Malleability of Attitudes or Malleability of the IAT?

H. Anna Han

St. Mary's College of Maryland

Sandor Czellar

HEC Paris School of Management

Michael A. Olson

University of Tennessee

Russell H. Fazio

Ohio State University

Please address correspondence to:

Russell H. Fazio
Department of Psychology
Ohio State University
1835 Neil Avenue
Columbus, OH 43210-1287
Phone: 614-688-5408
Fax: 614-688-5414
E-mail: fazio.11@osu.edu

Abstract

In the current set of experiments, we establish, and explore the consequences of, the imprecision that characterizes the attribute response labels typically employed in the Implicit Association Test (IAT). In Experiment 1, we demonstrate the malleability of the IAT, as conventionally implemented. IAT scores are shown to be influenced by perspective mindsets induced by an unrelated preceding task. Then, we explore how the malleability of the IAT can lead to the inference that attitude change has occurred even when there is very good reason to believe it has not (Experiment 2), and conversely, how it can obscure the detection of attitude change when such change is indeed likely to have occurred (Experiment 3). We provide conceptual explanations for these discrepancies and suggest methodological improvements to enhance the specificity of IAT measures.

Keywords: Implicit association test, attitudes, attitude change, extrapersonal associations

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 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 we had been posing. 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, the outcome of such efforts to disambiguate have the potential to produce less than meaningful data and can lead the researcher to draw inappropriate inferences. Hence, 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 extent to which these attitude measurement tools are influenced by efforts on the part of the participants to disambiguate the nature of the task before them has not yet been as carefully examined. 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; Fazio & Olson, 2003; Govan & Williams, 2004; Karpinski & Hilton, 2001). However, more research is needed so as to inform the scientific community about the appropriateness of inferences that might be drawn from IAT findings. Most importantly, the

implications of any disambiguating efforts by the participants for inferences regarding the

measurement procedure 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? We argue that the imprecise response labels typically employed in the IAT

allow for plasticity that can lead to erroneous inferences regarding the malleability of attitudes.

Specifically, the present research demonstrates that the IAT as conventionally

implemented is open to multiple interpretations, and therefore, can provide contextually

malleable measurement outcomes (Experiment 1). We then explore important consequences of

the IAT's malleability for the inferences that can be drawn regarding the presence or absence of

attitude change following a social influence attempt. Whereas past research often has interpreted

change in IAT scores as evidence of the malleability of attitudes, we argue that the malleability

of the IAT can sometimes lead to the mistaken inference that attitude change has occurred even

when there is very good reason to believe it has not (Experiments 2). However, we also show

that the traditional implementation of the IAT can sometimes obscure the detection of attitude

change when such change is indeed likely to have occurred (Experiment 3). We provide a

conceptual explanation for these discrepancies and suggest methodological improvements to

focus IAT measures more precisely.

The Sensitivity of the IAT to Extrapersonal Associations

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., Blair, 2002; Greenwald et al., 1998; Nosek, Greenwald, &

Banaji, 2005), health behaviors (e.g., Sherman et al., 2002), and self-esteem (e.g., Greenwald & Farnham, 2000; Jordan et al., 2003). Despite this wide usage, evidence indicates that the IAT as typically implemented may not be as robust to momentary, irrelevant contextual considerations as once believed (e.g., Karpinski & Hilton, 2001; Olson & Fazio, 2004).

IAT performance depends on one's ability to cope with a response mapping problem. During the critical blocks, any given response key has two meanings: one key is to be pressed, for example, whenever an exemplar of Category X or something "good" is presented, and the other key whenever an exemplar of Category Y or something "bad" is presented. To perform well, one needs to remember the dual meaning of any given key. The easier it is to associate X with good and/or Y with bad, the faster one can respond. The assumption is that the ease of remembering and working with a given response mapping is a function of individuals' attitudes toward X and Y. But, why would attitudes be the only form of information to affect the efficiency with which the dual meaning of a response key can be accommodated?

Olson and Fazio (2004) argued that IAT participants may be influenced by *any* information that can facilitate their management of the response mapping, including information that is inconsistent with their attitudes. They maintained that category labels such as the commonly employed "good/bad" or "pleasant/unpleasant" are open to multiple interpretations and, hence, allow for the activation of any information that might be useful for managing the demands of the categorization task. Most importantly, the perspective intended by such category labels is unspecified: "good"/"bad" or "pleasant"/"unpleasant" 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 traditional versions of the IAT can be influenced by extrapersonal associations (Han, Olson, & Fazio, 2006; 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. For example, the knowledge that my neighbor was a huge supporter of presidential candidate McCain did not impact my support for Obama. Nevertheless, such knowledge can facilitate a respondent's efforts to accommodate the dual meaning of the response keys during the IAT. Thinking about Aunt Mary or my neighbor's preferences, the researcher's presumed intent, or the cultural perspective may promote faster responding for a given response mapping.

A much more thorough and lengthy consideration of extrapersonal associations can be found in Olson, Fazio, & Han (2009). Responding to various questions about how "extrapersonal" might be conceptualized (Gawronski, Peters, & LeBel, 2008; Nosek & Hansen, 2008a), Olson et al. (2009) address issues concerning the conceptual and empirical distinction between personal and extrapersonal associations, including definitional matters, the development of extrapersonal knowledge, and the ensuing advantage that typically characterizes personal associations in terms of their likelihood of activation. In the present context, we wish to emphasize only three related points. The first is definitional, and concerns both attitudes and extrapersonal associations. Consistent with such classic definitions as those of Allport (1935) and Thurstone (1928), we view attitudes as inherently personal reactions to an attitude object. More specifically, we consider attitudes to be summary evaluations that have the potential to be automatically activated upon encountering an attitude object and that then guide construal of the object in the immediate situation and ultimately approach/avoidance decisions (see Fazio, 2007).

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 a severe allergy to peanuts 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 food items and menus for the presence of peanuts. A sibling's liking for peanut butter played no role in the development of this negative attitude, nor did other knowledge related to peanuts, such as the cuteness of the Mr. Peanut character (a peanut sporting a top hat and cane) that serves as the mascot for Planters Peanuts. Information of this sort is certainly available in memory, but it did not contribute to the person's negative attitude toward peanuts.

The second point to be highlighted follows directly from this conceptualization. Extrapersonal associations should not be regarded as corresponding directly to cultural associations (Nosek & Hansen, 2008a). 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 person. Most lovers of anchovies recognize that their positivity places them in the minority, but this does not stop them from anticipating delight at the prospect of requesting that their caesar salads be served with anchovies. For such individuals, knowledge of the more culturally predominant negativity assumes the status of an extrapersonal association.

Finally, we wish to emphasize that our focus on the concept of extrapersonal associations is not intended to deny in any way the essential social psychological principle that social influence is pervasive in nature. Individuals develop attitudes not only from their personal experiences when interacting with the attitude object, but also through the social transmission of information about the attitude object, the norms that characterize their reference groups, and even their mere awareness of others' attitudes. Social forces have long been known to produce not only public compliance, but also private acceptance, at least under certain specifiable conditions (e.g., Deutsch & Gerard, 1955). Informational social influence is undoubtedly a powerful force, as is direct persuasion. Yet, it also seems evident that at least sometimes the attitudes of at least some people are not affected by their knowledge of others' views or the communicated attributes of the attitude object. Social influence phenomena are not universal. It is when they fail to impact a given individual's attitudes that such information meets our definition of an extrapersonal association. To borrow an example from Olson et al. (2009), awareness of one's mother's obvious affection for floral-patterned wallpaper, bedspreads, and curtains represents an extrapersonal association to the extent that one's own reaction upon seeing a floral pattern is a sense of distaste. Even if the distaste had been facilitated by a desire to be different from mother, knowledge of her positivity is distinct from one's experienced negativity.

Given the idiosyncratic nature of attitudes and the varying bases from which individuals' attitudes might develop, it is difficult to identify a given item of information, in any a priori manner, as personal or extrapersonal for any given individual (see Olson et al., 2009, for further discussion of this issue). What is extrapersonal to one person may form the very essence of another's attitude. Hence, the most convincing evidence that extrapersonal associations can indeed influence performance on the IAT stems from research that has experimentally created

both attitudes towards novel stimuli and extrapersonal associations in the laboratory. With such

experimental control, Han et al. (2006) were able to engineer a situation in which it was highly

implausible that a particular piece of information would impact individuals' attitudes. Hence,

the information could confidently be considered extrapersonal in nature.

In these experiments, participants first formed attitudes towards two novel Pokémon

characters, which were characterized by attributes that made one 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, thus establishing a

very firm preference. Participants then were exposed to a video of two 10-years-old boys

commenting on 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, clearly erroneous, and easily dismissed preference for the objectively inferior

character (inconsistent or extrapersonal condition).

Across two experiments, results revealed that participants' IAT scores differed as a

function of the comments to which they had been exposed. When exposed to inconsistent

commentary (i.e., when the boys' opinions contrasted with objective reality and participants'

own attitudes), participants' IAT scores suggested a significantly reduced preference for the

objectively superior Pokémon character compared to when the boys agreed. This change in IAT

scores occurred despite the fact that participants in this extrapersonal condition both 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 video. Both measures revealed only a preference for the objectively superior Pokémon. Apparently, even though their own preferences were unaffected, participants in the extrapersonal condition were able to recruit the boys' opinion to solve the mapping problem posed by assignment of the inferior Pokémon character to the same key as "pleasant" and the superior Pokémon to the same key as "unpleasant." By momentarily adopting the boys' perspective, it proved easier to associate the inferior Pokémon with the key signifying positivity.

Goals and Overview of Studies

Since we argue that the valence labels of the traditional IAT are potentially ambiguous, we first sought to demonstrate the openness of these labels to multiple interpretations and the resulting malleability of the traditional IAT. In Experiment 1, we explored how an unrelated preceding task can change participants' perspective regarding the traditional IAT category labels. Our findings suggest that, depending on the perspective primed by the preceding task (self or other people), participants' IAT scores may differentially reflect more or less personal evaluations of the target categories. Then in Experiment 2, we explore the consequences of the IAT's susceptibility to these effects by replicating and shedding light on a study purportedly concerning the malleability of automatically-activated attitudes. Specifically, we explore whether the results outlined in this past experiment might simply reflect the malleability of the IAT measure itself rather than changes in participants' attitudes. Finally, in Experiment 3, we sought to show the converse of Experiment 2 and demonstrate that the malleability of the traditional IAT may sometimes mask the detection of attitude change that is actually likely to have occurred.

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 generally 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 that the upscale brand Perrier was rated 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 allow participants to shift their meaning as a function of the judgmental perspective 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 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 earlier, although now irrelevant, questionnaire labels and, hence, lead to different IAT scores. Consistent with arguments regarding the prevalence of negative cultural portrayals of Blacks (Goff, Williams, Eberhardt, &

Jackson, 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").

Method

   Participants. Fifty-three introductory psychology students (11 males, 42 females) 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).

   Procedure. 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 normative focus condition, the scale points were labeled "people like very much/people don't like very much," and participants were asked to ignore their own likes and dislikes and answer with what they thought people in general like/dislike. Conversely, in the personal focus condition, the labels were "I like very much/I don't like very much," and participants were asked to ignore what people in general like/dislike and focus on their own personal preferences.

   In the ostensibly unrelated second study, participants completed a traditional IAT

assessing racial attitudes, with the usual category labels of pleasant/unpleasant. 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.

Results and Discussion

The data from the IAT were aggregated in accord with the procedure originally

established by Greenwald et al. (1998).[1] The response latencies for the first two trials of each

block were dropped and latencies under 300 and over 3,000 ms were recoded 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).

These findings suggest that the IAT, as traditionally implemented, can easily be affected by the preceding context. It appears that the traditional IAT labels do allow for multiple interpretations. 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 different 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 or neutral concepts (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 not only to the importance of perspective, but also to the value of personalizing the IAT – a matter that is central to the argument we develop.

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 most important change with respect to the traditional IAT involves modifying 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. The perspective intended by the labels "pleasant/unpleasant" or "good/bad" is unspecified, 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. To maintain such a personal focus, error feedback is removed from the personalized IAT. Error feedback encourages a normative focus by indicating that a correct answer, independent of the person's own liking or disliking of the target objects, exists.

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. For example, Olson and Fazio's (2004) Experiments 1-2 found less evidence of racial bias on the personalized IAT than on the traditional IAT, suggesting that negative extrapersonal associations about Black people may enhance the bias observed on the traditional 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.

By personalizing the IAT, the aim was to refine, focus, and improve the measure. Extant research indeed shows evidence of improved validity. Personalized IATs have yielded stronger relations with reports of past behavior, behavioral intentions, 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). The personalized IAT also appears to more accurately assesses the in-group attitudes of stigmatized group members (e.g., African-Americans and homosexual men; Olson, Crawford, & Devlin, 2009). Although traditional IAT findings suggest such groups harbor negative views about themselves, explicit measures, priming measures, and the personalized IAT show that stigmatized groups prefer their in-groups just as other groups do. In all these cases, the dissociation between the traditional IAT and other measures appears to stem from the widespread availability of knowledge that is extrapersonal in nature.

This is not to imply that the traditional version of the IAT bears no relation to attitudes. We believe that the traditional version can measure one's personal attitudes, but the extent to which it does is inversely related to the extent to which extrapersonal associations are salient. For respondents for whom the IAT does not evoke any attitudinally-incongruent extrapersonal associations, the traditional version of the IAT should be indicative of attitudes. For example, Han et al. (2006) showed that when participants were exposed to comments that were consistent with the attitudes they had developed earlier, the traditional IAT scores successfully reflected their liking for the preferred Pokemon character. However, to the extent that some respondents recruit counterattitudinal extrapersonal associations to help disambiguate the IAT, we cannot be sure that attitudes are being reflected by the IAT. When the IAT is personalized, the labels are clearly focused on personal likes and dislikes, thus disambiguating the labels in the same way for all participants. In this way, the measure bypasses the potential influence of extrapersonal associations.[2]

Implications for IAT Findings Suggestive of Attitude Change

Experiment 1 suggests that the pliability of the labels in the traditional IAT can lead participants to interpret the task from different perspectives. The labels "pleasant/unpleasant" were influenced by an unrelated task that varied the information on which participants focused—either a personal focus or a normative focus. These results suggest that, depending on the previous task or instructions, participants may actually complete the IAT in ways different than those intended by the researcher. If the traditional IAT can be affected by momentarily salient contexts that have no bearing on the attitude object in question, one must carefully consider its role when interpreting outcome data. In particular, this malleability of the IAT may lead researchers to mistaken interpretations – for example, regarding inferences about attitude change.

For instance, 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 researchers 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, we question 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?

Yet, perhaps due to assumptions regarding implicit measures, such conclusions of change in automatic attitudes have been readily drawn on the basis of changes in IAT performance. For example, in one of the most frequently-cited illustrations of the presumed malleability of automatically activated attitudes, Dasgupta and Greenwald (2001) exposed participants to a set of famous well-liked African Americans and infamous 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). An IAT administered immediately thereafter revealed reduced racial bias in the former condition relative to the latter, prompting the conclusion that "exposure to admired and disliked group members produces substantial change in automatic intergroup evaluations" (Dasgupta & Greenwald, 2001, p. 808).

Although the above studies, as well as others (e.g., Blair et al., 2001; Carpenter & Banaji, 2001; Lowery, Harden, & Sinclair, 2001; Mitchell, Nosek, & Banaji, 2003; Wittenbrink, Judd, & Park, 2001), suggest the possibility that automatically-activated attitudes could be malleable, it is important to recognize that inferences about change in a memory representation 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 Experiment 1 prompted any change in participants' attitudes toward African-Americans. It seems unlikely that the consideration of personal versus normative likes and dislikes in a preceding task could affect a participant's evaluative representation of African-Americans, especially given that none of the items in the initial judgmental task related to race. Yet, there was an effect on the IAT. Thus,

Experiment 1 demonstrates malleability of the IAT in the absence of any reason to expect that attitudes themselves were changed in any way. Experiment 2 pursues a similar aim. We sought to examine whether IAT performance would prove sensitive to a brief contextual manipulation that had no plausible bearing on the relevant attitudinal representations.

Experiment 2

In Experiment 2, we explored whether the manipulations employed by Foroni and Mayr (2005), instead of influencing attitudes, may have provided participants with highly salient extrapersonal associations that then could be used to manage the response mapping problem posed by the IAT. 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, we closely mirrored Foroni and Mayr's (2005) paradigm with two crucial modifications. First, we presented the scenario and the IAT 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, our 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, we included the personalized IAT as a crucial comparison condition. If automatically activated attitudes were readily influenced by exposure to the scenario, we should see the changes 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).

Procedure. The experiment was conducted in laboratory conditions identical to the previous experiment. 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).[3] In the scenario, 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). After reading this scenario, participants completed a questionnaire regarding what foods they thought were

safe to eat in this imagined world and what their quality of life would be like. They 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

we had 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 us

to gauge effects of the scenario relative to baselines for each IAT version.

Results

The IAT data were prepared as in Experiment 1. The compatible blocks were identified

as flowers/ + vs. insects/ – and incompatible block as flowers/ – vs. insects/ +. Higher scores

thus indicated greater preference for flowers and greater dislike for insects.

The IAT scores were entered into a 2 (scenario: present vs. absent) x 2 (IAT type:

personalized vs. traditional) ANOVA. The results revealed a significant scenario x IAT type

interaction, $F(1, 115) = 10.49$, $p = .002$ (See Figure 1). 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.

Thus, exposure to the scenario significantly impacted the traditional IAT. This occurred even though the scenario was presented as a separate, unrelated, and completed 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. The focused labels of the personalized IAT inhibited any such use of the extrapersonal information. Paralleling Experiment 1, it seems very implausible that the manipulation – in this case, exposure to a brief and obviously fictional scenario – could have had any effect on participants' underlying attitudinal representations. Yet, IAT performance was affected.

Discussion

The results from Experiment 2 indicate that when participants read counterfactual scenarios regarding insects and flowers, their scores on the traditional IAT showed significantly less negativity toward insects. This did not occur on the personalized IAT. The implication is that the malleability effects evidenced in prior research using the traditional IAT may stem from this measure's imprecise labels. The labels "pleasant/unpleasant" seem sensitive to salient extrapersonal associations, and do not necessarily reflect the attitude representation itself. 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 change in participants' attitudes. In short, findings that have been interpreted as evidence of the malleability of attitudes might more appropriately be considered evidence of the malleability of the measure.

Experiment 3

Our previous findings show that the malleability of label construal leaves the traditional IAT susceptible to the influence of extrapersonal associations. As a result, the measure can suggest attitude change when none has occurred. What happens if, following a persuasion attempt, attitude change does occur? Our last experiment illustrates the reverse side of the IAT's malleability, namely that the IAT as conventionally implemented can sometimes fail to reveal attitude change when such change is likely to have occurred. We show that, under specific persuasion conditions, the influence of extrapersonal associations on the traditional IAT may obscure the detection of attitude change.

Recently, Briñol, Petty, and McCaslin (2009) provided a comprehensive review of research concerning the impact of persuasive messages on attitudes as measured implicitly by the traditional IAT. They offered a compelling explanation for the mechanism by which deliberative persuasion processes may affect subsequent IAT performance. When a message is processed in detail, each positive or negative argument is carefully assessed in terms of its validity. If the message presents strong convincing arguments, this thoughtful process should lead to attitude rehearsal, thus increasing the accessibility of one's positivity or negativity regarding the object. The effect of this attitude rehearsal should be apparent in the IAT, by facilitating the completion of either its compatible or incompatible task, whichever response mapping more closely corresponds to the individual's newly-developed or updated attitude. However, if the message

presents weak arguments, argument-consistent attitude rehearsal is less likely to occur. Instead, individuals would easily dismiss or counterargue the message. Briñol et al. (2009) report several experiments showing that attitude change can be evident on the traditional IAT and that evidence of such persuasion is more likely when the arguments are stronger.

Just like Briñol et al. (2009), we focus our attention in Experiment 3 on persuasion under high elaboration conditions. We extend Briñol et al.'s reasoning by considering the potentially differential sensitivity of IAT type (traditional vs. personalized) on the measurement outcome following persuasion attempts of varying strength. If the persuasion is sufficiently strong, it is likely to shape personal attitudes. However, in some persuasion situations, the traditional IAT may not fully reflect this attitude change. To the extent that a persuasive setting makes salient some extrapersonal associations that are contrary to the newly-revised attitude, the sensitivity of the traditional IAT to the changed attitude may be compromised. Even after a successful persuasion attempt, some message-inconsistent extrapersonal associations may remain salient and serve to manage the traditional, but not the personalized, IAT's mapping problem. In such situations, despite the occurrence of attitude change, the traditional IAT would fail to capture the full extent of that change.

In order to investigate this possibility, we needed to find an attitude domain where extrapersonal associations are prevalent. The likely candidates seemed to be domains in which there are obvious reasons for favoring *either* side of an issue. Our choice fell on the domain of luxury products, because such brands are characterized by a rich variety of potential associations, including quality-related beliefs, hedonic benefits and prestige connotations. In addition, given the substantial social visibility of luxury items, their evaluation may be subject to the influence of social norms, as well as concerns with socially desirable responding. Thus, people may

associate luxuries with information that does not itself contribute to their attitudes. For example, a person with a favorable attitude toward Mercedes-Benz autos may evoke a variety of thoughts about the brand, including its sporty engines, comfortable interior, and excellent reliability -- thoughts that contribute to his or her positive attitude toward the brand (personal associations). At the same time, this person may acknowledge that some people find Mercedes-Benz to be overpriced and snobbish (extrapersonal associations). Although these latter two associations may not affect this person's global positive attitude toward Mercedes-Benz, they may be salient at the time the IAT is administered.

In the present experiment, the persuasion setting involved reading other people's arguments about luxury brands, after which the extent of attitude change concerning luxuries was measured with either a traditional or personalized IAT. Sufficiently strong arguments should lead to attitude change and, when the strong arguments are not accompanied by extraneous material, such change should be detected with both IAT versions. However, what if the persuasion context were to include not only strong arguments in one direction but also a weak argument in the opposite direction? Although it is unlikely to influence attitudes, this single weak argument may act as a salient extrapersonal association at the time of attitude measurement. In such a situation, the traditional IAT may fail to capture the full extent of attitude change, relative to the personalized IAT.

Method

Design. Participants were randomly assigned to one of three persuasion conditions. The first condition was a control in which no arguments were presented; participants only completed a traditional or personalized IAT assessing preference for luxury versus common brands. In the other two conditions, we employed a strong positive message, but varied whether the context

also introduced a negative extrapersonal association. Accordingly, in the second condition, we presented an individual's comments arguing strongly in favor of luxury brands, along with a clear statement that the particular message had been selected because this person's views were very representative of collective opinion. We expected both IAT variants to detect positive attitude change in this condition. In the third condition, a negative extrapersonal association was made very salient immediately after the presentation of the strong positive arguments. In this condition, a brief message, explicitly labelled as unrepresentative of collective opinion, introduced mention of a negative comment about the snobbishness of luxuries. We expected this weak negative comment to increase the salience of a negative extrapersonal association; however, we did not expect this single unrepresentative comment to undermine the attitude change induced by the strong positive arguments. Thus, we expected the traditional IAT to be affected by this negative comment and show less evidence of attitude change than in the condition in which only the positive message was presented. On the other hand, we expected the personalized IAT to remain impervious to the brief negative comment and evidence the same attitude change as in the positive argument alone condition.

Experimental material development. The experiment was conducted in the Paris area in France and was preceded by two pilot studies. In a first pilot study, 16 individuals were interviewed with the aim of identifying material that could serve as the basis for the development of persuasive messages of varying strength. The participants were asked to remember and describe their recent acquisitions of, or experiences with, luxury items. They were then asked to describe their views about luxuries in general and the various product/service categories and brands associated with them. Most interviewees evaluated luxury brands positively and emphasized the quality, exceptional design and sophistication of those brands. A few viewed

luxuries somewhat negatively, largely in light of the snobbish aspects of some brands. We adapted characteristic sections from these interviews to develop our persuasive messages.

A second pilot study with 49 undergraduate students was designed to develop stimulus exemplars for a luxury brands/common brands IAT. Participants were asked to list brands that were "most" and "least" luxurious in their view. The six most frequently cited luxury brands and the six most frequently cited common brands were selected as target stimulus items for the IAT. As a result, the luxury brands stimuli featured the prestigious international names *Chanel*, *Vuitton*, *Dior*, *Gucci, Mercedes* and *Ferrari*. For common brands, the most frequently cited in the pretest and, hence, selected for use were *Zara* (clothing brand), *H&M* (clothing brand), *Ariel* (detergent), *Monoprix* (grocery retailer), *Bic* (office furniture) and *Carrefour* (retailer). Valence attributes included *peace*, *paradise*, *joy*, *love*, *pleasure* and *happiness* for positive attributes and *disaster*, *grief*, *accident*, *pain*, *bad* and *agony* for negative attributes. We employed the same IAT procedures as in the previous experiments. Thus, the compatible block of the traditional (personalized) IAT consisted of combined tasks with the labels luxury brands/pleasant (I like) and common brands/unpleasant (I don't like); the incompatible block featured the labels luxury brands/unpleasant (I don't like) and common brands/pleasant (I like).

Participants and procedure. Two-hundred and nineteen students participated in a 3 (message arguments: no arguments; positive arguments; positive arguments followed by a negative argument) × 2 (IAT type: traditional, personalized) between-subjects design. Participants were randomly assigned to the experimental conditions in a lab equipped with six workstations. Participants in the no arguments condition only completed either the traditional or personalized IAT regarding luxury brands/common brands. In the other two conditions, participants first completed a paper-and-pencil task. They were told that the present research

investigated how people formed impressions about unknown people on the basis of various

interview excerpts. Participants in the positive arguments condition read an introduction

explaining that the excerpts came from an interview with "Eric who is 50 and lives in Paris,"

with the note that his opinions were representative of what most people thought. They were then

presented with the interview excerpts and were asked to read them carefully so as to be able to

form a first impression of the source (see Appendix). After reading the entire set of interview

excerpts, they responded to a series of manipulation check questions on a 1 (strongly disagree) to

7 (strongly agree) scale: "This person thinks that luxury brands mean snobbery"(reverse-coded),

"This person thinks that luxury brands are well regarded in society", and "Overall, this person

thinks that luxuries are a good thing". These items were averaged to serve as checks for the

message valence manipulation. We also included three additional items measured on 1-7 scales

intended to assess perceptions of the arguments (not interesting/interesting, weak/strong, not

convincing/convincing). After this, participants completed either the traditional or personalized

IAT. Participants in the positive arguments + negative argument condition followed a similar

procedure, with the only difference that after reading and evaluating the representative positive

arguments presented by Eric, they also read a short quote from another person, Michel, 35,

commenting on the snobbery of luxury brands (see Appendix). Michel's comment was clearly

presented as unrepresentative of the people sampled. Identical manipulation checks were

administered after this interview excerpt, too.

Results

      Manipulation checks. In the condition in which positive arguments were followed by the

brief negative comment of another interviewee, participants judged the positive (vs. negative)

source to view luxuries as less snobbish ($M = 6.31$, $SD = .70$ vs. $M = 2.15$, $SD = 1.33$, $t(74) =$

23.60, $p < .001$), better regarded in society ($M = 5.16$, $SD = 1.24$ vs. $M = 2.45$, $SD = 1.01$, $t(74) =$

14.68, $p < .001$) and as globally good things ($M = 5.53$, $SD = 1.18$ vs. $M = 2.37$, $SD = .86$, $t(74) = 21.17$, $p < .001$). They also evaluated the positive (vs. negative) arguments as more interesting ($M = 4.59$, $SD = 1.48$ vs. $M = 3.60$, $SD = 1.58$, $t(74) = 4.25$, $p < .001$), stronger ($M = 3.72$, $SD = 1.35$ vs. $M = 3.00$, $SD = 1.39$, $t(74) = 3.39$, $p = .001$) and more convincing ($M = 3.96$, $SD = 1.47$ vs. $M = 3.22$, $SD = 1.45$, $t(74) = 3.13$, $p = .002$). These results suggest that the manipulation of message valence and strength operated as intended. The brief comment offered by the unrepresentative interviewee was regarded as more negative, but clearly weaker, than the lengthier arguments offered by the representative interviewee.

Evidence of attitude change with the two IAT versions. Data were prepared as in the previous experiments.[5] A 3×2 ANOVA was conducted with the IAT measure as the dependent variable and message arguments and IAT type as factors. The mean IAT scores as a function of message arguments and IAT type are presented in Figure 2. Significant main effects for both message arguments, $F(2, 213) = 8.57$, $p = 0.001$, and IAT type, $F(1, 213) = 6.64$, $p = 0.01$, emerged. Those main effects were qualified by a significant two-way interaction between message arguments and IAT type, $F(2, 213) = 3.10$, $p = 0.047$. Simple contrasts evidence that when only positive arguments are presented, the traditional IAT reflects a positive attitude change ($M = 254$, $SD = 160$) compared to the baseline ($M = 120$, $SD = 152$), $F(1, 213) = 10.26$, $p = 0.002$. However, if the positive arguments were immediately followed by a negative argument, scores on the traditional IAT measure were substantially reduced ($M = 141$, $SD = 230$) compared to the positive arguments only condition ($M = 254$, $SD = 160$), $F(1, 213) = 9.68$, $p = 0.002$. Indeed, if the positive arguments were immediately followed by a negative argument, no attitude change was detected with the traditional IAT ($M = 141$, $SD = 230$) compared to the baseline control condition ($M = 120$, $SD = 152$), $F < 1$.

The pattern of results was different with the personalized IAT. When only positive arguments were presented, the personalized IAT reflected a positive attitude change ($M = 278$, $SD = 274$) compared to the baseline ($M = 152$, $SD = 166$), $F(1, 213) = 7.25$, $p = 0.008$. Likewise, if the positive arguments were immediately followed by a negative argument, a positive attitude change was still detected ($M = 292$, $SD = 223$) relative to the baseline ($M = 152$, $SD = 166$), $F(1, 213) = 7.15$, $p = 0.008$. Thus, if the positive arguments were immediately followed by a negative argument, the personalized IAT scores reflected the same amount of attitude change ($M = 292$, $SD = 223$) as was evident in the positive arguments only condition ($M = 278$, $SD = 274$), $F < 1$. These results support our prediction that the negative extrapersonal association activated through a brief weak message may mask attitude change if the traditional IAT, but not if the personalized IAT, is used.

Discussion

We hypothesized that strong positive arguments about the quality of luxury brands would induce attitude change, but that a relatively weak negative argument would not influence attitudes. Although dismissed, the negative comment would heighten the salience of negative extrapersonal associations about the snobbishness of luxury brands. In this case of a persuasion attempt introducing a weak negative point in the presence of strong positive arguments, we expected the traditional (vs. personalized) IAT to no longer successfully reflect attitude change. The pattern of results observed in Experiment 3 supported these expectations. A strong positive message likely induced attitude change and this change was captured by both IATs. However, if this attitude change was followed by a weak counter message, the traditional IAT was affected by that brief counterpoint and failed to show evidence of the original attitude change. Not so the personalized IAT, which exhibited a measurement outcome still reflecting attitude change. From

the perspective of the framework adopted in this paper, these findings provide further evidence that, due to its unspecified labels, the traditional IAT is sensitive to extrapersonal associations that are salient at the time the attitude assessment is made.

It is important to consider the divergent findings as a function of IAT version within the context provided by the other data from the experiment. In the control condition, scores on both IATs were predominantly positive, reflecting a pre-experimental preference for luxury products. According to the manipulation check data, the lengthy positive comments of the interviewee who was described as representative of the sample as a whole were considered stronger and more convincing than the brief negative comment of the unrepresentative interviewee (see the Appendix for the actual text). Exposure to the former endorsement of luxuries, in and of itself, intensified participants' initial preferences, according to both versions of the IAT. Thus, the pro-luxuries message clearly was effective. Yet, when attitudes were assessed by the traditional IAT, this evidence of attitude change was completely overridden by the weak counterattitudinal message. The mean in this condition was equivalent to that in the control condition, and significantly lower than that in the condition in which only the positive arguments were presented (see Figure 2). Somehow, a brief, unsubstantiated, counterattitudinal comment – one that was perceived as weak and unconvincing and that stemmed from a source explicitly labelled as atypical – left participants seemingly unaffected by the documented persuasiveness of the proattitudinal arguments to which they had been exposed. How plausible is it that attitudes would revert back to their original position in response to such a message? Readers will have to judge for themselves, but to us, it seems highly doubtful. Yet, this is the portrait painted by the traditional IAT. In contrast, the personalized IAT displays a much more plausible pattern,

whereby the persuasive impact of the pro-attitudinal endorsement remains apparent even when it was followed by the ever so easily-dismissed counterpoint.

General Discussion

The goal of the present experiments was to explore the malleability that results from the multiple plausible meanings permitted by the evaluative labels of the traditional IAT. We also investigated 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. In addition, we showed that the malleability of the IAT can sometimes obscure the detection of attitude change when the occurrence of such change was likely.

In the first experiment, we demonstrated that interpretation of the labels "pleasant/unpleasant" can be influenced by a prior unrelated task that increased the salience of a personal vs. normative focus. The findings suggest 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. The accessibility of these differential perspectives can influence performance, and hence, the attitude estimates that are obtained.

In Experiment 2, we explored the consequences of the malleability of the IAT for interpretive issues regarding the occurrence of attitude change. Experiment 2 replicated Foroni and Mayr (2005) and showed that only the traditional IAT was impacted by counterfactual information made salient in a preceding task. Participants who completed the traditional IAT showed less of a preference towards flowers immediately after they read a hypothetical scenario in which flowers were negative and insects positive. The personalized IAT did not show such fluctuation, supporting our claim that the findings on the traditional IAT were driven by the

presence of extrapersonal associations rather than the formation of "new automatic associations" that overrode longstanding personal associations to flowers and insects.

In Experiment 3, we proposed that the traditional IAT implemented in most prior research can sometimes fail to capture the full extent of attitude change in settings where extrapersonal associations are prevalent. This experiment supports the view that the traditional and personalized IATs are differentially sensitive to persuasion in such attitude domains. These findings are all the more meaningful when considered in conjunction with the Han et al. (2006) research. That work demonstrated that the traditional IAT can be sensitive to a rather trivial, unjustified, and objectively incorrect remark offered by two 10-year old children. The remark did not produce attitude change as evidenced by its failure to affect either the personalized IAT or a priming measure, but it did serve as an attitudinally inconsistent extrapersonal association. As a result, the traditional IAT scores suggested that change had occurred in response to the boys' comment, when other measures, including assessments of the boys, suggested that it had not. The present research illustrates what might be considered the flip side: attitudes *did change* as a function of a strong positive message, and that change was consistently evident on the personalized IAT. However, in this experiment, a brief comment alluding to the snobbishness of luxury brands, by an interviewee whose opinion was explicitly labelled as unrepresentative, was sufficient to render the traditional IAT insensitive to the attitude change produced by the strong positive arguments. Only when that positive message had been identified as representative of all the interviewees sampled and not followed by the presumably rare comment to the contrary did the traditional IAT reveal the change that was apparent on the personalized IAT. Thus, due to its susceptibility to extrapersonal associations, the traditional IAT can suggest change where none

has occurred, as in Han et al. (2006) and the present Experiment 2, or fail to reveal change that has occurred, as in Experiment 3.

In sum, Experiments 2 and 3 revealed striking disparities between the personalized and traditional versions of the IAT. These findings provide further support for the value of the conceptual distinction between personal and extrapersonal associations. We have difficulty seeing how the consequences of the substitution of "I like/I don't like" for "pleasant/unpleasant" can be explained without some reference to the personal/extrapersonal distinction. Not all the knowledge an individual possesses about a given attitude object necessarily contributes to the evaluative reaction evoked by that object. Information that does not contribute meets our definition of extrapersonal and, on the basis of the present findings, appears less likely to affect IAT performance when the labels specify a personal perspective.

On the value of focusing the IAT

The first experiment regarding the malleability of the IAT adds 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. The mindset results suggest that, depending on the previous task or the instructions, participants may actually complete the IAT in different ways. It has already been shown that the IAT can be 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; Mitchell et al., 2003; Steffens & Plewe, 2001). The current findings add to this literature by demonstrating that the IAT is sensitive to unrelated contexts and mindset priming. It seems that a variety of factors may interfere with the capability of the traditional IAT to gauge participants' attitudes.

On the other hand, the present results also 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 culturally-oriented versions of the IAT could potentially be implemented to measure normative or cultural associations. In Experiment 1, IAT scores were significantly higher (and, hence indicative of more racial prejudice) when people were primed to think in a normative mindset ("People like/don't like"). These data suggests that individuals can complete the IAT with the perspective of "others" in mind. Indeed, in an experiment that examined such a culturally-oriented IAT, Han, Olson and Fazio (2006) observed stronger evidence of racial bias with the attribute labels "People like/don't like" than with the "I like/don't like" labels of the personalized IAT. Such culturally-oriented implementations may be especially useful when attempting to ascertain individuals' perceptions of general societal evaluations or predict compliance with normative expectations.

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 the possibility of multiple interpretations by using clearly defined category labels.

Inferences regarding attitude change

Experiment 2 demonstrated that the traditional IAT's susceptibility to extrapersonal associations 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. 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.

Our goal has not been 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. Corroborating the evidence reviewed by Briñol et al. (2009), Experiment 3 illustrates the influence that exposure to strong persuasive arguments can have on attitudinal associations. In addition, various conditioning procedures have been shown to produce changes in attitudinal representations (e.g., Kawakami, Phills, Steele, & Dovidio, 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 easily 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), we do 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 it to assume that reading a short counterfactual scenario

changes one's automatically activated attitudes towards insects or flowers? Could watching two minutes of a counter-stereotypical movie or seeing several images of atypical exemplars change attitudinal representations regarding Blacks? If such possibilities were 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.

In the present work, we have focused on the difficulty of distinguishing attitude malleability from measurement malleability. Before closing, however, we should highlight an additional distinction that relates to the confidence with which inferences can be drawn about a change in attitudinal representations upon observation of change on some outcome measure. As Asch (1948) so clearly articulated and demonstrated decades ago, a change in expressed evaluations is sometimes due, not to a change "in the judgment of the object," but to "a change in the object of the judgment" (p. 256). In other words, attitude objects themselves are quite malleable, in the sense that they often can be construed in multiple ways. The importance of such "change of meaning" is well-established in the impression formation literature. For example, the trait ascription "proud" can connote either "conceited" or "confident" depending upon the likability of the target person (Hamilton & Zanna, 1974). Similarly, current motivational states or contextual cues can foster varying construals of an attitude object. When an individual is thirsty, a hose may evoke much more positivity than it typically does, but this is

because it is now being construed as a thirst-quenching water fountain, instead of as a device for washing one's car. Likewise, an African-American male dressed in a suit will be viewed very differently than the same person dressed in a prison inmate uniform and, as a result, evoke different evaluative responses (Barden, Maddux, Petty & Brewer, 2004). How an attitude object is construed in the current situation determines the evaluation that is automatically activated, and such construals are certainly malleable (see Fazio, 2007, for further discussion of this issue). Thus, just as Asch demonstrated that an apparent change in one's evaluation of an object may be due to either a change in the object of judgment or a change in one's judgment of it, we have demonstrated that such apparent changes may be due to either malleability of the attitude or malleability of the measure used to assess the attitude.

Conclusion

In conclusion, the IAT's sensitivity to extrapersonal associations leaves it susceptible to momentary contextual influences – ones for which it is difficult to plausibly entertain that attitudes have changed. Presumably, no researcher would want their measure of attitudes to be influenced by the perspective mindset induced by a preceding task, or by a brief mental excursion into an obviously fictional future world. We doubt anyone would argue that these preceding events changed attitudes. Yet, they changed scores on the IAT, as it is traditionally implemented. Moreover, the last experiment shows that this sensitivity has the potential to obscure the detection of change in a situation where there is every reason to believe that attitude change has occurred. Fortunately, this problem can be addressed by focusing the IAT through the use of the more specific labels "I like" and "I don't like." Thus, the present work goes beyond the mere illustration of a problem with the IAT; it offers a way to improve the measure by the simple modification of personalizing the labels.

Appendix

Materials in Experiment 3

Positive Arguments

Please carefully read the following excerpts from an interview with Eric, 50, living in Paris. His thoughts are actually quite representative of how most people think in the sample we interviewed.

"For some things, I will take luxury. If it's for a gift, you have to buy a name brand. I can buy a prestige champagne if it's necessary, for very memorable events or as a gift. If I'm about to present a gift to someone, well, I will not buy an unknown brand. To simplify the purchase, I will go to a great department store, I'd for example only go to the Bon Marché in Paris"

"I have memories of exquisite, mysterious hotels, giving the impression of entering a cozy, closed world, and I'm very sensitive to the charm of that kind of hotel. If we take the Mamounia as an example, I was 15 when I first came into contact there with luxury hotels, the thing that really impressed me was that it was a hotel for heads of state. This was where Churchill had spent much of his time during the war. And in the end, what people enjoy about luxury hotels, it's being on the other side of the wall. Finally, passing to the other side of the wall is essential. At the San Regis hotel in Paris, a coffee costs 10 euros but you feel truly happy, you don't regret those 10 euros"

"Now if I go out for dinner with a friend, really to go out, in this case it will depend on the person, on what he or she will like But I may say, Well, why don't we go to the Tour d'Argent or something like that … If I know they would appreciate, then why not !"

"On an evening out, or an a day of invitation or reception, or something like that, I think you have to wear a certain attire … If we go to a gala dinner or something like that, banquets by friends, if we see someone who is not dressed according to our criteria, we say to each other, this one, well, this person could have made an effort to dress properly. If you wear luxury clothes, well-designed ones, you don't have to pay attention to a certain number of things. For me, luxury means basically taking a cab instead of taking the underground, eating out in restaurants often, staying in great hotels when I'm on a trip"

"Regarding great wines, we have some fine bottles but we only open them if we invite connoisseur friends that we want to make a special honor to. I sometimes drink fine wines with friends who are really knowledgeable about them. We sometimes order a thing like that in a restaurant, it may happen. If you are with people who share the same preferences as you or if you

are celebrating an event, then you can make an even bigger effort, you can buy something very very good. To drink a bottle of good wine alone or with people who don't care makes no sense. It's something rare, something that's exceptional. So something that needs exceptional moments with people who appreciate that. It is for quality guests only, in the sense of people who will like it. I can offer wine as a gift, I sometimes take a bottle or two out of the cellar to offer it to people who will really like them"

"If I buy cognac for my guests, in this case I will not buy the average quality, I will buy something better, of superior quality, that is, VSOP, something better than average. I have already given great cognacs to some people as a gift. I may choose it depending on the packaging, on the case … It's true that visual appearance counts a lot for the person who receives it … and also for one who is buying it. It must give pleasure to the recipient if it's nicely presented … The way it looks like counts a lot"

Positive Arguments Followed by a Negative Argument

Same as above, followed by:

Please carefully read the following excerpts from another interview with Michel, 35, who also lives in Paris. His thoughts are not really representative of how most people think in the sample we interviewed.

"I don't know, maybe I would avoid buying a frame with the name Dior, St-Laurent, Nina Ricci or whatever written on it … I'm not very familiar with those brands. Someone who is fond of luxury things would be someone who cares about brands in the first place, that is, buys a brand because it's a famous brand ... Someone who goes to a hotel because it's a five-star hotel, who will buy Mercedes. Dior, for example, it's about couture, fragrance, it's a little bit about show-off"

References

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

Asch, S. E. (1948). The doctrine of suggestion, prestige, and imitation in social psychology. *Psychological Review, 55,* 250-276.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87*, 5-22.

Bickart B.A. (1992). Question-order effects and brand evaluations: The moderating role of consumer knowledge. In N. Schwarz and S. Sudman (Eds.), *Order effects in social and psychological research* (pp. 63-79). New York: Springer-Verlag.

Billiet, J.B., Waterplas, L., & Loosveldt, G. (1992). Context effects as substantive data in social surveys. In N. Schwarz and S. Sudman (Eds.), *Order effects in social and psychological research* (pp. 131-147). New York: Springer-Verlag.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*, 828 - 841.

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

Bless, H. & Wänke, M. (2000). Can the same information be typical and atypical? How perceived typicality moderates assimilation and contrast in evaluative judgments. *Personality and Social Psychology Bulletin*, *26*, 306-314.

Briñol, P., Petty, R. E., & Mc Caslin, M. J. (2009). Changing attitudes on implicit versus explicit

    measures: What is the difference? In R. E. Petty, R. H. Fazio & P. Briñol (Eds.),

    Attitudes: insights from the new implicit measures. New York: Psychology Press.

Carpenter, S., & Banaji, M. R. (2001, February). *Malleability in implicit stereotype and attitudes.*

    Paper presented at the annual meeting of the Society for Personality and Social

    Psychology. San Antonio, TX.

Conner, M. T., Perugini, M., O' Gorman, R., Ayres, K., & Prestwich, A. (2007). Relations

    between implicit and explicit measures of attitudes and measures of behavior: Evidence

    of moderation by individual difference variables. *Personality and Social Psychology*

    Bulletin, *33*, 1727-1740.

Dasgupta, N., &Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating

    automatic prejudice with images of admired and disliked individuals. *Journal of*

    Personality and Social Psychology, 81, 800-814.

De Houwer, J. (2009). Comparing measures of attitudes at the functional and procedural level:

    Analysis and implications. In R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes:*

    Insights from the new implicit measures. New York, NY: Psychology Press.

De Houwer, J., Custers, R., & De Clercq, A. (2006). Do smokers have a negative implicit

    attitude toward smoking? *Cognition & Emotion, 20*, 1274-1284.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit

    measures. In B. Wittenbrink& N. Schwartz (Eds.), *Implicit measures of attitudes* (pp.

    179-194). New York: Guilford Press.

Deutsch, M., & Gerard, B. (1955). A study of normative and informational social influences

    upon individual judgment. *Journal of Abnormal and Social Psychology*, *51*, 629-636.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: determinants, consequences, and

correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude*

strength: antecedents and consequences (pp. 247-282). Mahwah NJ: Lawrence Erlbaum.

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

Cognition, 25, 603-637.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and

controlled components of racial prejudice. *Journal of Experimental Social Psychology,*

33, 451-470.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their

meaning and use. *Annual Review of Psychology*, 54, 297-327.

Foroni, F., & Mayr, U. (2005). The power of a story: New, automatic associations from a single

reading of a short scenario. *Psychonomic Bulletin & Review, 12*, 139-144.

Gawronski, B., Peters, K. R., & LeBel, E. P. (2008). What makes mental associations personal or

extra-personal? Conceptual issues in the methodological debate about implicit attitude

measures. *Social and Personality Psychology Compass*, *2*, 1002-1023.

Goff, P. A., Eberhardt, J. L., Williams, M., & Jackson, M. C. (2008). Not yet human: Implicit

knowledge, historical dehumanization, and contemporary consequences*. Journal of*

Personality and Social Psychology, 94, 292-306.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items

influences the IAT by re-defining the category labels. *Journal of Experimental Social*

Psychology, 40, 357-365.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and

stereotypes. *Psychological Review*, 102, 4-27

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

differences in implicit cognition: The Implicit Association Test. *Journal of Personality*

and Social Psychology, *74*, 1464-1480.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-

esteem and self-concept. *Journal of Personality and Social Psychology*, 79, 1022-1038.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit

Association Test: An improved scoring algorithm. *Journal of Personality and Social*

Psychology, 85, 197-215.

Hamilton, D. L., & Zanna, M. P. (1974). Context effects in impression formation: Changes in

connotative meaning. *Journal of Personality and Social Psychology*, *29*, 649-654.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally-created

extrapersonal associations on the Implicit Association Test. *Journal of Experimental*

Social Psychology, 42, 259-272.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006, January). *Responsivity of the IAT to*

conversational norms. Paper presented at the annual meeting of the Society for

Personality and Social Psychology, Palm Springs.

Houben, K., & Wiers, W. (2007). Are drinkers implicitly positive about drinking alcohol?

Personalizing the alcohol-IAT to reduce negative extrapersonal contamination. *Alcohol*

and Alcoholism, *42*, 301-307.

Jordan, C. H., Spencer, S. J., Zanna, M. P., Hoshino-Browne, E., &Correll, J. (2003). Secure and

defensive high self-esteem. *Journal of Personality and Social Psychology, 85,* 969-978.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 774-778.

Kawakami, K., Phills, C. E., Steele, J. R., & Dovidio, J. F. (2007). (Close) distance makes the heart grow fonder: Improving implicit racial attitudes and interracial interactions through approach behaviors. *Journal of Personality and Social Psychology*, *92*, 957-971

Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response order effects in survey measurement. *Public Opinion Quarterly*, 51, 201-219.

Krosnick, J. A. (1992). The impact of cognitive sophistication and attitude importance on response order effects and question order effects. In N. Schwarz and S. Sudman (Eds.), *Order effects in social and psychological research* (pp. 203-218). New York: Springer-Verlag.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842-855.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*, 455-469.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*, 166–180.

Nosek, B. A., & Hansen, J. J. (2008a). The associations in our heads belong to us: Searching for attitudes and knowledge in implicit evaluation. *Cognition and Emotion*, *22*, 553-594.

Nosek, B. A., & Hansen, J. J. (2008b). Personalizing the Implicit Association Test increases explicit evaluation of the target concepts. *European Journal of Psychological Assessment*, *25*, 226-236.

Olson, M. A., & Crawford, M. T., & Devlin, W. (2009). Evidence for the underestimation of implicit in-group favoritism among low status groups. *Journal of Experimental Social Psychology, 45,* 1111-1116.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667.

Olson, M.A., Fazio, R.H., & Han, H.A. (2009). Conceptualizing personal and extrapersonal associations. *Social and Personality Psychology Compass*, *3*, 152-170.

Peach, J. M., Yoshida, E., Zanna, M. P., & Spencer, S. J. (January 2006). *Cultural norms and personal attitudes as distinct influences on the IAT score*. Paper presented at the annual meeting of the Society of Personality and Social Psychology, Palm Springs, CA.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The meta-cognitive model (MCM) of attitudes: implications for attitude measurement, change, and strength. *Social Cognition, 25*, 657-686.

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

Salancik, G. R., & Conway, M. (1975). Attitude inferences from salient and relevant cognitive content about behavior. *Journal of Personality and Social Psychology*, *32*, 829-840.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*, *54*, 93-105.

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology, Vol. 1: Intraindividual processes* (pp.

436-457). Oxford, UK: Blackwell.

Shavitt, S., & Fazio, R. H. (1991). Effects of attribute salience on the consistency between attitudes and behavior predictions. *Personality and Social Psychology Bulletin,* 17, 507-516.

Sherman, S. J., Presson, C. C., Chassin, L., Rose, J. S., & Koch, K. (2002). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology, 22*, 13-39.

Shook, N. J., & Fazio, R. H. (2008). Interracial roommate relationships: An experimental field test of the contact hypothesis. *Psychological Science, 19*, 717-723.

Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *ZeitschriftfürExperimentellePsychologie, 48,* 123-134.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529-554.

Wilson, T. D. & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. Martin and A. Tesser (Eds.), *The construction of social judgment* (pp. 37-65). Hillsdale, NJ: Erlbaum.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815-827.

Yoshida, E., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2006). *Automatic evaluation of what most people like: Developing an implicit measure of norms.* Paper presented at the 7th Annual meeting of the Society for Personality and Social Psychology, Palm Springs, CA.

Author Notes

H. Anna Han, Department of Psychology, St. Mary's College of Maryland.

Sandor Czellar, Department of Marketing, HEC Paris School of Management.

Michael A. Olson, Department of Psychology, University of Tennessee.

Russell H. Fazio, Department of Psychology, Ohio State University.

Correspondence concerning this article should be addressed to H. Anna Han, Department of Psychology, St. Mary's College of Maryland, 18952 E. Fisher Rd., St. Mary's City, MD 20686. E-mail: hahan@smcm.edu.

Footnotes

[1] For the reasons delineated by Olson and Fazio (2004), we have chosen to present the

data in terms of the original IAT scoring procedure rather than D-score algorithm proposed by

Greeenwald, Nosek, & Banaji (2003). Most importantly, we have reservations about a scoring

system that imposes a 600ms penalty on the latency for each trial on which the participant has

made an error. We view this as an arbitrary means of weighting error and latency data to form a

single composite. In our experience, imposition of the 600ms error penalty can weaken the

statistical effects that are observed, especially in experiments with relatively small samples.

However, each of the present experiments involved sufficient power that all the reported effects

remain statistically significant when the D600 scoring algorithm is employed.

[2] One might ask whether the usual absence of error feedback in the personalized IAT also

plays a role. We suggest its role is indirect, simply lessening any tendencies to adopt a

normative perspective. We have not observed the absence of feedback to lessen participants'

concerns about making errors. As noted in the respective articles, 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). These null findings stand 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. We can only

conjecture that the discrepancy stems from lesser attentiveness 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 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.

[3] The scenario presented was identical to the appendix material of Foroni and Mayr

(2005), except for our having deleted the last three sentences in which the IAT was presented as

a training program.

[4]There was no evidence of differential error rates in the traditional ($M = 8.0\%$, $SD = 6.3$)

vs. personalized IAT versions ($M = 7.4\%$, $SD = 6.7$), $t(112) < 1$.

[5]Once again, error rates in the traditional ($M = 3.93\%$, $SD = 3.31$) vs. personalized IAT

versions ($M = 4.91\%$, $SD = 6.05$), did not differ significantly, $t(218) = 1.49$, $p = .14$.

Figure Captions

Figure 1. Experiment 2. IAT scores as a function of IAT type and story condition.

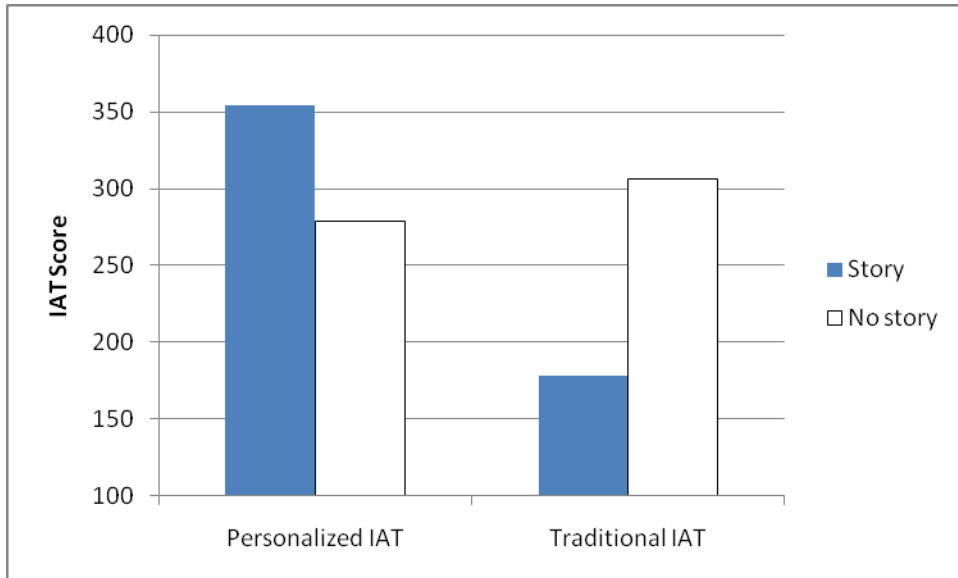*Figure 2.* Experiment 3. IAT scores as a function of message arguments and IAT type.

Fig.1

Fig. 2