration of a traditional IAT was preceded by a task that primed either self-related (i.e., by circling words “I” “me” “my” or “myself”) or neutral concepts (i.e., by circling words “the” and “a”) (Connor, Perugini, O’Gorman, & Prestwich, 2007). The predictive validity of the IAT was enhanced in the condition involving activation of the self. Across studies, this enhancement was observed in domains as diverse as ethnic attitudes, preference for science/arts, as well as alcohol and junk food consumption.

These findings point to the importance of perspective one adopts while completing the IAT, and the need to disambiguate the labels to avoid biased attitudes estimates. Experiment 2 sought to further establish the ambiguity of the IAT labels by considering how their interpretation may be affected by conversational norms.
EXPERIMENT 2

According to Grice (1975), conversational norms facilitate interpersonal communication because people try to take into account their conversation partners’ perspectives and knowledge, and, accordingly, make their responses clear, appropriately informative, and nonredundant. In more recent years, it has been shown that conversational norms apply not only to everyday conversation partners but also to questionnaire designers and respondents, such that people answer survey questions assuming that the norms are relevant to the “conversation” they are having with the survey researcher (see Holbrook et al., 2000; Schwarz, 1996). The purpose of Experiment 2 is to explore how the IAT might also be influenced by such conversational norms.

Specifically, the current experiment focused on the norm of nonredundancy, which dictates that the communicator must be succinct without being repetitive. Past research on this norm has shown that when participants’ marital and life satisfaction were gauged in different conversational contexts (nonredundant condition), the correlation between the two variables was high compared to a condition in which the questions were asked in the same conversational context (redundant condition). This occurred presumably because participants in the latter condition interpreted the second question as a request for new information (Schwarz, Strack, & Mai, 1991).

Similarly, Haddock and Carrick (1999) demonstrated that when participants indicated their liking for the Queen Mother (“Britain’s best loved royal,” p. 125) before reporting their liking for British Royal Family in the same survey (redundant condition), the rating for the British Royal Family was less favorable, because participants had excluded the Queen Mother from their rating of the British Royal Family. Presumably, this occurred because participants assumed that
the second question was referring to different aspects of the British Royal Family than had been posed earlier. However, when the questions were asked in two separate surveys, hence indicating nonredundancy, the rating of the Queen Mother and British Royal Family were highly correlated, resulting in more favorable ratings of the Royal Family. In this case, participants appeared to include the Queen Mother in their rating of the Royal Family because doing so no longer violated the nonredundancy norm.

Experiment 2 directly investigated how the IAT, as traditionally implemented, may be influenced by the nonredundancy norm. It was expected that asking participants to perform two race-related IATs involving different labels within the same experiment would prompt them to reconstrue the labels to avoid redundancy. That is, when participants first complete an IAT with clear unambiguous labels and then are asked to complete a second IAT with ambiguous labels, their interpretation of the ambiguous labels may shift to conform to the rules of conversational norms.

Two IATs with varying labels were implemented in order to demonstrate that differentially focused labels can yield very different IAT scores. To do so, participants first completed either a personalized IAT (labels “I like/I don’t like”) or a culturally-oriented IAT (labels “people like/people don’t like”). The use of the personalized IAT is potentially valuable in this context because its labels carry a meaning that focuses on personal likes and dislikes. The culturally-oriented IAT should make the conceptually reverse change, by focusing the IAT labels on people’s normative representations of the target objects. That is, by modifying the labels to “people like/people don’t like”, participants should focus not on their own personal evaluations, but on more general, culturally-shared information. After completing one of these modified IATs, participants completed a traditional IAT in which the category labels were not
disambiguated (labels “pleasant/unpleasant”) so that we can examine whether the initial IAT experience and participants’ desire to provide nonredundant information affect subsequent IAT performance.

We expected to observe greater racial bias on the culturally-oriented IAT than the personalized IAT. Society’s negative portrayals should facilitate the process of assigning “Black” and disliked items to the same response key when the labels focus on “People” as opposed to “I”. However, and more critical to the hypothesis regarding the impact of conversational norms, we also predicted that the scores on the subsequent traditional IAT would fluctuate as a function of the first type of IAT completed. That is, when asked to perform two IATs, participants will presume that the researchers are following the rules of conversational norms and will therefore interpret the “pleasant/unpleasant” category labels of the traditional IAT in such a way as to avoid redundancy. Presumably, the experimenter would not be asking one to perform the task a second time with different labels unless those labels were to convey a different meaning than the previously encountered labels. Thus, in the condition in which participants first completed the personalized IAT, they should be more likely to view the second category label “pleasant/unpleasant” in terms other than personal likes and dislikes. Similarly, in the condition involving the cultural IAT first, they should construe “pleasant/unpleasant” to refer to something other than general normative evaluations. Such a change in interpretation of category labels is implied by the conversational norm and leads, in this case, to the prediction that participants will display more racial bias on the traditional IAT after having performed the personalized version than after having completed the cultural version.
Method

Participants

Sixty-nine introductory psychology students (30 males, 39 females) participated in this experiment in partial fulfillment of their course requirements. Participants were randomly assigned to one of two experimental conditions. Data from two participants were excluded because they indicated they had lived in the U.S. less than three years and had difficulty categorizing black/white names, and data from another participant was deleted due to a high error rate (> 42%). The final sample consisted of 66 participants (28 males, 38 females).

Materials and Procedures

All participants completed two different black/white race IATs (see Experiment 1). Depending on the condition, participants first completed either the personalized version of the IAT, with the labels, “I like/ I don’t like” (Olson & Fazio, 2004), or a culturally-oriented IAT with the labels “people like/ people don’t like.” Participants then completed a short computerized filler survey whose major purpose was to demarcate the conclusion of the first IAT from the beginning of the second. Finally, they then completed the traditional IAT with “pleasant/unpleasant” labels.

In all conditions, the stimulus materials were identical, with the following exceptions: (1) the category labels differed on each IAT, and (2) participants did not receive error feedback on the personalized IAT (Olson and Fazio, 2004). Participants completed two IATs, each of which consisted of seven blocks with 20 trials in noncritical blocks and 40 trials in critical blocks. Depending on condition, the positive/negative category labels appeared as “I like/ I don’t like” (personalized IAT), “people like/ people don’t like” (culturally-oriented IAT), or “pleasant/unpleasant” (traditional IAT).
Results and Discussion

The IAT data were prepared as in Experiment 1.

*Original Scoring Method*

The IAT scores were then entered into a 2 (IAT scores: first vs. second) × 2 (first IAT type: personalized vs. culturally-oriented) ANOVA with repeated measures on the first factor. As expected, the results revealed a significant interaction between the IAT scores factor and the type of IAT that was completed first, $F(1,64) = 18.59, p < .001$. Decomposition of the interaction indicated that scores on the first IAT (personalized or cultural IAT) differed as predicted. Participants who completed the personalized IAT showed a reduced IAT score, indicating smaller racial bias ($M = 142, SD = 191$) compared to participants in the culturally-oriented IAT condition ($M = 239, SD = 202$), $t(64) = 2.01, p < .049$. More importantly, the second (traditional) IAT scores differed as a function of the preceding IAT, $t(64) = 2.26, p < .029$. Participants who completed the personalized IAT first exhibited a significant increase in race bias on the second traditional IAT compared to the first personalized IAT ($M = 235, SD = 179$), $t(35) = 3.63, p < .001$, while the opposite occurred for participants who completed the culturally-oriented IAT first. These individuals demonstrated a significantly smaller traditional IAT score and reduced racial bias compared to the first culturally-oriented IAT ($M = 151, SD = 179$), $t(29) = 2.53, p < .017$.

*D-Score Algorithm*

The D-scoring algorithm of the IAT (see Greenwald, Nosek, & Banaji, 2003) produced similar results. The algorithm first involved elimination of data from participants with trials greater than 10,000 ms (0 cases) and for whom more than 10% of trials have latencies less than 300ms (2 cases). Since participants in the personalized IAT did not receive error messages, we did not impose the 600ms error penalties for either personalized and traditional IAT, so as to
keep the traditional D-score equivalent to the personalized. Instead, we included only the
correct trials in the analysis. The means for each of the 6 critical blocks and three pooled
standard deviations for all trials were computed. The different scores of mean latencies were
computed and then divided by its corresponding pooled standard deviations. Finally, they were
averaged to form a D-score. We submitted the two D-scores into a 2 (IAT scores: first vs. second)
× 2 (first IAT type: personalized vs. culturally-oriented IAT) ANOVA with repeated measures on
the first factor. The analysis revealed a significant D-IAT scores × First IAT type completed
interaction, $F(1,64) = 14.51, p = .001$. A t-test revealed that participants who completed the
personalized IAT showed a smaller race bias ($M = .34, SD = .58$) on the first IAT than those who
completed the culturally-oriented IAT ($M = .60, SD = .53$), $t(64) = 1.83, p = .067$. The converse
occurred for the second (traditional) IAT; participants who had completed a personalized IAT
first showed greater bias against Blacks than did those who had first completed the cultural IAT
($M = .67, SD = .50$ vs. $M = .44, SD = .43$, respectively), $t (64) = 1.92, p = .062$. Moreover,
participants’ traditional D-IAT scores increased significantly relative to their first IAT when they
had completed the personalized IAT first, $t(36) = 3.63, p < .001$, but decreased significantly
when they had completed the culturally-oriented IAT first, $t(29) = 1.82, p = .08$.

Experiment 2 demonstrated that scores on the traditional IAT can fluctuate as a function
of the preceding IAT type. When participants completed the personalized IAT first, traditional
IAT scores increased significantly, indicating a racial bias similar in magnitude to that revealed by
the culturally-oriented IAT. However, when the culturally-oriented IAT preceded the traditional
IAT, the latter showed a significantly reduced racial bias, resembling the personalized IAT. These
results suggest that the traditional IAT can be influenced by contextual information such as
conversational norms. It appears that participants interpreted the second traditional IAT
differently as a function of the type of IAT they completed first so as to avoid redundancy with the previous IAT. For example, if participants completed the personalized IAT with labels “I like/I don’t like” first, when then faced with “pleasant/unpleasant” labels, they appear to have concluded that the researchers were no longer interested in their own personal evaluations, but instead were more interested in other people’s associations. Otherwise, why would the experimenter have changed the labels? Likewise, if participants completed the culturally-oriented IAT first, the change in labels appears to have suggested that the researcher no longer wished the task to concern general evaluations. Thus, it appears that the labels of the traditional IAT are ambiguous enough to allow participants to shift their meaning as the situation calls for—in this case, to avoid redundancy and comply with the rules of conversational norms.

GENERAL DISCUSSION

The first two experiments demonstrate the inherent ambiguity present on the IAT. Specifically, the ambiguity of the evaluative labels, “Pleasant/Unpleasant.” It appears that these labels are ambiguous enough to allow for shift in its meaning depending on the context. Experiment 1 showed that participants can construe the labels differently depending on the mindset or the perspective they have adopted in a previous unrelated task, i.e., a personal or a normative focus. And Experiment 2 illustrated that the IAT can be influenced by “informal” contextual factors such as conversational norms. When participants first complete an IAT with clear evaluative labels, either personalized or culturalized, they adopt a different meaning of the labels “pleasant/unpleasant” to avoid redundancy with the previous IAT. These results suggest that, depending on the previous task or instructions, participants may actually complete the IAT in ways different than that intended by the researcher, raising concerns about the reliability and validity of those measures. If the traditional IAT can indeed be affected by various context
effects and factors other than respondent's personal associations, one must carefully consider its role when interpreting the outcome data.

These two experiments also illustrate how implicit measures, such as the IAT, are not as robust to the ambiguity problems that plague the explicit survey measures. It appears that implicit measures, at least the IAT, can be influenced by the same problems that impact the quality of the data obtained in surveys, such as preceding question issues and the conversational norms. Together the findings emphasize the importance of fully understanding a measure and its assumptions before it is employed to test theory. The next chapter addresses the important consequences of the IAT’s malleability for the inferences that can be drawn regarding the presence or absence of attitude change.
CHAPTER 3

CONSEQUENCES OF MALLEABILITY OF THE IAT: MALLEABILITY OF ATTITUDES OR THE MEASURE?

Experiments 1 and 2 have illustrated that the ambiguity of the labels in the traditional IAT can lead participants, just like in ambiguous surveys, to interpret the measure in different ways. These results emphasize the importance of carefully considering the role of context effects and other influences on the traditional IAT when interpreting the data. It is possible that the malleability of the IAT may result in misleading interpretations of the data, especially regarding inferences about attitude change.

For example, a growing body of literature has demonstrated that the attitude estimates provided by implicit measures can be quite malleable and often depend on situational and contextual factors (see Blair, 2002 for a review). Although many theoretical accounts can explain the context effects, such as the attitudes-as-construction approach, which suggests that attitudes are constructed on the spot (Schwarz & Bohner, 2001; Wilson & Hodges, 1992), or the change-of-object perspective, which claims that the attitude object itself is construed as a different object depending on the context (e.g., Asch, 1948; Barden, Maddox, Petty, & Brewer, 2004; Fazio, 2007; Fazio & Dunton, 1997), perhaps the most popular account is that of the malleability of automatically-activated attitudes. According to this perspective, automatically-activated attitudes are often impacted and changed by “self and social motives, specific strategies, the perceiver’s focused attention, and configuration of stimulus cues” (Blair, 2002, p. 242).
One of the most frequently-cited illustrations of the malleability of automatically activated attitude as induced by situational cues was provided by Dagupta and Greenwald (2001). In their first study, participants were exposed to a set of famous and well-liked African Americans and infamous and disliked White Americans (e.g., Michael Jordan and Ted Bundy) or conversely, famous well-liked White Americans and infamous disliked African Americans (e.g., John F. Kennedy and Mike Tyson). Immediately after this task, participants completed an IAT assessing racial attitudes. Results revealed that those participants who were exposed to positive Black and negative White exemplars showed a significantly reduced racial bias compared to participants in the opposite condition (negative Black and positive White exemplars). In their second study, Dasgupta and Greenwald (2001) replicated the finding using positive elderly/negative young exemplars (e.g., Mother Teresa/Tonya Harding) or negative elderly/positive young exemplars (e.g., Ted Kaczynski/Prince William). Again, they demonstrated that when participants were exposed to the pro-elderly exemplars before the IAT, the bias against the elderly on the IAT was significantly reduced compared to those exposed to the pro-youth exemplars. On the basis of these studies, the authors concluded that automatic preferences and attitudes were malleable. That is, “implicit evaluations of historically stigmatized groups such as African Americans and older people may be modified, at least temporarily, by repeatedly reminding people of admired members of those groups and of disliked members of high-status reference groups” (Dagupta & Greenwald, 2001, p. 806).

Another study frequently cited as evidence for the malleability of automatically-activated attitudes is that of Wittenbrink, Judd, and Park (2001). In their first experiment, participants watched video clips depicting African Americans in either a stereotypically positive or negative fashion (e.g., a harmonious barbecue scene vs. a violent gang scene) after
completing a pre-manipulation race IAT. Following exposure to the movie, they completed another race IAT as a “distracter task” with several 20-second clip reminders of the movie between the critical blocks. The results showed that participants who were exposed to the stereotypically positive movie clip showed a significantly reduced race bias on the post-movie IAT compared to those who were exposed to the stereotypically negative movie. Similar to Dasgupta and Greenwald (2001), the results from this study were interpreted as a demonstration of the malleability of automatically activated racial attitudes. According to the authors, the findings demonstrate the “variability of automatic responses due to changes in stimulus context” (Wittenbrink et al., 2001, p. 815).

In another example, in a paper entitled, “The power of a story: New, automatic associations from a single reading of a short scenario,” Foroni and Mayr (2005) reported research in which participants read a counterfactual scenario regarding insects and flowers, one in which participants were induced to imagine a post nuclear war world where insects were more favorable than flowers. The authors found that participants showed a greater preference for insects on the IAT after reading the scenario, and interpreted the result as evidence for attitude change, as highlighted by their provocative title. However, how reasonable it is to conclude that longstanding negativity towards insects can be changed after reading a short imaginary scenario about positive insects? Would anyone maintain that, after such a mental exercise, an individual would be less likely to step on ants or swat flies, or be less appreciative of flowers? If an effect of exposure to the scenario were apparent on an explicit measure, would one conclude that attitude change has occurred? Or, would one be more likely to question the validity of the measure itself, possibly wondering whether participants had interpreted the measure as intended?
Although the above studies, as well as others (e.g., Blair et al., 2001; Carpenter & Banaji, 2001; Lowery, Harden, & Sinclair, 2001; Mitchell et al., 1999), suggest the possibility that automatic activated attitudes could be malleable, it is important to recognize that inferences about change in a construct in memory are being drawn from the observation of changes on a measurement outcome (De Houwer & Moors, 2007). These need not be isomorphic. Change in observed scores may occur through mechanisms other than change in the attitudinal representation in memory.

Indeed, it would be difficult to argue that the manipulation we employed in Experiments 1 and 2 prompted any change in participants’ attitudes toward African-Americans. It seems exceedingly unlikely that the consideration of personal versus normative likes and dislikes in a preceding unrelated task, or that conforming to the implicit rules of conversational norms, could affect a participant’s representation of Blacks. Yet there were effects on the IAT. Thus, Experiments 1 and 2 demonstrate malleability of the IAT in the absence of any reason to expect that attitudes themselves were changed in any way.

Experiments 3 and 4 focus on one particular way in which a disparity between attitude representation and IAT performance may occur. As mentioned above, prior studies exposed participants to contextual and situational information and found such exposure to influence subsequent performance on an IAT. I explored whether these contextual manipulations, instead of influencing attitudes, may provide participants with highly salient extrapersonal associations that can then be used to solve the IAT’s mapping problem. The question addressed in the next two experiments is: might these contextual manipulations increase the salience of extrapersonal associations on the IAT instead of “changing” participants’ attitudes? That is, are these studies demonstrating the malleability of the IAT labels and not necessarily the malleability of the
attitude? Specifically, the experiments examine whether IAT performance would prove sensitive
to a brief contextual manipulation that had no plausible bearing on the relevant attitudinal
representations.

EXPERIMENT 3

Experiment 3 explored whether the manipulations employed by Foroni and Mayr
(2005), instead of influencing attitudes, may have provided participants with highly salient
extrapersonal associations that could then be used to solve the IAT’s mapping problem. As
noted earlier, we do not find it very plausible that a brief imaginary excursion into a
counterfactual world could produce a change in longstanding attitudinal representations
regarding insects and flowers. To explore this hypothesis, the current experiment closely
mirrored Foroni and Mayr’s (2005) paradigm with two crucial modifications. First, the scenario
and the IAT were presented as two clearly separate experiments, and hence, unrelated tasks.
The original version asked participants to keep the scenario in mind when completing the IAT.
Indeed, the IAT was presented as a “tutoring program” intended to help individuals firmly
establish like for insects and dislike for flowers. Given this instruction, one might argue that
Foroni and Mayr’s findings reflect the IAT’s sensitivity to the contextual demand that insects be
momentarily viewed positively and flowers negatively. However, the present view of the IAT’s
malleability maintains that it would be affected even if the imagination exercise were concluded
prior to the administration of the IAT. The mental imagery, even though obviously based on
fiction and, hence, unlikely to affect individuals’ attitudes, would provide salient extrapersonal
associations that would facilitate mapping insects and “pleasant” on to one key and flowers and
“unpleasant” on to the other. Just as the normative mindset in Experiment 1 facilitated
associating Blacks with negativity, the scenario imagined in the preceding task would facilitate associating insects (flowers) with positivity (negativity).

Second, because the personalized version of the IAT is more robust to the influence of extrapersonal associations (Olson & Fazio, 2004; Han et al., 2006), the personalized IAT was included as a crucial comparison condition. If automatically activated attitudes were readily influenced by exposure to the scenario, the changes should be reflected in both the personalized and traditional IATs. However, if the malleability effects were driven by salient extrapersonal associations, we should observe the changes only in the traditional version.

Method

Participants

One hundred and seventeen psychology students participated in this experiment in partial fulfillment of their course requirements. The data from one participant was excluded because she spent less than 1.5 seconds reading the scenario (mean reading time = 42 seconds). The final sample consisted of 116 participants (49 males, 67 females).

Materials and Procedure

The experiment was conducted in laboratory conditions identical to the previous experiments. In the experimental condition, participants learned that they would be participating in two different studies. They were told that the first study was about one’s ability to imagine and generalize from a short story and the second was about one’s categorization abilities. After they answered some demographic questions, participants read a scenario regarding insects and flowers adapted from Foroni and Mayr (2005).2 In the scenario,

2 The scenario presented was identical to the appendix material of Foroni and Mayr (2005), except for the last three sentences in which the IAT was presented as a training program.
participants were asked to imagine that they were survivors of nuclear war and, because of radioactivity, flowers were no longer safe to eat. Insects, however (due to their quick mutation) were safe to eat and could be used to feed higher-level animals (e.g., sheep, cows).

Imagine you are a survivor of a nuclear war. Years after the war, the radiation still affects your life. The only food resources you have are certain higher-level animals (e.g., sheep, cows). You can breed them but you have to make sure that they don’t eat or touch any flowers. All flowers are radioactive and they cannot be used as nutrition for your animals. The only way you have to raise your animals is to feed them with all kinds of insects. In fact, because of their fast metabolism and genetic mutations, insects have already overcome the radiation problems. In this desolate scenario your survival completely depends on the insects. They are the most positive things you can imagine. In contrast, flowers are poison for you and your animals. They are the most negative things you can imagine. To survive under such circumstances, it is extremely important to make quick and accurate choices regarding flowers versus insects.

After reading this scenario, participants completed a 20-item questionnaire regarding what foods they thought were safe to eat in this imagined world and what their quality of life would be like on a 7-point Likert scale (e.g., Think carefully about the potential food sources for the animals. Answer whether following food sources would be a good source of food and water: ROSES. Anchored by (1) definitely NOT good to eat, and (7) definitely good to eat; What you think your quality of life will be in this new world? (1) extremely bad, and (7) extremely good).

These items were included to further bolster the cover story, making it appear that the initial imagination experiment had come to its conclusion (see Appendix C).

As in Experiment 1, participants were then directly told that the first experiment had ended, and that it was time for the second experiment on categorization. All participants then completed either a traditional or a personalized insects/flowers IAT as part of the categorization
task experiment. The two versions of the IAT were identical except for the evaluative labels and the type of error feedback provided. The evaluative labels appeared as “I like/ I don’t like” (personalized IAT), or as “pleasant/unpleasant” (traditional IAT). Furthermore, those in the personalized IAT condition received error feedback only in the insects/flowers categorization block and did not receive feedback during the practice block concerning the categorization of valence, nor during the critical blocks. The parameters of the IAT were identical to Experiment 1 except it involved categorization of insects and flowers (e.g., “Tulips” “Roaches”).

The experiment also included a control condition in which participants did not read a scenario but only completed either a personalized or traditional insects/flowers IAT. Although there was no reason to expect a difference between personalized and traditional versions of an IAT regarding pre-existing preference for flowers versus insects, this control condition permitted gauging of effects of the scenario relative to baselines for each IAT version.

Results and Discussion

The IAT data were prepared as in the previous two experiments. The compatible blocks were identified as flowers/ + vs. insects/ – and incompatible block as flowers/ – vs. insects/ +. Higher scores thus indicated greater dislike for insects or greater preference for flowers.

Original Scoring Method

The IAT scores were entered into a 2 (story or no story) X 2 (IAT type: personalized vs. traditional) ANOVA. The results revealed a significant scenario x IAT type interaction, $F(1, 115) = 10.49, p = .002$. As expected, the scenario influenced the traditional IAT. When the scenario was present, the preference for flowers (or dislike for insects) was significantly reduced compared to when no story was present, ($M = 178, SD = 170$ vs. $M = 306, SD = 190$), $t(58) = 3.16, p = .011$. However, the personalized IAT was not impacted by the scenario. Regardless of whether the
scenario was present or not, the personalized IAT showed a strong dislike for insects (and
preference for flowers) (scenario: $M = 354, SD = 164$; control: $M = 279, SD = 178$), $t(54) = 1.39$,
p = .18 (Figure 1).

**D-Score Algorithm**

D-scores showed similar results. No participants displayed any latencies greater than
10,000 ms. The data from two were excluded from analyses because they violated the scoring
algorithm’s requirement that no more than 10% of trials have latencies less than 300ms. ³

The D600 scores were entered into a 2 (scenario: present vs. absent) × 2 (IAT type:
personalized vs. traditional) ANOVA. The results revealed a significant condition main effect,
$F(1, 113) = 5.36, p = .039$, such that the control condition was characterized by a higher
preference for flowers (vs. insects) than the experimental condition. However, the main effect
was qualified by the predicted scenario × IAT type interaction, $F(1, 113) = 4.57, p = .035$. As
expected, the scenario influenced the traditional IAT. When the scenario was present, the
preference for flowers (or dislike for insects) was significantly reduced compared to when no
story was present, ($M = .57, SD = .50$ vs. $M = .89, SD = .42$), $t(57) = 2.63, p = .011$. However, the
personalized IAT was not impacted by the scenario. Regardless of whether the scenario was

³ As discussed in Chapter 1, one might ask whether the usual absence of error feedback in the
personalized IAT played a role in observing the difference between the two versions of the IAT.
Replicating previous research, there was no evidence of differential error rates in the traditional ($M =$
8.0%, $SD = 6.3$) vs. personalized IAT versions ($M = 7.4%, SD = 6.7$), $t(115) < 1$. Thus, the differences with
respect to the validity of the two IAT versions cannot be dismissed as due to a lesser understanding of the
task instructions or a lesser emphasis on accuracy when the IAT is personalized.
present or not, the personalized IAT showed a strong dislike for insects (and preference for flowers) (scenario: $M = .84$, $SD = .25$; control: $M = .84$, $SD = .39$), $t(53) < 1$(Figure 2).

Thus, exposure to the scenario significantly impacted the traditional IAT. This occurred even though the scenario was presented as a separate and unrelated experiment. Presumably, having read the scenario earlier in the session increased the salience of extrapersonal associations which participants could use to solve the mapping problem posed by the traditional IAT.

EXPERIMENT 4

In the interest of convergent validity and generalizability, Experiment 4 sought to explore the malleability of the traditional IAT in a social domain and in a manner that did not rely on the imagining a fictitious world.

Experiment 4 sought to replicate and explore one of the most well-known and cited demonstrations of the malleability of automatically activated attitudes— that of Dasgupta and Greenwald’s (2001, Experiment 2). As mentioned above, that experiment presented participants with either positive elderly/ negative young or negative elderly/positive young exemplars, following which participants completed a traditional version of the young/elderly IAT with the usual unpleasant/pleasant attribute labels. A substantially reduced IAT effect was observed for participants exposed to the positive elderly/ negative young exemplars compared to participants who had viewed the negative elderly/positive young exemplars. On the basis of the perspective underlying the current research, this result may more appropriately reflect a malleability of the IAT measure rather than genuine attitude change. It is possible that, for some participants, the positive elderly exemplars temporarily made some extrapersonal associations salient, to which the ambiguous “pleasant/unpleasant” labels were sensitive. If this was the
case, then with the same procedure, the use of the more focused “I like/I don’t like” labels would not detect a change in the IAT scores. To test these predictions, Dasgupta and Greenwald’s (2001) methods were mirrored, but with the personalized IAT as a crucial comparison condition. It was predicted that only responses on the traditional IAT would be affected by the potential extrapersonal associations introduced by the exemplars to which participants were exposed.

Method

Participants

One hundred twenty-eight introductory psychology students (42 males, 86 females) participated in this experiment in partial fulfillment of their course requirements. They were randomly assigned either to the elderly positive or control condition. We excluded 13 Asian participants, due to evidence suggesting that Asians’ stereotypes of the elderly are mostly positive and one non-native speaker who could not correctly complete the general knowledge task. The final sample consisted of 114 participants (39 males, 75 females).

Procedure

Participants were randomly assigned to one of four conditions of a 2 (exemplar condition: elderly +/young – vs. elderly –/young +) × 2 (IAT type: personalized vs. traditional) between-subjects design. After they answered some demographic questions, participants learned that the purpose of the study was to explore one’s general knowledge and categorization abilities. As in Dasgupta and Greenwald (2001), contextual information was first presented as a “general knowledge” task. In this task, participants were presented with an image of a target person, the person’s name, and both a correct and incorrect description of the person (e.g. Mother Teresa: “Was a true champion of the poor and destitute” or “A famous
novelist") (see Appendix D). The participant’s task was to identify the correct description, and
the valence of the description was held constant for each picture in case of error. Depending on
the condition, half of the participants saw positive elderly (e.g., Albert Einstein) and negative
young (e.g., Ted Bundy) exemplars while the other half saw negative elderly (e.g., Leona
Helmsley) and positive young (e.g., Prince William) exemplars. Ten exemplars of each type were
presented, and they each were shown twice for a total of 40 trials.

After the “general knowledge task,” all participants completed either a personalized or a
traditional version of the IAT. Both versions of the IAT were identical except for the evaluative
labels and the type of error feedback provided. The evaluative labels appeared as “I like/ I don’t
like” (personalized IAT), or as “pleasant/unpleasant” (traditional IAT). Those in the personalized
IAT received error feedback only in the name categorization block and did not receive feedback
on the categorization of valence or the critical blocks. To parallel Dasgupta and Greenwald
(2001), we used the same 12 block version of the IAT, involving the categorization of old and
young names (e.g., “Ethel”, “Kyle”) that they employed (see Appendix E). The order in which the
participants performed the critical combined blocks was counterbalanced.

Results

The IAT data were prepared as in previous experiments with adjustments for the 12-
block version of the IAT (e.g., 6 critical blocks and 3 pooled standard deviations). Compatible
blocks were identified as young / pleasant (I like) vs. elderly / unpleasant (I don’t like) and

4 The valence of the two alternatives was held constant so that participants would be exposed to
the desired valence even if they were unaware of the correct description.
incompatible blocks as young / unpleasant (I don’t like) vs. elderly / pleasant (I like) with a higher IAT score indicating greater bias against the elderly or greater preference for the young.  

_original scoring method_

The IAT scores were entered into a 2 (exemplar condition: elderly+/young– vs. elderly–/young+) × 2 (IAT type: personalized vs. traditional) ANOVA. The analysis revealed a significant exemplar main effect, \( F(1, 110) = 7.32, p = .008 \) such that the positive young/negative elderly exemplar condition resulted in greater elderly bias. However, this main effect was qualified by an interaction. As predicted, IAT scores varied as a function of IAT type and the exemplars that were presented in the initial phase of the experiment. This was evidenced by a significant Exemplar × IAT type interaction, \( F(1, 110) = 3.91, p = .05 \). Decomposition of the interaction confirmed that we successfully replicated Dagupta and Greenwald’s (2001) finding in the traditional IAT condition. When liked elderly and disliked young exemplars had been presented, the elderly bias was significantly reduced (\( M = 197, SD = 97 \)) compared to when disliked elderly and liked young were presented (\( M = 281, SD = 108 \), \( t(53) = 3.83, p < .001 \). Scores on the personalized IAT, however, were unaffected by the exemplar presented, \( t (57) < 1 \), (\( M = 254, SD = 141 \) for elderly + and \( M = 257, SD = 121 \) for elderly –), suggesting that only the traditional IAT was affected by the context manipulation (Figure 3).

₅ Although the personalized version of the IAT (\( M = 7.2\% , SD = 6.05 \)) had slightly higher error rates than the traditional version (\( M = 6.2\% , SD = 3.31 \), \( t(109) = 1.70, p = .084 \), even this marginal difference was not consistent across studies and, hence, cannot be characterized as a reliable pattern. Recall that the error rate differential in Experiment 3 did not even approach significance; indeed, the means themselves were in a direction opposite to that observed here.
Once again, the D-score showed similar results. No latencies exceeded 10,000 ms. However, the data from three participants for whom more than 10% of the trials had latencies less than 300 ms were excluded from the D600 analysis.

The D600 scores were entered into a 2 (exemplar condition: elderly+/young– vs. elderly–/young+) × 2 (IAT type: personalized vs. traditional) ANOVA. The analysis revealed a significant exemplar main effect, $F(1, 107) = 6.80$, $p = .01$, such that the elderly–/young + exemplar condition resulted in greater elderly bias than the reverse condition. As predicted, however, IAT scores varied as a function of IAT type and the exemplars that were presented in the initial phase of the experiment. This was evidenced by a significant exemplar × IAT type interaction, $F(1, 107) = 4.18$, $p = .043$. Decomposition of the interaction revealed that when liked elderly and disliked young exemplars had been presented, the elderly bias was significantly reduced ($M = .57$, $SD = .28$) compared to when disliked elderly and liked young were presented ($M = .81$, $SD = .22$), $t(52) = 3.53$, $p < .001$. Scores on the personalized IAT, however, were unaffected by the exemplar presented, $t(55) < 1$, ($M = .64$, $SD = .29$ for elderly + and $M = .67$, $SD = .31$ for elderly –) (Figure 4).

The results from Experiment 4 indicate that when participants were shown positive elderly exemplars, their scores on the traditional IAT were significantly reduced to show less elderly bias. This did not occur on the personalized IAT. The absence of an effect on the personalized IAT suggests that the context effects demonstrated on the traditional IAT may stem from its sensitivity to salient extrapersonal associations, instead of representing change in the attitude representation itself.
GENERAL DISCUSSION

Experiments 3 and 4 suggest that the malleability effects evidenced in prior research using the traditional IAT may stem from this measure’s inherently ambiguous labels. The labels “pleasant/unpleasant” seem sensitive to salient extrapersonal associations, and do not necessarily reflect changes in the attitude representation itself. Especially informative in this respect is the apparent absence in either experiment of any contextual effects on the personalized IAT—a version of the IAT using more focused labels that are less susceptible to the influence of extrapersonal associations. It appears that what sometimes has been interpreted as indicative of the malleability of automatically-activated attitudes may have been driven by changes in the information participants considered while completing the IAT, rather than reflecting a real change in participants’ attitudes.

Experiment 3 showed that imagining atypical scenarios, or thinking about insects and flowers in an atypical manner, before completing the IAT can significantly reduce the scores obtained on the traditional version. In Experiment 4, the context manipulation that preceded the IAT made specific atypical (or typical exemplars) salient. It appears that these atypical exemplars or scenarios can serve as extrapersonal associations that facilitate management of the response mapping problem that the IAT poses. Although such exemplars are unlikely to have been the basis for one’s attitude toward the general category, keeping one or more of these atypical exemplars (e.g., Mother Teresa or Einstein) in mind will make it easier to associate the dual meanings of a given response key (e.g., elderly/pleasant). From the perspective of these atypical exemplars, the category is relatively more positive. The likelihood of such a perspective shift is attenuated by the personalized focus of the “I like/I don’t like” labels.
A second mechanism could have contributed to the findings observed in Experiment 4. The exemplars from the earlier task may have disambiguated the target category labels (elderly versus young). Several experiments have demonstrated that the presentation of atypical exemplars *during* the IAT can impact the resulting attitude estimates, presumably because participants redefine the category labels in terms of the exemplars included as stimuli (e.g., Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003). Experiment 4’s presentation of atypical exemplars in a preceding context may have promoted a similar redefinition of the category labels, even though these exemplars were encountered in a completely separate task bearing no connection to the IAT and, hence, should not have had any implications for the meaning of the category labels within the IAT task.

Importantly, however, the results indicate that the personalized IAT was not affected by such category redefinition processes. Why not? One might reason that because the category labels for the personalized and traditional IATs were the same (elderly/young), the personalized IAT should also be impacted by contextual factors that can help disambiguate the labels. Experiments in Chapter 4 explore these very questions.

A second mechanism could have contributed to the findings observed in Experiment 4. The exemplars from the earlier task may have disambiguated the target category labels (elderly versus young). Several experiments have demonstrated that the presentation of atypical exemplars *during* the IAT can impact the resulting attitude estimates, presumably because participants redefine the category labels in terms of the exemplars included as stimuli (e.g., Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003). Experiment 4’s presentation of atypical exemplars in a preceding context may have promoted a similar redefinition of the category labels, even though these exemplars were encountered in a completely separate task.
bearing no connection to the IAT and, hence, should not have had any implications for the meaning of the category labels within the IAT task.

Importantly, however, the results indicate that the personalized IAT was not affected by such category redefinition processes. Why not? One might reason that because the category labels for the personalized and traditional IATs were the same (elderly/young), the personalized IAT should also be impacted by contextual factors that can help disambiguate the labels. Experiments in Chapter 4 explore these very questions.
CHAPTER 4

STIMULUS EXEMPLAR EFFECTS: AMBIGUITY OF THE CATEGORY LABELS?

Experiments 3 and 4 investigated the evidence purported to demonstrate the malleability of the automatically activated attitudes when participants are exposed to varying contextual information prior to completing the IAT. In Experiments 5 and 6, I explore what happens when the contextual information is presented as stimuli during the administration of the traditional vs. personalized IAT. That is, the aim was to explore stimulus exemplar effects on the IAT as a function of the assumed malleability of the measure.

Many previous studies have explored stimulus exemplar effects, by examining the consequences of using atypical stimuli within the traditional version of the IAT (Steffen & Plewe, 2001; Govan & Williams, 2004; Mitchell, Nosek, & Banaji, 2003; Bluemke & Friese, 2005). Often, these stimulus exemplar effects are interpreted as illustrative of some form of attitude malleability - not strictly as attitude change but rather as demonstrating the online construction of automatic attitudes depending on the context. For example, Mitchell et al. (2003) presented participants with individually tailored well-liked blacks and disliked whites, or disliked blacks and liked whites, as stimuli on the IAT and found that participants showed less racial bias in the former case. The authors interpreted these findings as “initial evidence of rapid automatic attitude change” (p. 457) and concluded that automatic attitudes must be defined and constructed by the situation and context.
In yet another illustration of stimulus exemplar effect, Govan and Williams (2004) presented participants with atypically valenced flowers (poison ivy) and insects (butterfly) and found that the typical insects/flowers IAT effect, preference for flowers, could be reversed. The results were interpreted as indicating that participants temporarily redefined the intended meaning of the category labels insects and flowers. To directly test the idea of category redefinition Govan and Williams (2004) had participants first complete a nasty plant/nice animal IAT (e.g., poison ivy/ puppy), or a nice plant/nasty animal IAT (e.g., lily/pit-bull) in the second study. Then participants completed a “category” IAT in which the stimuli were pleasant/unpleasant items and the mere words “animal” and “plant.” Results revealed that on the category IAT, plants and animals were categorized in accord with the exemplars experienced earlier. Shortly after completing the nasty plant/nice animal IAT, participants found it easier to associate plant with “unpleasant” and animal with “pleasant.” The reverse was true for participants who completed the nice plant/nasty animal IAT. These results were interpreted as supporting the redefinition of the category hypothesis, whereby atypical stimuli can, at least temporarily, redefine how participants view the category as a whole.

So far, the research on stimulus exemplar effects has demonstrated its effects using only the traditional version of the IAT. In light of the findings from Experiments 3 and 4 attesting to the value of personalizing the IAT, one has to wonder what would happen to the stimulus exemplar effect if the personalized IAT were employed. Would the personalized IAT also be sensitive to the presentation of atypical stimuli during the task and prompt participants to temporarily refine the category? Recall that results from Experiment 4 revealed that when atypical exemplars were presented before the IAT in the context of a preceding task, the personalized IAT was not impacted whereas the traditional IAT was. Would the same occur
when the atypical exemplars are presented within the IAT? That is, can personalizing the IAT also have the benefit of inhibiting category redefinition and, hence, be robust to stimulus exemplar effects?

There is a sense in which this does not appear very likely. After all, the only category labels changed by personalizing the IAT are the ones referring to valence (I like/I don’t like replacing Pleasant/Unpleasant). The other category labels (e.g., Black/White or elderly/young) remain the same. Thus, even when the IAT is personalized, atypical exemplars may prompt reconstrual of the category.

However, there is another possibility. Although no research has explored this question, the provision of clearly focused labels such as “I like/don’t like” may provide a perspective-specific mindset that makes it difficult for participants to construe a category in any way other than implied by that perspective. Thus, evaluative labels that focus the participant on “I” may inhibit any tendency to view the category label differently than one typically does. Not only does the “I” lead to a focus on personal likes and dislikes, but it also may restrict the potential meanings of the category labels to those that are most dominant for the participant. In other words, it may be difficult to think of the category label “elderly” any differently than one typically does when the “I” labels encourage the adoption of a personal mindset.

Experiments 5 and 6 explore these two possibilities. The experiments ask whether stimulus exemplar effects and redefinition of the category are promoted by the ambiguity present in category labels vis-a-vis the ambiguity inherent in the evaluative labels. That is, does the perspective ambiguity in the evaluative labels (Pleasant/unpleasant to whom) also allow for shifting construal of the category labels (Elderly/young according to whom)? Thus, Experiments 5 and 6 compare the susceptibility of the traditional versus personalized IAT to the stimulus
exemplar effect. They investigate whether focusing the valence perspective of the IAT might also have the consequence of limiting construals of the category labels.

EXPERIMENT 5

Method

Participants

Eighty-one introductory non-African American psychology students (43 males, 38 females) participated in this experiment in partial fulfillment of their course requirements. Two participants were removed from the analysis because they had lived in the U.S. less than 2 years and were unable to categorize Black/White names. The final sample consisted of seventy-nine participants (43 males, 36 females).

Procedure

The procedure for Experiment 5 was similar to the one used by Govan and Williams (2004, Study 1b), with one crucial difference – we also included the personalized IAT. Depending on the condition, participants completed either a personalized or traditional black/white race IAT with typical (e.g., Tyrone/Jason) or atypical exemplars. Stimulus items for atypical Black exemplars were Michael Jordan, Bill Cosby, Eddie Murphy, Oprah Winfrey, and Danzel Washington, and atypical White exemplars were Charles Manson, Adolph Hitler, Hannibal Lechter, Jeffrey Dahmer, and Timothy McVeigh (Appendix E).6

---

6 The stimuli items were identical to Govan and Williams (2001) except for four names which were only well-known in Australia, the location in which their participants resided. Paulin Hanson and Martin Bryant were changed to Jeffrey Dahmer and Timothy McVeigh, and Cathy Freeman and Ernie Dingo to Oprah Winfrey and Danzel Washington.
Results and Discussion

Original Scoring Method

The IAT scores were entered into a 2 (IAT type: Personalized vs. Traditional) X 2 (Stimulus Typicality: Typical vs. Atypical) ANOVA. The analysis revealed a significant main effect of typicality, such that the typical stimuli conditions showed greater race bias, $F(1,75)=10.21$, $p=.002$. However, this main effect was qualified by a Typicality X IAT type interaction, $F(1,75)=10.58$, $p=.002$. Racial bias was reduced when atypical Black and White stimuli were employed on the traditional IAT ($M=51.02$, $SD=153.87$) compared to when typical stimuli were presented ($M=259.50$, $SD=134.13$), $t=4.50$, $p<.001$. This result replicates Govan and Williams’ (2004) finding. In contrast, no effects of stimulus typicality were apparent on the personalized IAT (for the typical condition, $M=179.38$, $SD=132.46$; for the atypical condition, $M=180.57$, $SD=161.87$), $t<1$ (Figure 5).

Additional comparisons showed that the personalized IAT with typical stimuli reflected a smaller race bias ($M=179.38$) than the traditional IAT with typical stimuli, ($M=259.47$), $t=2.3$, $p=.05$, thus replicating Olson and Fazio’s (2004) findings. With the atypical exemplars, this effect of IAT type was reversed, $t(41)=2.62$, $p=.012$, such that the traditional IAT showed a smaller race bias ($M=51.02$, $SD=153.8$) than the personalized version ($M=180.57$, $SD=161.87$).

D-Score Algorithm

D-score results mirrored that of the original scoring method. No data needed to be removed due to latencies greater than 10,000 ms or to participants for whom more than 10% of the trials involved latencies less than 300 ms.\textsuperscript{7} The analysis revealed a main effect of typicality of

\textsuperscript{7} Again, there was no evidence of differential error rates in the traditional ($M = 8.0\%, SD = 7.8$) vs. personalized IAT versions ($M = 8.1\%, SD = 7.5$), $t(77) < 1$.  

46
the stimulus exemplars such that typical exemplars produced greater race bias, $F(3,75)= 11.47$, $p<.001$. However, the effect was qualified by an IAT type X Stimulus exemplar interaction, $F(3,75)= 8.16$, $p=.006$. On the traditional IAT, D-scores differed significantly as a function of the stimulus exemplars presented, $t(38)=4.16$, $p<.001$, such that atypical exemplars revealed a smaller race bias ($M=.12$, $SD=.44$) than did typical stimulus exemplars ($M=.60$, $SD=.24$). The personalized IAT, on the other hand, was not impacted by the stimulus exemplars, $t(37) < 1$, ($M=.35$, $SD=.31$ for atypical and $M=.40$, $SD=.32$ for typical). Analysis also showed that the typical stimuli conditions differed significantly, $t(34)=-2.17$, $p<.037$, such that the traditional IAT showed greater race bias compared to the personalized IAT. As expected, significant differences on the atypical IAT conditions were also found, $t(41)=2.01$, $p<.052$, such that the personalized IAT showed a greater race bias than the traditional, suggesting that the personalized was not influenced by the atypical stimuli (Figure 6).

It appears that the personalized IAT is not susceptible to atypical exemplar effects. That is, it seems that stimulus items do not matter on the personalized version of IAT, and what seems to drive the personalized IAT effect is participants’ own evaluations toward the category. When the personally focused labels are present, participants seem to be construing and evaluating the category as a whole, as they typically define it, and remain unaffected by the construal implied by the context. The findings suggest that once participants have a clear perspective-focus in mind, the construal of the category becomes more rigid and less susceptible to multiple interpretations.

Experiment 6 sought to replicate this experiment with another social domain, that of elderly/young. On the basis of the findings from Experiment 5, it was predicted that the use of
atyical exemplars would attenuate the usual bias against the elderly on the traditional IAT but have less impact on the personalized version of the IAT.

EXPERIMENT 6

Method

Participants
Forty three introductory non-Asian psychology students (16 males, 27 females) participated in this experiment in partial fulfillment of their course requirements.

Procedure
The procedure for Experiment 5 was identical to the one employed in Experiment 5, with the exception that participants completed the young/old IAT (See Experiment 4). In the typical IAT elderly/young IAT, the exemplars were old (e.g., Oscar, Agnes) and young names (e.g., Brittany, Ryan) while the atypical stimuli condition employed positive elderly/negative young exemplars such as Barbara Walters, Mother Teresa, Eleanor Roosevelt, Frank Sinatra, and Albert Einstein for the elderly, and Erik Menendez, Timothy McVeigh, Jeffrey Dahmer, Ted Bundy, and Tonya Harding for the young (Appendix F).

Results

Original Scoring Method
Again, the IAT scores were entered into a 2 (IAT type: Personalized vs. Traditional) X 2 (Stimulus Typicality: Typical vs. Atypical) ANOVA, and again the analysis revealed a significant main effect of typicality, $F(1,39)=11.41, p=.002$, such that typical stimuli conditions resulted in greater elderly bias. However, as in the earlier experiment, this main effect was qualified by a significant Typicality X IAT type interaction, $F(1,39)=4.50, p=.034$. Consistent with Experiment 5, a significant effect of stimulus typicality on the traditional IAT was found, such that the atypical
condition showed less bias against the elderly ($M=84.05$, $SD=134.62$) than the typical condition did ($M=352.07$, $SD=166.73$), $t(19)=3.76$, $p<.001$. However, on the personalized IAT, the typicality of the stimuli did not matter; the atypical ($M=287.09$, $SD=161.66$) and typical ($M=312.08$, $SD=163.32$) conditions displayed a similar magnitude of elderly bias, $t<1$. As before, we also found a difference in the atypical exemplar condition, $t(15)=2.18$, $p=.046$, such that the personalized IAT showed a higher level of elderly bias than the traditional IAT (Figure 7).

*D-Score Algorithm*

D-score results mirrored that of the original scoring method.9 Data from participants with trials greater than 10,000 ms (0 cases) and for whom more than 10% of trials have latencies less than 400ms (3 cases) were removed. The analysis revealed a main effect of typicality of the stimulus exemplars such that typical exemplars produced greater elderly bias, $F(1,36)=12.83$, $p<.001$. However, the effect was qualified by IAT type X Stimulus exemplar, $F(1,36)=4.17$, $p=.048$. T-tests showed that traditional IAT D-scores differed significantly as a function of the stimulus exemplars presented, $t(17)=3.6$, $p=.002$, such that atypical exemplars revealed smaller elderly bias ($M=.10$, $SD=.33$) than typical stimulus exemplars ($M=.62$, $SD=.29$).

---

8 At a first glance, it might appear surprising to not have observed a significant difference between the traditional and personalized IATs when typical elderly exemplars were employed, as we had with the race IAT in Experiment 5. However, upon further reflection, this pattern of data may reveal an interesting difference between the domains of race and age. Although personal evaluations of blacks vs. whites vary greatly in terms of their coherence with cultural views, it is possible that most people’s personal associations regarding the elderly may cohere with the general societal view that “old is negative.”

9 Again, there was no evidence of differential error rates in the traditional ($M = 10\%, \ SD = 6.5$) vs. personalized IAT versions ($M = 11\%, \ SD = 7.0$), $t(41) < 1$. 
The personalized IAT, on the other hand, was not impacted by the stimulus exemplars, \( t(19)=1.2, (M=.42, SD=.18 \text{ for atypical and } M=.42, SD=.18 \text{ for typical}) \) (Figure 8).

DISCUSSION

Although Experiments 5 and 6 replicated the previously observed effects of stimulus atypicality on the traditional IAT, the personalized IAT proved less susceptible to stimulus exemplar effects. These results seem to support the conclusion that participants were temporarily redefining the categories on the traditional IAT to “fit” the category that is implied by the stimulus exemplars, presumably to accord with what they perceived the researchers intended the category labels to mean. That is, when confronted with atypical exemplars, participants may interpret the category labels “Black/ White (young/old)” and the task as concerning special cases of the category subtype X (“Oh, this is what they [the researchers] mean by Blacks/Whites (young/old).”) This explanation closely follows the logic of conversational norms similar to what was demonstrated in Experiment 2. If people are following the rules of conversational norms, especially the maxim of relevance, they should be (implicitly) asking, “Why am I being asked to categorize X and Y yet being presented with odd, atypical examples? It must be that researchers want me to think of category X as this special case.”

The personalized IAT, on the other hand, has more focused and clear evaluative labels. As suggested by Olson and Fazio (2004), the labels “I Like” and “I don’t like” focus people’s attentions on their own idiosyncratic perspectives and personal evaluations as opposed to cultural perspectives or the perspective of some specific others (e.g., extrapersonal associations). Furthermore, the personal focus may also promote consideration of the category labels in the way one typically does – one’s default representation of that category. Thus, when
adopting a personal perspective, the category label may have more restricted latitude and be less open to multiple interpretations. As a consequence of the personalized IAT’s evaluative labels, construing the category in any way other than one’s default representation may be relatively difficult.

However, this argument can be extended. It is possible that the focus on the self per se is not essential. The personalized IAT may be robust to stimulus effects not because it focuses on the self, but simply because it promotes a focused representation. To the extent that this is the case, using any specific and constrained labels on the IAT, not just the self, should make the measure robust to any stimulus exemplar effects. Chapter 5 explores this possibility.
Experiment 7 explores a probable mechanism underlying the apparent robustness of the personalized IAT to stimulus exemplar effects: its specific, constrained, and focused label. As mentioned in the previous chapter, if this is the key factor, then any IAT label that provides a specific focus should also prove relatively robust to atypical exemplar effects. This experiment capitalizes on Asians’ view of the elderly as a positively stereotyped social group (Palmore, 1975; Sung, 2001). In Asian cultures, the elderly are often viewed as sage and wise individuals who are to be respected and revered. It is not uncommon in some parts of Asia to see young individuals touch their grandparents or elderly in order to “tap” into their sage knowledge, wisdom, and secret to longevity (see Plate, 2004). Moreover, many non-Asian Americans tend to be aware of the Asian cultural reverence for the elderly (see Method section).

To test the idea that the presence of constrained and focused labels is what drives the robustness of the personalized IAT, an “Asianized IAT” with labels “Asians like/Asians don’t like” was created. It was hypothesized that focusing participants on Asians (what Asians like/don’t like), could constrain participants’ representation of the category “elderly” as positive – to the extent that the participants are aware of the Asian’s stereotype of the elderly. That is, no matter what the stimulus exemplars are, people should focus on Asians and how Asians view the category of elderly. It was expected that participants completing the “Asianized IAT” would
show less elderly bias compared to those participants completing the traditional IAT. It was also expected the Asianized IAT to prove relatively less sensitive to stimulus exemplar effects. The current experiment presented somewhat of a methodological challenge. It was not desirable to provide participants with information regarding Asians’ stereotypes of the elderly at the beginning of the experiment, nor to provide any cues that might encouraged them to generate such thoughts. Any such information would simply serve as salient cues or extrapersonal associations regarding the link between Asians and the elderly. Hence, it would not be surprising to find that participants for whom such associations were made salient would find it easier to map “Asians like” and elderly names on to the same response key, even if atypical exemplars were presented. For this reason, participants were not queried regarding their knowledge of stereotypes at the beginning of the experiment. Nor were they provided with information regarding the relevant stereotype. Instead, a screening instrument was administered after the IAT to identify participants who were aware of the cultural stereotypes.

Method

Participants

One hundred and twenty eight non-Asian introductory psychology students (47 males, 81 females) participated in this experiment in partial fulfillment of their course requirements. We excluded 23 participants for failing a screening measure regarding their awareness of the elderly stereotype among Asians (see below). The final sample consisted of 105 participants (36 males, 69 females).

Procedure

Similar to Experiment 5, participants completed an elderly/young IAT with either atypical or typical exemplars. They also were randomly assigned to either a traditional version of the IAT
(with labels “Pleasant/Unpleasant”) or an “Asianized version” (with the labels “Asians like/Asians don’t like” with the instruction “The categories that you are to distinguish are: Things Asians generally don’t like vs. Things Asians generally like”). To make the Asian perspective more relevant to the task at hand, we changed several of the evaluative stimulus items (those that are to be categorized in terms of valence). On the basis of a pilot study in which participants were asked to generate items Asians were especially likely to like or dislike, several items were identified as applicable to Asians and, hence, selected for inclusion in the stimulus lists. Thus, the final list of stimuli for “like” items included love, smartness, intelligence, happy, honor, success, rice, sushi, tea, and pleasure; the stimuli for the “don’t like” items were death, sickness, laziness, hatred, failure, stupidity, atomic bomb, disloyalty, agony and dishonor (See Appendix G).

After the IAT, participants completed a “cultural knowledge” questionnaire in which they were asked to rate how much various cultures (U.S., Asian, Hispanics, Arabs, Germans, and Swedish) place emphasis on various values (equality and respect for elderly). The crucial item of interest was how much Asians were viewed as endorsing the value of respect for the elderly, compared to other cultures (Appendix H).

Participants were selected for inclusion in the data analysis on the basis of a simple rule. A rank-ordering of the participant’s ratings of the extent to which the various cultures value the elderly had to place Asians in a position that was at least tied for first place. Thus, eligible participants needed to have rated Asians’ respect for the elderly as greater than or equal to that shown by any other culture. As noted earlier, this selection criterion led to the exclusion of 23 of the 128 participants. Thus, the overwhelming majority were able to report this aspect of Asian culture.
Results

Original Scoring Method

A 2 (IAT type: Asian vs. traditional IAT) X 2 (Typicality: Atypical and Typical stimuli) ANOVA revealed a significant main effect of stimuli typicality, $F(1,104)=28.64, p<.001$, such that atypical stimuli showed a smaller elderly bias compared to typical stimuli ($M=170.67$ vs. 296.07, respectively). However, the main effect was qualified by a significant IAT type X stimuli typicality interaction, $F(1,104)=5.01, p=.027$. Consistent with our previous study, further analysis of the interaction revealed that the traditional IAT was impacted by the stimuli, such that the atypical traditional IAT showed a reduced elderly bias ($M=155.42, SD=155.49$) compared to the typical traditional IAT ($M=338.34, SD=136.43$), $t(49)=5.76, p<.001$. On the other hand, the Asianized IAT revealed only a marginally significant difference between the atypical ($M=184.83, SD=176.64$) and typical stimuli ($M=258.50, SD=127.62$), $t(52)=1.77, p=.083$. The comparison of the IAT types also revealed a significant difference between the traditional and Asian IAT when typical exemplars had been employed as stimuli, $t(49)=2.50, p=.016$. As expected, the traditional IAT with typical stimuli showed a greater bias against the elderly than did the Asianized IAT. However, we did not find such differences on atypical Asianized and traditional IAT conditions, $t<1$ (Figure 9).

D-Score algorithm

Once again, the D-score showed similar results. Data from participants with trials greater than 10,000 ms (0 cases) and for whom more than 10 % of trials have latencies less than 300ms (1 case) was removed.\textsuperscript{10} The analysis revealed a main effect of typicality of the stimulus

\textsuperscript{10} There was no evidence of differential error rates in the traditional ($M = 7\%, SD = 5.0$) vs. Asianized IAT versions ($M = 8\%, SD = 5.4$), $t(103) = 1.2, p=.23$. 

55
exemplars such that typical exemplars produced greater elderly bias, $F(1,103)= 22.53, p<.001$, such that typical stimuli resulted in greater elderly bias. However, the effect was qualified by IAT type X Stimulus exemplar, $F(1,103)= 4.29, p=.041$. The traditional IAT D-scores differed significantly as a function of the stimulus exemplars presented, $t(48)=5.16, p<.001$, such that atypical exemplars revealed a smaller elderly bias ($M=.34, SD=.27$) than typical stimulus exemplars ($M=.71, SD=.23$). The Asianized IAT, on the other hand, was only marginally impacted by the stimulus exemplars, $t(52)=1.82, p=.077$ ($M=.41, SD=.31$ for atypical and $M=.55, SD=.28$ for typical) (Figure 10).

**DISCUSSION**

Results from Experiment 7 demonstrate that clearly defining the category labels on the IAT, by “Asianizing” it, limits the redefinition of the categories suggested by the stimulus exemplars. Compared to the traditional IAT, the Asianized IAT showed a much smaller stimulus exemplar effect. These results indicate that when the focus of the labels on the IAT is constrained, one relies on the specific representations suggested by the currently active focus and is less influenced by the atypical stimuli. Perhaps this is the mechanism for the robustness of the personalized IAT. The clear and constrained evaluative labels focus on one’s own personal likes/dislikes, and by doing so, also limit the extent to which the category labels are open to redefinition.

To the extent the IAT labels are specific and people can elicit any group or person’s representation of a category in mind, they can complete the IAT accordingly. In this case, we have shown that people can complete the IAT with Asians’ stereotypes of the elderly in mind. But as Experiment 2 demonstrated, people can complete the IAT with a society in mind, such as in the cultural IAT, where the labels are “people like/people don’t like.”
Interestingly, this is the first experiment to demonstrate that participants who were aware of the Asians’ stereotype of elderly could also complete the IAT with Asians’ representation of elderly in mind and show a much reduced elderly bias on the typical Asian IAT. Although the data did not reveal a reversal of the IAT score, a preference for the elderly, a significant reduction of elderly bias was evident relative to what was observed with the traditional IAT. Similar trends in the data are also evident in many IAT studies. For example, Foroni and Mayr (2001) did not show a preference for insects compared to flowers, but rather a reduction in preference for flowers. Similarly, Govan and Williams (2004) and Mitchell et al. (2003) showed a reduction or no differences in race bias instead of a preference for blacks. Our data are consistent with these patterns.

I can only speculate as to why the current experiment did not show a preference for elderly (rather than reduced bias). It is possible that, for most of the participants, the associations and the knowledge of positive Asian elderly stereotypes may be somewhat weak. For example, almost 66% of participants thought other cultures value respect of the elderly just as much as the Asian culture. Although consistent with the criterion for inclusion in the study, this high percentage seems to suggest that a majority of the participants in the current experiment were not fully aware of Asian’s reverence for the elderly. It is also possible that the negative stereotype of the elderly in the U.S. is so pervasive and strong (Cuddy, Norton, and Fiske, 2005) that eliciting positivity by reminding how Asians view elderly alone is not sufficient to override the strong negativity.
CHAPTER 6

GENERAL DISCUSSION

The main goals of the present experiments were three fold: 1) explore the inherent ambiguity resulting from the evaluative labels traditionally employed on the IAT; 2) investigate the consequences of the IAT’s malleability by reexamining prior research interpreting that malleability as an evidence of attitude change when no such change seemed likely; and 3) explore how the ambiguity present in the evaluative labels may also impact the construal of the category labels (ambiguity in the category labels). Another ancillary goal underlying the current research was to illustrate how that the IAT, as traditionally implemented, can be influenced by some of the same issues that are relevant to survey methodology. Finally, the research also can and should be viewed as having the goal of refining and improving the IAT as a measure of attitudes or other associations.

Experiments 1 and 2 demonstrated that the labels “pleasant/unpleasant” are ambiguous, and therefore, can be influenced by unrelated tasks. Experiment 1 increased the salience of a personal vs. normative focus via an unrelated evaluative task, and found that the traditional IAT can be susceptible to mindset priming effects. An earlier experience in an unrelated task can affect how people disambiguate the labels and whether they adopt a normative or a personal focus while completing the IAT. Experiment 2 suggested that the traditional IAT can be influenced by the implicit rules of conversational norms - the scores on the traditional IAT fluctuated as a function of the preceding IAT type. When participants
completed a personalized IAT first, traditional IAT scores increased significantly, indicating a racial bias similar in magnitude to that revealed by the culturally-oriented IAT. However, when the culturally-oriented IAT preceded the traditional IAT, the latter showed a significantly reduced racial bias, resembling the personalized IAT. It appears that the labels of the traditional IAT are ambiguous enough to allow participants to shift their meaning as the situation calls for – in this case, to avoid redundancy and comply with the rules of conversational norms. These two experiments demonstrate that the IAT is malleable. The accessibility of these differential perspectives can influence performance on the IAT, and hence, the attitude estimates that are obtained.

Experiments 3 and 4 explored the consequences of the malleability of the IAT and how it can influence the quality of the data and its interpretations. By employing the methods of Foroni and Mayr (2005), Experiment 3 showed that only the traditional IAT was impacted by salient contextual information. Participants who completed the traditional IAT showed less of a preference towards flowers immediately after they read the negative flowers/positive insects scenario. The personalized IAT did not show such fluctuation, again supporting the hypothesis that the findings of the original Foroni and Mayr (2005) study were driven by the presence of extrapersonal associations rather than the formation of “new automatic associations” to flowers and insects.

Experiment 4 replicated a prominent study by Dasgupta and Greenwald (2001); the traditional version of the IAT was impacted by a “general knowledge” task that provided participants with positive elderly/negative young exemplars or the reverse. When positive elderly/negative young exemplars were made salient, the traditional IAT showed a significantly
reduced elderly bias compared to the negative elderly/positive young exemplars condition. However, such differences were not found in the personalized version of the IAT, which is less susceptible to extrapersonal associations. The personalized IAT reflected a similar magnitude of elderly bias in the two exemplar conditions. Such a pattern of data supports the idea that the original Dasgupta and Greenwald (2001) finding may have been due to the impact of salient extrapersonal associations on the IAT and suggests that the findings reflect the malleability of the IAT rather than actual attitude change.

Experiments 5 and 6 explored whether the stimulus exemplar effect might occur due to the presence of ambiguous labels on the traditional IAT, which enable multiple construals of the category. Experiment 5 replicated Govan and Williams’s (2004) findings; when atypical black and white exemplars were presented as stimuli, participants showed a significant reduction of race bias on the traditional IAT. However, the personalized IAT was not affected by the atypicality of the stimulus exemplars. Experiment 6 showed an identical pattern of results to Experiment 5, but using atypical vs. typical elderly and young exemplars. Both of these experiments support the idea that, unlike the traditional IAT, the personalized IAT is robust to stimulus exemplar effects and category redefinition. That is, the clear and focused evaluative labels of the personalized IAT prevented participants from redefining what the category signifies and prevented shifting interpretation of the category labels.

Lastly, Experiment 7 investigated the possible mechanism that might drive the robustness of the personalized IAT to context or the typicality of stimuli. Mainly, it explored whether constrained focus labels lead to less shifts or movement in the how the categories are represented. Using an Asianized IAT, it was shown that when the labels are specific as to whose
category representation is to be used (in this case, Asians), the IAT became less susceptible to stimulus exemplars used as well as less malleable. The results illustrate the value of clearly defining the category labels. It appears that when the IAT labels are constrained by focusing on a specific mindset (personal or other specific people’s, such as Asians’), it may be difficult to construe a category in multiple ways. This experiment also makes a significant contribution because it is the first to show that people can successfully complete the IAT with “others” in mind. This idea lends support to the growing evidence that the IAT is not necessarily a measure of automatically activated personal attitudes (Olson & Fazio, 2004; Han et al., 2006; Karpinski & Hilton, 2001), but can be implemented in a focused manner and, hence, measure people’s beliefs about any group’s presumed likes or dislikes.

The findings from the last three experiments (Experiments 5, 6, & 7) also illustrate the importance of selecting prototypical exemplars on the traditional IAT to avoid biased attitude estimates of the categories. And, especially in cases where selection of prototypical exemplars is difficult or unknown, it may be crucial to use a version of the IAT that can be robust to the impact of stimulus exemplars.

On the value of specificity

The first two experiments regarding the malleability of the IAT add to a growing body of evidence suggesting that the IAT, as traditionally implemented, may measure different constructs as a function of how respondents interpret the labels. These results suggest that, depending on the previous task or the instructions, participants may actually complete the IAT in very different ways than what the researcher had intended. It has already been shown that the IAT can be easily influenced by extrapersonal associations (Han et al., 2006; Olson & Fazio,
2004), cultural associations (Karpinski & Hilton, 2001), and the representativeness of stimulus exemplars (Govan & Williams, 2004; Michelle et al., 2003; Steffens & Plewe, 2001). The current findings add to this literature by demonstrating that the IAT is sensitive to unrelated contexts, mindset priming, and implicit conversational norms. It seems that influences from a variety of factors may interfere with the capability of the traditional IAT to accurately gauge participants’ attitudes. On the other hand, the present results suggest that appropriately chosen labels can improve the IAT by focusing it more specifically on the association of interest.

The present findings also indicate that “other” oriented versions of the IAT can be implemented to assess people’s associations of groups. For example, culturally-oriented versions of the IAT could potentially be used to measure normative or cultural associations. In Experiments 1 and 2, IAT scores were significantly higher (and, hence indicative of more racial prejudice) when people were primed to think in a normative mindset or when the evaluative labels were changed to (“People like/don’t like”). In Experiment 7, participants in the “Asianized” IAT revealed a significantly smaller elderly bias. These data suggests that individuals can complete the IAT with the perspective of “others” in mind.

The value of employing IATs with more focused labels is illustrated further by recent research in which participants completed three different versions of an IAT in counterbalanced order one week apart. Peach, Yoshida, Zanna, & Spencer (2006) administered traditional, personalized, and cultural IATs (see also Yoshida, Peach, Spencer, & Zanna, 2006) across a seven-day span. They found scores on the personalized and cultural IATs to each predict unique variance in responses to a traditional IAT. That is, participants’ responses on the traditional IAT appeared to be uniquely influenced by both their personal and cultural associations. Again, the
results emphasize the importance and the benefits of employing a well-labeled, focused IAT – one that reduces ambiguity and the possibility of multiple interpretations by using clearly defined category labels.

*Malleability of attitudes?*

Experiments 3, 4, 5, and 6 demonstrated that the traditional IAT’s susceptibility to contextual effects can pose serious problems for data interpretation. That is, responses on the traditional version of the IAT may suggest the presence of attitude change when no such attitude change has actually occurred. This set of experiments reveals that some context effects demonstrated on the traditional IAT may stem from its sensitivity to salient extrapersonal associations due to unfocused category labels, instead of representing a change in the attitude representation itself. The current results call into question the typical interpretation of past findings as evidence for the malleability of automatically-activated attitudes. It is possible that the findings simply reflect the malleability of the measure itself and not any underlying change in attitudes.

Specifically, Experiments 3 and 4 illustrate and emphasize that what might seemingly be attitude change on the traditional version of the IAT may not actually be a change, but a mere reflection of the malleability in the measure itself. The goal was not to argue that automatically-activated attitudes are incapable of change. Obviously, attitudes can be changed. Decades of research on persuasion, dissonance, and self-perception processes attest to that fact. For example, evaluative conditioning has been shown to produce actual change in one’s representations and attitudes (Kawakami et al., 2007; Olson & Fazio, 2006). Also, in cases like a semester-long diversity class, in which students have ample time to learn about counter-
stereotypes and automatize their non-prejudicial responses (Rudman, Ashmore, & Gary, 2001), attitude change at the automatic level could result. Indeed, recent experimental evidence demonstrates that interracial dormitory roommate relationships promote favorable changes in White students’ automatically-activated racial attitudes (Shook & Fazio, 2008).

However, given what is known about the stability, accessibility, and the resistance of automatically-activated attitudes to change (see Fazio, 1995), one must question how reasonable is it to assume that automatically-activated attitudes are subject to change with every situation and context. For example, how reasonable is to assume that watching two minutes of a counter-stereotypical movie or seeing several images of atypical exemplars changes one’s automatically activated attitudes towards Blacks? If this is true, then automatically activated attitudes should change in every situation – whenever the interaction setting has changed, someone voices a different view or an opinion, or every time one watches TV or read the news. Just as decades of research indicate that attitudes can be changed, they also point to how difficult it can be to accomplish such change. The same message is evident with respect to the success of clinical interventions. Maladaptive fears and dysfunctional attitudes can be changed effectively with appropriate treatment intervention, but success can be very difficult to achieve and is certainly not guaranteed.

Obscuring the presence of attitude change

The current set of experiments focused on how the traditional version of the IAT may reflect attitude change when no such attitude change has actually occurred. However, it is also possible that the traditional IAT can also obscure real attitude change when it does occur. In fact, using a persuasion paradigm, Han, Czellar, Olson, and Fazio (2009) have demonstrated that
the traditional IAT can sometimes fail to capture the full extent of attitude change in settings where extrapersonal associations are prevalent. When a strong positive message regarding luxury brands was presented, it induced attitude change and this change was captured by both a traditional and a personalized IAT. However, if the positive message was followed by a very weak counterpoint, a brief message that was unsupported by argument or data and was explicitly labelled as unrepresentative of the typical interviewee’s thoughts, the traditional IAT failed to reveal any evidence of attitude change. On the other hand, the personalized IAT exhibited a measurement outcome still reflecting attitude change regardless of the presence of the weak counter message. Apparently, the weak counterpoint served as an extrapersonal association that interfered with the ability of the traditional IAT to reveal the persuasive impact of the strong positive message.

On the parallels between the IAT and survey research

The present findings demonstrate that, contrary to popular belief, the IAT, although an implicit measure, can be influenced by some of the same issues that afflict survey questions. Experiment 1 showed that IAT can be influenced by conversational norms. Experiment 2 illustrated that, just as in survey measures, the mindsets induced by an earlier task can influence how people interpret and complete the IAT. As true in explicit survey measures, the accessibility of these differential perspectives can influence performance, and hence, the attitude estimates that are obtained.

Moreover, results from Experiments 5, 6, and 7 draw an interesting parallel between the wording biases that are present on explicit measures and the stimulus exemplar selection on the IAT. For example, when Time magazine asked six hundred Americans whether “The U.S.
government spends too much money on assistance to the poor” or “The U.S. government spends too much money on welfare,” there was a shocking disparity in people’s answers depending on the phrasing of the question. While “assistance to the poor” is the definition of welfare and hence rendering the two questions equivalent, the survey revealed that only 23% of Americans agreed to the “assistance to the poor” version of the question and a full 53% agreed to the latter version of the question regarding welfare (Time, 1994, as cited in Brehm et al., 2006). Although seemingly illogical, these results are not surprising. Consider the prototypical exemplars that are likely to come to mind while thinking of the poor who need assistance (perhaps, the elderly or the disabled?) vs. those on welfare (perhaps, teenage mothers or lazy individuals?). Indeed, years of research in survey methodology have shown that subtle changes in wording can significantly impact how people construe questions and impact both the quantity and the quality of the answers (see Schwarz, 1999). The results from the last three experiments seem to point to a similar finding such that selection of stimulus exemplars matter on the IAT – that it impacts the quality of the attitude estimates (via category redefinition).

Thus, the current research demonstrates the importance of considering the communicative and interpretation issues that are relevant to designing clear and effective explicit measures of attitudes when implementing an IAT.

Conclusion

In conclusion, researchers must carefully evaluate their implementation of the IAT. The current experiments show that information presented before and during the IAT, even in the context of an ostensibly unrelated study, can influence how people interpret and complete the
IAT. Consequently, one may see “attitude change” where it does not exist, or fail to see change when it did occur.

However, it should be emphasized that when properly implemented, the IAT can be a versatile and an effective tool. The personalized IAT is robust to the influence of extrapersonal associations, momentary irrelevant contextual information, and stimulus exemplar effects, and therefore, can provide better estimates of attitudes. Moreover, depending on the experimental purpose, researchers can employ more specific, focused category labels to assess other important associations -- cultural ones, or beliefs about what various social groups (Asians, Democrats, etc.) value.
LIST OF REFERENCES


70


*Time*, (1994, June 27). P. 26


APPENDIX A

FIGURES
Figure 1: Original IAT scores as a function of IAT type and scenario (Experiment 3)
Figure 2: D-scores as a function of IAT type and scenario (Experiment 3)
Figure 3: Original IAT scores as a function of IAT type and exemplars presented (Experiment 4).
Figure 4: D-scores as a function of IAT type and exemplars presented (Experiment 4).
Figure 5: Original Scores on the race IAT as a function of IAT type and stimulus exemplars presented (Experiment 5).
Figure 6: D-scores on the race IAT as a function of IAT type and stimulus exemplars presented (Experiment 5).
Figure 7: Original scores on the Age IAT as a function of IAT type and stimulus exemplars presented (Experiment 6).
Figure 8: D-scores on the Age IAT as a function of IAT type and stimulus exemplars presented (Experiment 6).
Figure 9: Original scores on the Age IAT as a function of IAT type and stimulus exemplars presented (Experiment 7).
Figure 10: D-scores on the Age IAT as a function of IAT type and stimulus exemplars presented (Experiment 7).
APPENDIX B

EXPERIMENT 1: PERSPECTIVE INSTRUCTIONS
Normative focus Instruction:

In this first task, we are interested in how people, in general, perceive other people's preferences. That is, we are interested in what you think most other people's (or culture's) likes and dislikes are on a variety of social issues, ideas, and everyday items. The items are randomly generated by the computer, so you will see a random mix of various items.

Remember, we are NOT interested in your personal likes and dislikes, but rather, we are interested in what you think the society or the majority of people in general like and dislike.

Please try not to let your personal likes and dislikes influence your ratings of these items. You responses will be confidential.

Personal focus instruction:

In this first task, we are interested in YOUR OWN personal preferences. That is, we are interested in your personal likes and dislikes are on a variety of social issues, ideas, and everyday items. The items are randomly generated by the computer, so you will see a random mix of various items.

Remember, we are interested in YOUR OWN personal likes and dislikes. Please be as honest as you can. All answers will be confidential.

Please try not to let others or culture's likes and dislikes influence your ratings of these items. You responses will be confidential.
EXPERIMENT 3: SCREENING QUESTIONNAIRE

INSTRUCTION:
Did you read and understand the story?
If you have accidently hit continue, you can go back and finish reading. If you are ready to answer some question regarding the story, you can hit continue to get started.

1) How easy was it for you to imagine the scenario?
   (1) Very easy to (7) Very difficult

2) How do you think you will feel in this world where you've survived a nuclear war?
   (1) Very bad to (7) Very Good

3) Now think carefully about the potential food sources for the animals. Answer whether following food sources would be a good source of food and water.
   a) Water from a well
   b) Earth worm
   c) Roses
   d) Beetles
   e) Fish
   f) Mosquito
   g) Chicken
   h) Purified water
   i) Maggots
   j) Seaweed
   k) Daisies
   l) Lettuce
   m) Apples
   n) Insects
   o) Flowers
   (1) Definitely NOT good to eat/drink to (7) Definitely good to eat/drink

4) How confident do you feel that you can survive in this world?
   (1) Not at all confident to (7) Extremely confident

5) What do you think your quality of life will be in this new world?
   (1) Very bad to (7) Very good
APPENDIX D

EXPERIMENT 4: GENERAL KNOWLEDGES TASK STIMULI
EXPERIMENT 4: GENERAL KNOWLEDGES TASK STIMULI

POSITIVE ELDERLY EXEMPLARS

Dr. Benjamin Spock

Famous golf champion
Former pediatrician famous for his books on childrearing

Albert Einstein

Popular jazz musician
World famous physicist who won the Nobel Prize
Eleanor Roosevelt

Beloved former First Lady of the United States

Award winning artist

Frank Sinatra

Famous American singer

Famous American painter
<table>
<thead>
<tr>
<th>Jessica Tandy</th>
<th>Mother Teresa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Famous singer</td>
<td>Famous novelist</td>
</tr>
<tr>
<td>Actress who won the Oscar for <em>Driving Miss Daisy</em></td>
<td>Was a true champion of the poor and destitute</td>
</tr>
</tbody>
</table>
Barbara Walters
Co-anchor of a popular news show on ABC.  
Award winning writer.

Phil Donahue
Former tennis champion  
Popular talk show host
Walter Cronkite

Former news anchor for CBS

Famous actor

Willie Nelson

A popular DJ on the radio

Famous country music singer
NEGATIVE ELDERLY EXEMPLARS

Bob Packwood

Former senator who resigned after being accused of sexual assault and harassment.  Embezzled millions of taxpayers' money

James McDougal

Convicted of fraud and conspiracy in the recent Whitewater development case. Recently convicted for a series of bank robberies.
Jim Bakker

Member of an international terrorist organization
Television evangelist convicted of embezzling millions of dollars from his church

Leona Helmsley

Infamous real estate tycoon convicted of felony and mail fraud
School teacher who stole thousands from the teacher's retirement fund
Marge Schott

Daycare worker accused of neglecting children at her daycare center
Owner of the Cincinnati Reds team infamous for her racist and anti-Semitic comments

Marie Noe

Housewife accused of suffocating and killing 8 of her 10 children
Accused of insurance fraud
Marshall Applewhite

Convicted of dealing drugs
Leader of the Heaven’s Gate cult who convinced his group to commit mass suicide.

Sam Bowers

Bombed the World Trade Center in New York City.
Former Ku Klux Klan head convicted for the brutal murder of civil rights worker in Mississippi.
Ted Kaczynski

The Unabomber who injured and killed using letter bombs.

Convicted pedophile.

Tammy Faye Bakker

Ex-wife of TV evangelist who helped him embezzle money from his church

Convicted pedophile
POSITIVE YOUNG EXEMPLARS

Jodie Foster
Oscar-winning actress
Famous novelist

Casey Martin
Up-and-coming professional golfer who has succeeded despite his ability
Famous actor
Leonardo DiCaprio

A popular D.J. on the radio
Actor who played a leading role in the *Aviator*.

Matt Damon

Popular actor who starred in *Good Will Hunting*
Famous golf champion
Princess Diana

Award winning artist  Beloved princess who died in a car accident

Prince William

Heir to the British throne  Up-and-coming gold champion
Tobi MacGuire

Former Tennis champion

Popular actor

Todd Eldridge

Famous American painter

One of the finest American ice-skaters
<table>
<thead>
<tr>
<th>Sarah McLachlan</th>
<th>Steffi Graf</th>
</tr>
</thead>
<tbody>
<tr>
<td>Award-winning writer.</td>
<td>Famous singer.</td>
</tr>
<tr>
<td>Famous singer and organizer of the Lilith Fair.</td>
<td>Famous singer.</td>
</tr>
<tr>
<td>One of the best female tennis players.</td>
<td></td>
</tr>
</tbody>
</table>
NEGATIVE YOUNG EXEMPLARS

Andrew Cunanan

Embezzled millions of taxpayers' money. Murderer who killed fashion designer Gianni Versace.

Diane Zamora

Army cadet who planned the brutal killing of her boyfriend's other lover. School teacher who stole thousands from the teachers' retirement fund.
Erik Menendez
Leader of antigovernment militia.
Serving life in prison for killing his own parents with a shotgun.

Jeffrey Dahmer
Bombarded the World Trade Center in New York City
Serial killer who cannibalized his victims
Louise Woodward

Convicted of dealing drugs.

Nanny who was convicted of murder for killing the child in her care.

Lyle Menendez

Brutally killed his parents with a shot gun.

Convicted for a series of bank robberies.
Susan Smith
Young mother in North Carolina who drowned both her children.
Convicted pedophile.

Ted Bundy
Serial rapist and killer from the Northwest
Member of an international terrorist organization
Mafia member. Bombed the federal building in Oklahoma City, killing hundreds.

Daycare worker accused of neglecting children at her daycare center. Olympic skater accused of plotting to injure her competitor in order to secure the Olympic gold for herself.
APPENDIX E

EXPERIMENT 5: RACE IAT STIMULI
## EXPERIMENT 5: RACE IAT STIMULI

<table>
<thead>
<tr>
<th>Typical</th>
<th>Atypical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black</td>
<td>White</td>
</tr>
<tr>
<td>&quot;THEO&quot;</td>
<td>&quot;CHIP&quot;</td>
</tr>
<tr>
<td>&quot;LEROY&quot;</td>
<td>&quot;JOSH&quot;</td>
</tr>
<tr>
<td>&quot;TYRONE&quot;</td>
<td>&quot;TODD&quot;</td>
</tr>
<tr>
<td>&quot;LAKISHA&quot;</td>
<td>&quot;AMBER&quot;</td>
</tr>
<tr>
<td>&quot;EBONY&quot;</td>
<td>&quot;BETSY&quot;</td>
</tr>
</tbody>
</table>
APPENDIX F

EXPERIMENT 6 & 7: AGE IAT STIMULI
## EXPERIMENT 6 & 7: AGE IAT STIMULI

<table>
<thead>
<tr>
<th>Typical</th>
<th>Atypical</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Old</strong></td>
<td><strong>Young</strong></td>
</tr>
<tr>
<td>&quot;ETHEL&quot;</td>
<td>&quot;RYAN&quot;</td>
</tr>
<tr>
<td>&quot;AGATHA&quot;</td>
<td>&quot;BRITTANY&quot;</td>
</tr>
<tr>
<td>&quot;BERNICE&quot;</td>
<td>&quot;BRANDON&quot;</td>
</tr>
<tr>
<td>&quot;LUCILLE&quot;</td>
<td>&quot;CHRISTINE&quot;</td>
</tr>
<tr>
<td>&quot;ALFRED&quot;</td>
<td>&quot;JULIE&quot;</td>
</tr>
</tbody>
</table>
APPENDIX G

EXPERIMENT 7: ASIANIZED IAT STIMULI
**EXPERIMENT 7: ASIANIZED IAT STIMULI**

<table>
<thead>
<tr>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian</td>
<td>Asian</td>
</tr>
<tr>
<td>“Love”</td>
<td>“Death”</td>
</tr>
<tr>
<td>“Smartness”</td>
<td>“Sickness”</td>
</tr>
<tr>
<td>“Rice”*</td>
<td>“Laziness”</td>
</tr>
<tr>
<td>“Intelligence”</td>
<td>“Hatred”</td>
</tr>
<tr>
<td>“Happy”</td>
<td>“Failure”</td>
</tr>
<tr>
<td>“Honor”</td>
<td>“Stupidity”</td>
</tr>
<tr>
<td>“Sushi”*</td>
<td>“Atomic Bomb”</td>
</tr>
<tr>
<td>“Success”</td>
<td>“Disloyalty”</td>
</tr>
<tr>
<td>“Tea”*</td>
<td>“Pleasure”</td>
</tr>
<tr>
<td></td>
<td>“Agony”</td>
</tr>
<tr>
<td></td>
<td>“Dishonor”</td>
</tr>
</tbody>
</table>
APPENDIX H

EXPERIMENT 7: SCREENING MEASURE
EXPERIMENT 7: SCREENING MEASURE

Brief Cultural Knowledge Inventory

Please read the instructions carefully and answer the following questions.

Please rate the extent to which each of the following cultures or nations are viewed as abiding by the listed value:

*All questions are rated from (1) Not at All to (7) Very Much*

Americans and the value “Equality for all”
Are Americans viewed as following the value "Equality for all?"

Germans and the value “Equality for all”
Are Germans viewed as following the value "Equality for all?"

Arabs and the value “Equality for all”
Are Arabs viewed as following the value "Equality for all?"

Swedish and the value “Equality for all”
Are Swedish viewed as following the value "Equality for all?"

Asians and the value “Equality for all”
Are Asians viewed as following the value "Equality for all?"

Hispanics and the value “Equality for all”
Are Hispanic viewed as following the value "Equality for all?"

Americans and the value “Respect for the elderly”
Are Americans viewed as following the value “Respect for the elderly?”

Germans and the value “Respect for the elderly”
Are Germans viewed as following the value “Respect for the elderly?”

Arabs and the value “Respect for the elderly”
Are Arabs viewed as following the value “Respect for the elderly?”

Swedish and the value “Respect for the elderly”
Are Swedish viewed as following the value “Respect for the elderly?”

Asians and the value “Respect for the elderly”
Are Asians viewed as following the value “Respect for the elderly?”
Hispanics and the value “Respect for the elderly”
Are Hispanics viewed as following the value “Respect for the elderly?”