Auditory Implicit Association Tests

Mark E. Vande Kamp

A dissertation submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2002

Program Authorized to Offer Degree: Psychology

UMI Number: 3072151

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of the dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, MI 48106-1346, or to the author.

Signature *Mark Vande Kamp*

Date 12/17/02

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Mark E. Vande Kamp

and have found that it is complete and satisfactory in all respects, and that any and all revisions required by the final examining committee have been made.

Chair of Supervisory Committee:

A. G. Greenwald

Anthony G. Greenwald

Reading Committee:

Earl Hunt

Earl B. Hunt

F. Beth Kerr

F. Beth Kerr

A. G. Greenwald

Anthony G. Greenwald

Date: December 13, 2002

University of Washington

Abstract

Auditory Implicit Association Tests

Mark E. Vande Kamp

Chairperson of the Supervisory Committee:
Professor Anthony G. Greenwald
Psychology

Three studies demonstrated that auditory implicit association tests (auditory IATs) using speech and environmental-sound stimuli could be used to measure implicit attitudes and stereotypes. Study 1 showed a preference for songbirds over insects in auditory IATs that used a) speech stimuli (spoken names of songbirds and insects), and b) stimuli consisting of sounds made by songbirds and insects, in conjunction with text stimuli consisting of positive and negative words. Study 2 showed that two IATs using only auditory stimuli could measure implicit gender stereotypes. The first used male and female voices in conjunction with high-power and low-power words spoken by a non-gendered computerized voice to measure sex differences in gender stereotypes. The second used the same high- and low-power words in conjunction with male and female names spoken by the non-gendered voice. Study 3 showed implicit attitudes favoring European-Americans over African-Americans using an auditory IAT in which participants classified speech segments based on the accent of the speaker, thus measuring personally and/or socially undesirable implicit racial attitudes that were not detected through explicit self-report measures. The studies demonstrate that the auditory IAT is an extension of the IAT that can measure implicit associations in a variety of situations in which IATs consisting entirely of visual or text stimuli are not applicable. Exploration of the relationships between the various IATs suggested that their results reflect automatic cognitive associations that are influenced by factors within the IAT procedure, such as the stimuli chosen to represent the target categories, and external contextual factors such as associations made salient by preceding experiences. For

example, Study 3 found that a picture IAT using neutral faces measured significantly stronger implicit attitudes favoring European-Americans over African-Americans than were measured by a sound IAT using friendly greetings. However, the difference was only observed when the picture IAT was completed first. When the sound IAT was first, the associations with friendly targets apparently remained salient, and implicit attitudes did not differ. Such findings suggest possibilities for better understanding automatic associations and the way they are measured by the IAT.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Of the many attributes that I have come to appreciate in the people who have helped me achieve this goal, the foremost is patience.

I thank Tony who long ago said that there was no reason why I couldn't complete a dissertation. You had to wait a long time, but as in most cases, you were proven right.

I thank Darryll, my friend and supervisor, for providing a flexible work environment and for being a partner in many stimulating conversations. Your support has been essential.

I thank my parents who watched me go off to graduate school but were always there to be a home. They say home is where they have to take you in. You didn't have to, but I always knew you would.

Finally, to Jane, my partner in life, we share so much, thank you for giving me the time and space to do the parts of this I needed to do alone. I love you.

DEDICATION

For my Dad.

The nut doesn't fall as far from the tree as you might have thought.

INTRODUCTION

A potential employer picks up the telephone and hears the voice of a job applicant with an African-American accent. Is the employer's impression of the applicant affected by negative evaluations or other associations triggered by the accent? Thousands of travelers walk through an airport concourse when a female voice broadcasts the need for an immediate evacuation. Would a male voice motivate quicker response due to stronger associations with authority or power? Is the female voice more effective because of associations with safety and security that reduce the likelihood of panic? An advertiser is selecting the background music for a new television commercial. Of the candidate songs, which is more highly associated with the attractive image that the customer wishes to link with the product? Could the cognitive associations that are of interest in all three of these situations act at a level that is outside of consciousness, altering the judgments of employers, travelers, and consumers without their awareness?

The Implicit Association Test (IAT; Greenwald, McGhee and Schwartz, 1998) is a recently developed measure of automatic associations that has been used to investigate research questions like those described above. For example, IAT studies have, a) examined implicit racial attitudes elicited by African-American names (e.g., Greenwald et al., 1998) and faces (e.g., Dasgupta & Greenwald, 2000; Percodani, Haines, Adames, Fisher, & Greenwald, 2000), b) investigated gender stereotypes associated with male and female names (Rudman, Greenwald, & McGhee, 2002), and c) collected marketing information in the form of implicit attitudes toward brands of juice and soda (Maison, Greenwald, & Bruin, 2001). However, existing versions of the IAT can not address the questions asked above because, a) voices and music are poorly represented by abstract stimuli such as text, so tests of their effects require the presentation of sound, and b) no IATs to date have incorporated sound stimuli. An auditory IAT would be necessary to investigate associations elicited by racial accents, male or female voices, and music.

This research demonstrates that auditory IATs can be used to measure automatic associations such as attitudes and stereotypes. Such measures are potentially useful in a wide variety of situations that require the use of sound stimuli. In addition, development and application of auditory IATs can yield insights into the nature of automatic associations and the way they are measured using the IAT method. Before considering these potential applications and other aspects of auditory IATs in more detail, it is important to understand the IAT in general.

The Implicit Association Test

The IAT can be illustrated by the following thought experiment. Imagine sorting a mixed pile of photographs into two piles. The photographs fall into two pairs of categories, those that show a single glass of red wine or lager beer, and those that show purple grapes or golden barley. You are asked to place the wine and grape photos on your left, and the beer and grain photos on your right. Your ability to quickly sort through the pile will reflect the strength of association within the two pairs of categories (i.e., the beverages and the foodstuffs) that have to be sorted together. In this example, the task will be easy because the categories on the left and right are associated in at least two ways, a) on a surface level, because the wine and grapes are shades of red/purple while the beer and barley are shades of yellow/gold, and b) on a functional level, because grapes provide a main ingredient of wine and grain provides a main ingredient of beer.

What happens if the associations cannot be used to formulate a grouping strategy? If your second task is to sort photos of wine and grain to the left, and photos of beer and grapes to the right, then the red/purple = left, yellow/gold = right strategy no longer works, and your sorting speed should deteriorate. The strategy based on ingredient/product relationships would also be counterproductive. In either case, the difference in difficulty of the two sorting tasks is an indication of the association between the beverages and foodstuffs.[1]

Put in abstract terms, the IAT is an implicit measure of the strength of association between two categories of stimuli. For example, when measuring implicit attitudes, the IAT measures the association between attitude objects and evaluative attributes. Figure 1 shows the stimuli and set of tasks that comprise an attitude IAT. As in the wine and beer

[1] Interestingly, when subjects use the ingredient/product relationship as their sorting strategy, the difference in performance between the two tasks should be largest for beer and wine connoisseurs who strongly associate grapes with wine and grain with beer. Thus, it is likely that sorting strategy would interact with connoisseur status to determine IAT results in this example.

example, the critical portion of the procedure calls for subjects to simultaneously complete two sorting tasks; one task requires them to categorize two attitude concepts

	BIRD	INSECT	PLEASANT	UNPLEASANT
IAT Items in the Four Categories	cardinal	cicada	cuddle	abuse
	warbler	locust	happy	crash
	blackbird	bee	smile	disaster
	robin	mosquito	joy	grief

	respond left	respond right
Task 1	UNPLEASANT	PLEASANT
Task 2	INSECT	BIRD
Task 3	INSECT OR UNPLEASANT	BIRD OR PLEASANT
Task 4	BIRD	INSECT
Task 5	BIRD OR UNPLEASANT	INSECT OR PLEASANT

Figure 1. Illustration of the Implicit Association Test (IAT). The IAT starts by introducing subjects to the four categories used in the task. In this example, the categories are introduced in Tasks 1 and 2. In Task 1, subjects are asked to respond 'left' to *pleasant* words and 'right' to *unpleasant* words. In Task 2, subjects respond 'left' to *insect* names and 'right' to *bird* names. The IAT measure is obtained by comparing response latencies in Task 3 and Task 5. In Task 3 *insect* and *unpleasant* are assigned to 'left' and *bird* and *pleasant* to 'right', while in Task 5 *bird* and *unpleasant* are assigned to 'left' and *insect* and *pleasant* to 'right'. (In Task 4 subjects practice responding 'left' to *bird* names and 'right' to *insect* names.) If the subject responds more rapidly when *bird* and *pleasant* share a response, this indicates that the *bird-pleasant* association is stronger than the *bird-unpleasant* association.

(e.g., birds v. insects) and the other involves categorization of two attributes (e.g., pleasant v. unpleasant). In both tasks, subjects indicate their responses by pressing pre-assigned response keys with the right or left hand. When highly associated concepts and attributes share the same key (e.g., insects or unpleasant), subjects are expected to categorize stimuli more quickly and easily than when weakly associated concepts and attributes share the same key (e.g., insects or pleasant). The difference in categorization speed for the blocks of trials in which subjects use these two concept + attribute combinations is the IAT effect, indicating relative differences in the strength of association (readers can complete several IATs at http://www.tolerance.org/hidden_bias/index.html).

The IAT is a Useful Measure of Implicit Cognitions

The IAT was developed to gain access to implicit cognitions, a domain not reached reliably by self-report measures (Greenwald, Banaji, Rudman, Farnham, Nosek, & Mellott, 2002). Implicit cognitions are thought to influence behavior "in a fashion not introspectively known by the actor" (Greenwald and Banaji, 1995, p. 4). In contrast, self-reported explicit cognitions are presumed to reflect conscious processes (see Wegner & Bargh, 1998 for a discussion of the interplay between implicit and explicit cognitions). Theory often advances after innovations in measurement techniques (Kuhn, 1970), and the IAT has provided empirical support for advancement in theories of implicit cognition (e.g., Greenwald, Banaji, et al., 2002). The IAT has also been used as a diagnostic tool in applied situations (e.g., Teachman & Woody, 2001).

A rapidly increasing body of research has generally supported the usefulness of the IAT (Greenwald & Nosek, 2001). For example, a) the IAT has been shown to be resistant to self-presentational strategies (Greenwald, Banaji, Rudman, Farnham, Nosek, & Rosier, 2000; Kim & Greenwald, 1998), b) the internal validity of the IAT has been supported by studies ruling out alternate explanations based on higher familiarity with liked versus disliked targets (Dasgupta, McGhee, Greenwald, & Banaji, 2000), c) potential threats to internal validity involving procedural variables have either been found to have no effect on IAT results (e.g., the number of stimuli used per target category or the hand subjects use to indicate pleasant or unpleasant stimuli) or to have effects that can be addressed through counterbalancing (e.g., the order of administering the two mixed-categorization tasks affects the strength of IAT effects; see Greenwald & Nosek, 2001), d) convergent validity between the IAT and affective priming measures has been supported by Cunningham, Preacher, and Banaji (2001) who found the two measures to be correlated ($r = .55$) after correcting for measurement error, e) IATs have been shown to be sensitive to known group differences in gender stereotypes (Greenwald & Farnham, 2000; Nosek, Banaji, & Greenwald, 2000; Rudman, Greenwald, & McGhee, 2002) and attitudes toward homosexuality (Banse, Seise, & Zerbes, 2001), and f) IATs measuring racial attitudes have been shown to correlate with behavior toward Black and White experimenters (McConnell & Lebold, 2001). Because they differ from existing IATs only in the modality of the stimuli presented, auditory IATs are expected to prove useful in similar ways.

Two Stimulus Modalities Have Been Used in IATs

Originally established using only text stimuli (Greenwald et al, 1998), IATs were soon reported in which the target concept was represented by pictures (e.g., Dasgupta & Greenwald, 2000). No published works have directly addressed the question of whether IATs using picture and text stimuli measure the same constructs. However, in an unpublished study, Swanson and Greenwald (1997) found correlations between word and picture IATs ($r=.70$ for meat/vegetable-pleasant/unpleasant word and picture IATs; $r=.60$ for meat/vegetable-me/not-me word and picture IATs) that were comparable to test/retest correlations commonly reported for identical IATs (Greenwald & Nosek, 2001). These results, in conjunction with the continued widespread use of picture IATs (e.g., Dasgupta & Greenwald, 2000, Percodani et al., 2000, Mitchell, Nosek, & Banaji, 2002, Swanson, 2001) suggest that IAT effects can be readily obtained whether the stimulus modality is text or pictures. The apparent success of this extension in modality suggests that auditory stimuli can also be used to create a successful auditory IAT.

Auditory Implicit Association Tests

Humans utilize sounds either as a means of monitoring the environment or as communication (Kroemer & Grandjean, 1997). Environmental-sounds and speech correspond to these functions. Hereafter, environmental-sounds will be referred to simply as *sounds* and the term *auditory stimuli* will be used to refer inclusively to speech and environmental-sounds. Two forms of auditory IAT, a sound IAT (in which environmental sounds are sorted) and speech IAT (in which spoken words are sorted) were developed in this research.

Potential Applications of the Sound IAT

Many situations require the presentation of sound stimuli for psychological measurement because they involve constructs or attitude objects that cannot be well represented by text or verbal description and are best represented by sounds. Simple examples are bird songs, with textual descriptions such as "kli quiquiquiqui koo" or "pidi pew pew pedrrrrrr" (Sibley, 2000) that fail to capture reality. An example with more important societal implications involves speech. African-American and Euro-American voices are consistently identifiable by untrained participants based on accents not evident in the semantic content of the words (Purnell, Idsardi, & Baugh, 1999; Walton & Orlikoff, 1994). Such identification can lead to racial discrimination, as in studies showing that Euro-Americans more frequently turned down telephone requests for assistance that were made by African-Americans than those made by Euro-Americans (Gaertner, 1973; Katz, Cohen, & Glass, 1975), and in more recent studies showing that landlords discriminate against prospective tenants on the basis of racial dialect (Purnell et al., 1999). An IAT using sound stimuli (hereafter, sound IAT) could be used to investigate implicit cognitions associated with racial accents, or cognitions associated with other constructs that are best represented by sounds, such as music, gender (as indicated by vocal tone and pitch), or noise from sources such as aircraft or automobile traffic.

Potential Applications of the Speech IAT

Auditory stimuli in the form of speech are necessary for psychological measurement when constructs not readily represented by pictorial stimuli but readily presented as words (e.g., loyalty, honesty) must be presented to subjects who cannot read, such as young children. An IAT using speech stimuli might be constructed by simply replacing the text in a standard IAT with spoken words. Such speech IATs could prove useful in a variety of studies, including investigations of the early development of automatic associations. Speech IATs would also be necessary for studies involving blind subjects. Again, a variety of studies might be conducted in this situation, including studies of automatic associations with colors in subjects who can or cannot see color

(e.g., is the association of red and warm dependent on the experience of seeing the wavelength of light called red?)

Conclusion

The IAT is a promising new measure of implicit cognitions. Using auditory stimuli in IATs will allow measurement of automatic associations in a variety of situations where the IAT method could not otherwise be employed.

Overview of the Present Research

Three experiments sought to assess the ability of sound and speech IATs to measure automatic associations. Results were expected to confirm that auditory IATs measure differences in evaluative and stereotypic associations between pairs of semantic or social categories. Results were also intended to support preliminary assessment of the convergent validity between IATs incorporating sound, speech, picture, and text stimuli and to provide an opportunity for exploring the nature of automatic associations and the way they are measured using the IAT method.

Study 1 used sound and speech IATs, as well as text and picture IATs to measure implicit attitudes toward target concepts that were expected to be highly similar across persons (i.e., insects = negative and birds = positive). In Study 2 men and women completed sound, speech, and text IATs measuring gender stereotypes that were assumed to differ by sex. Study 3 used an accent-based sound IAT and a picture IAT to measure personally or socially undesirable evaluative associations in the form of implicit attitudes toward African-American and European-American racial categories.

STUDY 1

Study 1 used auditory and visual IATs to assess implicit attitudes toward a pair of target attitude concepts for which subjects were expected to have relatively uniform evaluative associations. The attitudinally positive concept was birds, and the negative concept was insects. The auditory IATs included a speech IAT in which spoken names of birds and insects were categorized, and a sound IAT in which the sounds made by birds and insects were categorized. The visual IATs were a text IAT and a picture IAT in which written names and photographs of birds and insects (respectively) were categorized. In all the IATs the target-concept categorization task was used in combination with a task requiring categorization of pleasant and unpleasant attributes (in all cases, presented in text). All IAT procedures were expected to show superior performance for response-key pairings that were compatible in evaluation (bird + pleasant) than for non-compatible combinations (insect + pleasant).

Method

Subjects

Ninety University of Washington undergraduates received course credit for participating. All reported that their vision and hearing were normal or corrected-to-normal and that they were fluent English speakers.

Stimuli

The study's four computerized categorization tasks used text, computer-generated speech, sounds, and pictures as stimuli. Text stimuli included four pleasant-meaning words (cuddle, happy, smile, joy) and four unpleasant-meaning words (abuse, crash, disaster, grief) selected from stimulus sets used by Greenwald, et al. (1998). Text stimuli also included the common names of the same four songbirds (blackbird, cardinal, robin, warbler) and four insects (bee, cicada, locust, mosquito) characterized by the speech, picture, and sound stimuli.

The speech stimuli were computer generated by a publicly available program (http://www.bell-labs.com/project/tts/voices.html) that accepted typed input and returned a digital sound file (.wav format). The default voice ("Man") was selected. A single file was obtained by providing all the names as input, with words separated by two commas to create a substantial pause between them. Dummy words were included at the beginning and end of the list to eliminate the possibility that the program would alter the inflection for those positions. Individual stimulus files were excerpted from the original such that there was no silence preceding sound onset. Durations ranged from 320ms (bee) to 615ms (cicada).

Sound stimuli were obtained from http://Naturesongs.com (Van Gausig, 2000) and were used with permission. Digital sound files (.wav format) were edited so that peak volumes were comparable and so that continuous sounds (e.g., the mosquito buzz) faded to silence in the last 100ms of the recording rather than ending abruptly. Sounds ranged from 1200ms (warbler) to 2300ms (redwing blackbird) in length.

Picture stimuli were obtained from public sources on the web, including http://www.mbr-pwrc.usgs.gov/id/framlst/infocenter.html, and http://www.insects.org/entophiles/. Images were cropped so that each songbird or insect filled the image frame without extending beyond the image boundary. Thus, birds and insects were not to scale. All stimuli were converted to gray-scale,[2] sized at 180x180 pixels (yielding 58mm x 58mm images at the 800 x 600 pixel screen resolution used), and saved as Windows bitmap (.bmp) files (see Appendix A for stimuli).

Explicit measures were paper-and-pencil tasks adapted from measures used in prior IAT research (cf., Swanson, Rudman, & Greenwald, 2001; see Appendix B). The attitude thermometer was a 100-point scale presented graphically as a vertical thermometer. The "0" point was labeled "Cold, or unfavorable", "50" was labeled "Neutral", and "99" was labeled "Warm, or favorable". Subjects were asked to make a mark on the scale indicating "...how you feel about BIRDS in general" (the same type of rating for INSECTS was marked on a second thermometer). On the semantic differential measure, subjects rated BIRDS and INSECTS on five seven-point scales (Ugly/Beautiful, Bad/Good, Unpleasant/Pleasant, Harmful/Harmless, Awful/Nice).

[2] Gray-scale was used to eliminate variability in the pleasantness of the stimuli based solely on the colors present in the photographs.

Apparatus and Program

Study one was administered on desktop computers using the Windows operating system to run Inquisit software (Draine, 1998). Subjects viewed the display from about 65 cm. They gave responses with the left and right forefingers, using the "a" and ";" keys, respectively. Subjects wore lightweight stereo headphones. White-noise generators were used to mask extraneous sound. Each computer station was located in an approximately 2 m by 3.5 m experimental cubicle.

Procedure

Prior to subjects' arrival, the experimenter readied the individual cubicles by adjusting the headphone volumes to be clearly audible but not uncomfortable and by setting the white-noise generators to a volume that made normal conversations outside the cubicle inaudible. Upon entering the lab, each subject was assigned to an individual cubicle and seated in front of a monitor. The first task was completion of the explicit measures, after which subjects were instructed to wear the headphones and were asked to follow the instructions presented on the computer monitor. Doors to the individual cubicles were then closed.

The computer-administered portion of the study consisted of three IATs preceded by associated blocks of practice trials. The IATs completed by each subject were selected permutations of the four IATs taken three at a time, with permutations balanced across 12 conditions such that each IAT appeared equally often in the first, second, or third position. Subjects first did a 20-trial practice block of single-category distinctions for the pleasant and unpleasant text stimuli. The text, picture, and auditory stimuli to be used in the following IATs were then presented. Text and picture stimuli were shown on a single computer screen with the category labels above. Speech and sound stimuli were played in sequence through the headphones while the category labels and names of the respective birds and insects were shown on the monitor. Subjects then completed three blocks of 16 single-category distinctions for birds and insects; one set for each type of stimuli to be used in the three following IATs. Response key pairing of BIRD and INSECT categories (i.e., placement of category labels on the right or left side of the monitor) alternated between blocks such that the third block pairing was opposite the pairing to be used in the first block of the combined-task. Finally, subjects completed each IAT in the following sequence: 1) a block of 16 trials to practice the combined-task; 2) a block of 40 test trials; 3) a block of 16 trials to practice the combined-task after switching the BIRD and INSECT response keys); and 4) a block of 40 test trials in the new configuration. Initial placement of the BIRD and INSECT categories on the right or left was balanced across subjects.

Each trial began with the presentation of the target stimulus (the appearance of a word or picture in the center of the monitor or the generation of a target sound) and ended with a correct response. The subsequent trial began after a 400 millisecond inter-trial pause. Category labels were shown in the upper-left and upper-right portion of the monitor. Stimulus presentation was halted when a categorizing response was made. If a target was incorrectly categorized, a red "X" appeared in the lower-center of the monitor to indicate an error and the subject was obliged to give the correct response to continue. Response times were recorded from stimulus onset to correct response.

Only the relevant category labels were presented during the single-category distinctions. During the combined-task all the category labels were arranged appropriately on the left or right. The center of the screen remained blank when speech or sound stimuli were presented.

In the combined-task test blocks, the four types of trials were presented in a specific order such that strings of more than two correct key-presses on the same key did not occur. The order also specified that consecutive strings of more than two Bird/Insect or Pleasant/Unpleasant did not occur.[3] Practice blocks used random presentation orders.

Results and Discussion

Data Preparation

Preparation of the IAT data followed the recommendations of Greenwald, Nosek, and Banaji (2002) who used large datasets obtained using web-based IATs to evaluate

[3] This presentation order was adapted from procedures used in earlier IAT studies (e.g., Swanson et al., 2001; Rudman et al., 2001). Stronger IAT effects are obtained using alternating concept and attribute trials with random stimulus selection (A. G. Greenwald, personal communication, November 1, 2001) and such procedures were used in Study 2 and 3.

many methods of scoring the IAT. They sought to improve upon the performance of conventional IAT scoring (see Greenwald et al., 1998) in meeting desirable criteria related to internal consistency, correlations with self-report measures, sensitivity to known influences, and resistance to extraneous influences. Based on the scoring procedure they recommended, trials with response latencies longer than 10,000 ms were excluded from analyses in this study. Other aspects of the recommended scoring procedure are described below in the section concerning implicit measures.

Implicit and explicit data from one subject were excluded from analyses because two of the three IATs he completed had very fast latencies consistent with flippant response (i.e., latencies less than 300 ms for more than ten percent of trials in the combined task blocks).[4] Due to experimenter error, explicit measures were not collected for two subjects.

Error Rates and Latencies: Subjects' Ability to Categorize Auditory Stimuli

A concern in designing the auditory IATs was that the stimuli might be difficult to categorize and that such difficulty could interfere with the IATs ability to detect automatic associations. The ease with which subjects categorized the auditory stimuli is best indicated by the speed and accuracy of the practice blocks in which they completed single-category distinctions for birds and insects.

Error rates. The mean error rates during the single-category distinctions for birds and insects were low, ranging from 3.4% for pictures, 3.6% for sounds, 4.9% for text, and 5.5% for speech. Error rates for auditory stimuli were not significantly higher than those for visual stimuli. The largest such difference observed was between pictures and speech, $t(41) = 1.45, p = .154; d = .22$.[5]

Latencies. The single-category distinctions for birds and insects took longer for auditory stimuli than text or pictures. Latencies averaged 569 ms for pictures, 668 ms for text, 773 ms for sound, and 864 ms for speech. Latencies for all four modalities differed

[4] Greenwald, Nosek, et al. (2002) tested a variety of exclusion rules and determined that this rule improved the performance of IAT measures in relation to a variety of criteria.

[5] Paired-comparison tests included approximately half the sample because each subject completed only three of the four IATs.

significantly. The smallest statistical difference was observed between speech and sound stimuli, $t(38) = 5.00, p = 10^{-4}; d = .80$.

Latencies for the text and picture stimuli may have been faster because all the features necessary for categorization were present at stimulus onset. In contrast, the auditory features necessary to categorize auditory stimuli (i.e., sound-wave patterns) play out over small but potentially important periods of time. In particular, it may have been impossible to categorize speech stimuli until one or more full syllables of the spoken words were pronounced (after durations as long as several hundred ms). Thus, longer latencies are not a sufficient basis to conclude that auditory stimuli were more difficult to categorize.

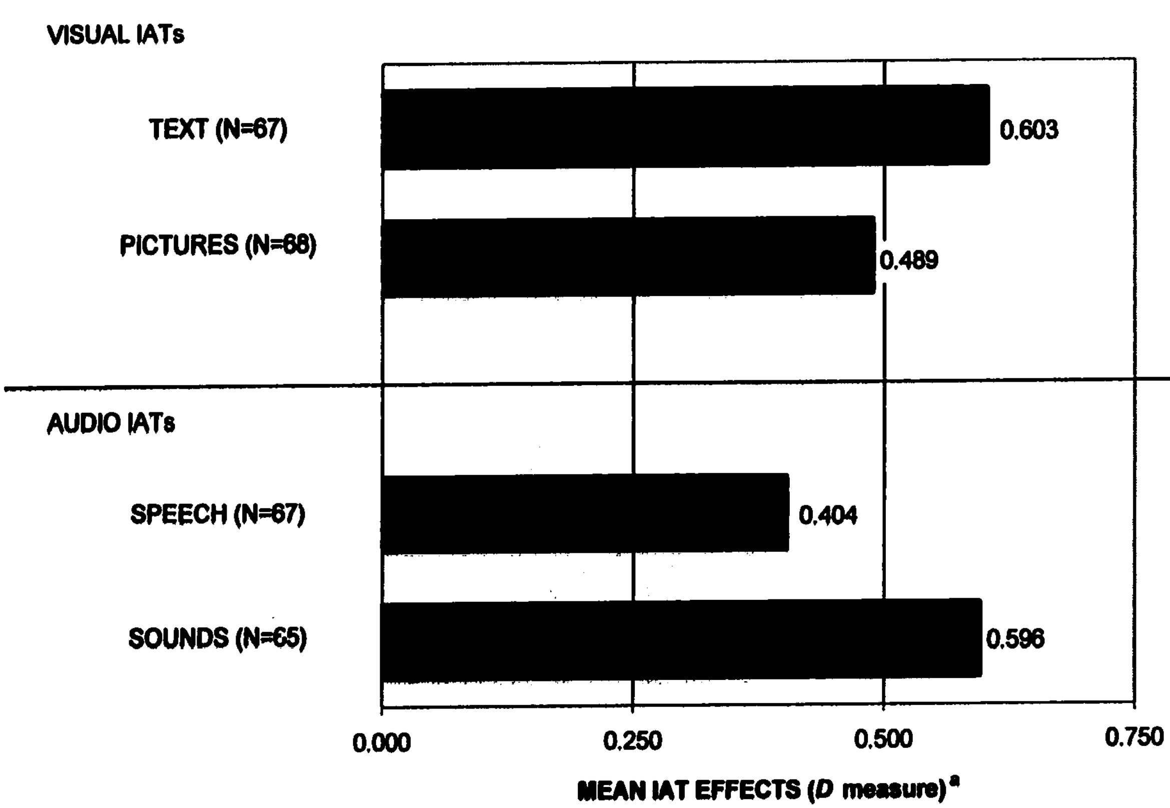
The mix of error and latency data suggests that categorizing the auditory stimuli was sufficiently easy to allow measurement of automatic associations using auditory IATs. The error rates were low, and satisfactory IAT effects have been reported using stimuli with categorization latencies longer than the 864 ms needed to categorize speech stimuli (Swanson, 2001a).

Implicit and Explicit Measures of Attitudes Toward Birds and Insects

Implicit measures. A measure of the IAT effect (i.e., the *D* measure) recently proposed by Greenwald, Nosek, et al., (2002) was calculated for the IATs used in this research. When applied to these IAT procedures, there are two primary differences between the *D* measure and the scoring method originally proposed by Greenwald et al. (1998): 1) practice-block data from the combined-task are used in addition to the test-block data, and 2) individual-respondent standard deviations are used as the measure's scale unit. In Study 1, the *D* measure was calculated by: a) subtracting the mean latency of the practice block for the Pleasant-or-Bird vs. Unpleasant-or-Insect task from the mean latency of the practice block for the Unpleasant-or-Bird vs. Pleasant-or-Insect task, b) dividing the difference calculated in a) by the pooled standard deviation of the latencies in both practice blocks, c) repeating step a) and b) for the corresponding test blocks, and d) calculating the average of the practice-block and test-block measures. The *D* measure can not be readily converted to a difference in response latency. However, analyses conducted by Greenwald, Nosek, et al. found that it exhibited better performance than the conventional scoring method on a variety of criteria including correlations with explicit measures, internal consistency, sensitivity to known influences, and resistance to undesired influences.

It was expected that the IATs would measure implicit attitudes favoring birds over insects. The results shown in Figure 2 are consistent with that expectation. Mean IAT effects for all four implicit measures were significantly larger than zero (as a result of faster response in the Pleasant-or-Bird vs. Unpleasant-or-Insect categorization task than in the Unpleasant-or-Bird vs. Pleasant-or-Insect task), and effect sizes were large, ranging from $d = 1.07$ for the speech IAT, to $d = 1.62$ for the text IAT. All auditory and visual IATs detected differences in association of the concepts *Bird* and *Insect* with valence (i.e., differences in attitudes) that were almost universal in the population.

Statistical comparisons between the four IAT effects showed that the effect measured by the speech IAT ($D = .404$) was significantly smaller than the text IAT ($D = .603$; $t(44) = 2.46$, $p = .018$), and the sound IAT ($D = .596$; $t(42) = 2.62$, $p = .012$). However, the effects measured by the sound IAT and the visual IATs did not differ significantly, and none of the observed differences alter the conclusion that all the IATs detected strong differences in implicit attitudes.



	SD	IAT Effect[a] p^b	Cohen's d^c
TEXT (N=67)	.371	10^{-19}	1.62
PICTURES (N=68)	.413	10^{-11}	1.18
SPEECH (N=67)	.379	10^{-13}	1.07
SOUNDS (N=65)	.386	10^{-18}	1.55

Figure 2. Summary Statistics for Implicit Attitude Measures (Study 1). [a]Mean IAT effects are represented in terms of the *D* measure (Greenwald, Nosek, & Banaji, 2002). All Bird/Insect IATs were scored so more positive scores indicate more positive attitudes toward birds than toward insects. [b]*p*-values correspond to *t*-tests of each IAT measure against 0. [c]Effect sizes, *d*, were computed by dividing mean IAT scores (represented as the *D* measure) by their standard deviations. Conventional small, medium, and large effects for *d* are .2, .5, and .8, respectively.

Procedural variables in the IATs. IAT effects (i.e., implicit attitudes toward birds vs. insects) were significantly different for the speech IAT depending on whether the Pleasant-or-Bird vs. Unpleasant-or-Insect task was the first or second block of critical trials completed (Pleasant-or-Bird first, $D = .214$, Pleasant-or-Bird second, $D = .577$), $t(65) = 4.45, p = 10^{-4}$. Although "pairing-order effects" have been common in prior IAT research (Greenwald & Nosek, 2001), the observed difference was opposite that commonly reported. It is not entirely clear why the pairing order effect should be reversed for the speech IAT. The number of combined-task practice trials was somewhat smaller than in other IAT studies (cf., Greenwald et al., 1998; Greenwald & Farnham, 2000), and the order of the practice blocks for each IAT completed in Study 1 differed from the five-step procedure used in most IATs (Greenwald & Farnham, 2000). However, without replication, the most likely explanation of this unusual result is type 1 error. In any case, the speech IAT showed the expected implicit attitudes favoring birds over insects even when the Pleasant-or-Bird pairing was completed first, $t(31) = 4.25$, $p = 10^{-4}$, $d = .75$.

Regression equations were used to test the linear trend in IAT effects based on the ordinal position of each IAT in the study. The IAT effects measured by the sound, text, and picture IATs were smaller the later that they were completed in the study (sound IAT when 1st $D = .749$, 2nd $D = .587$, 3rd $D = .450$; $F(1,63) = 8.00, p = .006$; text IAT when 1st $D = .724$, 2nd $D = .604$, 3rd $D = .450$; $F(1,65) = 6.19, p = .015$; picture IAT when 1st $D = .645$, 2nd $D = .464$, 3rd $D = .385$; $F(1,66) = 4.44, p = .039$).[6] The linear trend for the speech IAT was not significant (when 1st $D = .509$, 2nd $D = .329$, 3rd $D = .384$), $F(1,65) = 1.14, p = .291$ but showed some evidence of a decreasing pattern.[7] Studies using multiple IATs have previously noted such reductions in effect size for later IAT completion (Swanson 2001a). Such completion-order effects might be explained by factors ranging

[6] The completion order effects for the text and picture IATs were moderated by two-way interactions with pairing-order (when the interaction was added to the regression equation for text and picture IATs, *F* Change (1, 63) = 6.00, p = .017; *F* Change (1, 64) = 4.80, p = .032; respectively). In both cases, IAT effects decreased with later completion only when the bird+pleasant pairing was completed first. It is not clear why such interactions should be observed. Because they were found only in the visual IATs these results will not be discussed further.

[7] Tests for differences in the linear trends between IATs were precluded by the design of the study.

from subject fatigue to practice effects that increase subjects' skill. Apparently the sound IAT is susceptible to such effects in much the same way as visual IATs. However, even when completed last, the sound IAT showed the expected implicit attitudes favoring birds over insects, $t(23) = 6.83$, $p = 10^{-6}$, $d = 1.39$.

The analyses of procedural variables unexpectedly showed that none of the IATs exhibited the pairing-order effect commonly observed in prior research and that the speech IAT actually showed a significant effect in the opposite direction. If replicated, such an effect might have implications for the internal validity of the speech IAT as compared to text or picture IATs. However, even if it proved reliable, such an effect could be addressed by counterbalancing (in the same manner the more commonly observed effect is addressed). Analyses of IAT completion order showed that when completed after other IATs the effects observed in sound IATs were diminished. However, this limitation was shared by visual IATs (in this research and in previous studies, see Greenwald & Nosek, 2001) and is of little concern in most studies. In general, the current findings do not suggest important differences between the internal validity of auditory and visual IATs.

Explicit measures. As expected, attitudes were more favorable toward birds than insects on all explicit measures. Attitude thermometer ratings were 65.2 for birds and 35.0 for insects, $t(87) = 12.2$; $p = 1.5^{-20}$; $d = 1.30$. The semantic differential scales averaged 5.3 for birds and 3.3 for insects, $t(87) = 14.7$; $p = 2.6^{-25}$; $d = 1.56$.

Convergent Validity of IATs and Explicit Measures

Intercorrelations of IATs and explicit measures. Four inferences are suggested by the correlations between the various attitude measures. The correlations in Table 1 are divided into four boxes corresponding to those inferences. First, the text, picture, and speech IATs were moderately correlated (see box A in Table 1). Given that the uniformity in implicit attitudes toward birds and insects may have limited the inter-correlations through a restriction of variability, the average correlation of .45 among these three different IATs compares favorably to the test-retest correlations commonly reported for identical IATs (i.e., above r = .6; Greenwald & Nosek, 2001). In addition, the correlations were larger than a cross-method correlation reported by Cunningham, et al., (2001) who found that a Picture IAT and a Response-Window IAT (identical except that responses were constrained to a specific latency range) had an average correlation of .27. These data support the convergent validity of the speech IAT.

Table 1. Correlations between IATs and Explicit Measures of bird vs. insect attitudes.

		TEXT IAT	PICTURE IAT	SPEECH IAT	SOUND IAT	THERMOMETER	SEMANTIC DIFF.
TEXT IAT	r	1					
	N	67					
PICTURE IAT	r	.46**	1				
	N	46	68				
SPEECH IAT	r	.47**	.42** A	1			
	N	45	46	67			
SOUND IAT	r	.22	.24	.36* B	1		
	N	43	44	43	65		
THERM.	r	.26*	.13	.11	-.09 C	1	
	N	65	67	65	64	88	
SEM. DIFF.	r	.21	.17	.18	.06	.63** D	1
	N	65	67	65	64	88	88

* Correlation is significant at $p < .05$.
** Correlation is significant at $p < .01$.

Second, the sound IAT was weakly to moderately correlated with the other implicit measures (see box B in Table 1). Only the correlation between the sound and speech IATs was significantly greater than zero. However, the correlations with the sound IAT were not significantly lower than the correlations among the other implicit measures (i.e., those in box A). Given the restricted variability in the measured implicit attitudes, the positive trend of these correlations suggested some convergent validity between the results of the sound IAT and the other implicit measures, but were also consistent with the possibility that the sound IAT was measuring a different set of automatic associations.

Finally, the correlations between explicit and implicit measures were weak (see Table 1, box C), while the explicit measures were correlated strongly with each other (see Table 1, box D). The variable and generally weak correlations are consistent with the low explicit/IAT correlations reported when implicit attitudes toward target concepts are expected to be highly similar across persons (cf., Greenwald, et al., 1998). The restriction in variability associated with such near-universal attitude measures is likely to be responsible for the lack of correlation. Thus, the low correlations between explicit and implicit measures, in particular, the low correlations with the auditory IAT measures, should not be interpreted as evidence against their convergent validity.

Summary of Study 1.

The study demonstrated that auditory IATs, like visual IATs, could measure attitudes that were distributed nearly universally in the population (i.e., attitudes favoring birds over insects). Subjects could readily categorize auditory stimuli, and the combined-task was more difficult in the attitudinally inconsistent pairing condition (i.e., when bird and unpleasant were paired on the left response key). Pairing-order had an unexpected effect on the speech IAT, raising some question about the internal validity of that auditory IAT. However, the effects of the other procedural variable, IAT completion-order, were no greater for auditory than for visual IATs.

The restricted variability in attitudes toward birds and insects and low statistical power in comparisons of correlations limited the ability to draw conclusions about convergent validity based on the correlations among the implicit and explicit measures. The speech IAT was moderately correlated with the text and picture IAT, supporting its convergent validity. The sound IAT was weakly to moderately correlated with the other implicit measures. It was possible that a combination of restricted variability and measurement error masked evidence of convergent validity, or the sound IAT may have measured a different set of automatic associations. The implicit measures were weakly (if at all) correlated with the explicit measures. The variability in the observed correlations is congruent with past research showing variable correlations between: a) implicit and explicit measures (Greenwald & Nosek, 2001), b) IATs using different stimuli (e.g., Percodani et al., 2000; Swanson & Greenwald, 1997), and c) IATs using different methods (Cunningham et al., 2001).

STUDY 2

Study 2 sought to extend the auditory IAT method to the domain of gender stereotypes through a conceptual replication of a study reported by Rudman et al. (2002). They used a Gender-Potency IAT in which subjects categorized names as male or female and words as powerful or weak to show strong implicit gender stereotypes for male subjects but no evidence of such stereotypes for females. Two similar auditory IATs were used in Study 2 to discriminate sex differences in gender stereotypes -- a sound IAT in which spoken words were categorized as male or female based on the gender of the speaker, and a speech IAT in which names spoken by a gender-neutral voice were categorized as male or female. In addition, subjects completed a text IAT in which the same names were categorized by gender, as well as explicit measures of gender stereotypes. Based on Rudman et al. (2002), male subjects completing the sound, speech and text IATs were expected to find the Male-or-Low Power vs. Female-or-High Power combined task to be more difficult than the Female-or-Low Power vs. Male-or-High Power task. In contrast, women were expected to find the tasks similar in difficulty.

Method

Subjects

Ninety University of Washington undergraduates received course credit for participating in this study. All reported that their vision and hearing were normal or corrected-to-normal. All but one subject reported that they were fluent in English. The exception had been speaking English less than five years and his data were excluded from analyses.

Stimuli

The study's four categorization tasks used text and spoken words as stimuli. Text stimuli included four high-power words (command, confident, dominant, strong) and four low-power words (follow, submissive, timid, weak), and also four male names (ERIC, JASON, PETER, STEVE)[8] and four female names (DIANNE, KAREN, LISA,

---

[8] Attributes were presented in lower-case and names in upper-case in order to provide subjects with an additional cue concerning the type of categorization required.

SANDRA). All text stimuli were selected from stimulus sets used by Rudman et al. (2002).

The speech stimuli were computer generated by a publicly available program (http://www.research.att.com/~ttsweb/cgi-bin/ttsdemo) that accepted typed input and returned a digital sound file (.wav format). Two files were obtained for use in the sound IAT, one spoken by the program's male voice and one by the female voice. For each, a list of four neutral words associated with measurement (center, gallon, halfway, inch; selected from stimulus sets used by McGhee (2002) was provided as input, with the words separated by two commas to create a substantial pause between them. Dummy words were included at the beginning and end of the list to eliminate the possibility that the program would alter the inflection for those positions. Individual stimulus files were excerpted from the two original files such that there was no silence preceding sound onset. Durations ranged from 347 ms (inch) to 586 ms (halfway).

Two files were also obtained for use in the speech IAT, one for high-power and low-power words, and one for male and female names (the same stimuli used in the text IAT). Both files were generated by the program's female voice but subsequently manipulated to make the voice gender uncertain -- a sound-editing program was used to lower the pitch of the sounds while retaining their durations. Individual stimulus files were cut from the two original files such that there was no silence preceding sound onset. Durations ranged from 283 ms (weak) to 686 ms (confident) and from 390 ms (John) to 595 ms (Sandra).

Explicit measures were presented using MediaLab software (Empirisoft, 2001). Measures included stereotype thermometers and semantic-differential scales adapted from measures used in prior IAT research (Swanson et al., 2001; see Appendix B). The stereotype thermometer was a 100-point scale presented graphically as a vertical thermometer. The "0" point was labeled "Low Power", "50" was labeled "neutral", and "100" was labeled "High Power". Subjects were asked to enter a number from the scale indicating "...how you perceive MEN" (the same type of rating for WOMEN was entered for a second thermometer). On the semantic differential measure, subjects rated

MEN and WOMEN on four seven-point scales corresponding to the high-power and low-power words used in the text IAT and speech IAT (Strong/Weak, Confident/Timid, Dominant/Submissive, Command/Follow).

Apparatus and Program

Study 2 was administered using the same experimental cubicles and apparatus described for Study 1.

Procedure

Experimental procedures prior to the computer-administered portion of the study were the same as in Study 1 except that explicit measures were not completed at that time.

The first computer-administered portion of the study consisted of the sound IAT, the speech IAT, the text IAT, and a repeat of the sound IAT. The sound IATs were always completed first and last to maximize temporal separation for test-retest purposes. The order of the speech IAT and text IAT was balanced across subjects.

Blocks of single-categorization trials intended to familiarize subjects with the various stimuli were completed first. After initial instructions, the sound, speech, and text stimuli representing the concepts *male* and *female* in the implicit measures were presented. Text stimuli were shown on a single computer screen with the category labels shown above. Speech and sound stimuli were played through the headphones while the words being spoken and category labels were shown on the monitor. After initial presentation, subjects' completed three blocks of 16 single-category distinctions; one set for each type of stimuli. The speech stimuli to be categorized as high-power or low-power were presented after the concept practice. These attribute stimuli were played through the headphones while the words being spoken and category labels were shown on the monitor. Subjects subsequently completed a 16-trial practice block of single-category distinctions for the high-power and low-power speech stimuli. Finally, the high-power and low-power text stimuli were shown on a single computer screen with the category label above, and then subjects completed a 16-trial practice block of single-category distinctions for the high-power and low-power text stimuli.

Subjects completed each IAT in the following sequence: 1) a block of 20 trials to practice the combined-task; 2) a block of 40 test trials; 3) a block of 20 trials to practice the combined-task with the Male and Female categories on the opposite sides of the screen (i.e., switching the associated response keys); and 4) a block of 40 test trials in the new configuration. The Female and Low Power categories were placed on the left in the initial trial blocks for all subjects.[9]

Presentation of stimuli and category labels within trials, as well as recording of responses were the same as in Study 1, except that a 300 ms inter-trial pause was used to create a smoother transition from one trial to the next.

In all combined-task blocks, trials alternated between the FEMALE/MALE and low power/high power tasks. Otherwise, trial orders in practice and test blocks were randomized.

In the final portion of the study, subjects completed computer-administered explicit measures in the following order: 1) Male Semantic Differential Scales, 2) Female Semantic Differential Scales, 3) Male and Female power thermometers. The explicit measures are described in more detail above (also see Appendix B). Responses were made using the computer keyboard or mouse.

Results and Discussion

Data Preparation

The improved method of scoring the IAT used in Study 1 was employed in this study (see Greenwald, Nosek, et al., 2002). As in Study 1, trials with response latencies longer than 10,000 ms were excluded from analyses .

Implicit and explicit data from one subject were excluded from analyses based on the same exclusion rule used in Study 1 (i.e., latencies less than 300 ms for more than ten percent of trials in the combined task blocks). Because two subjects incorrectly reported

[9] Initial placement of the concept categories is commonly counterbalanced to control for the pairing-order effect observed in many IAT studies (see Greenwald & Nosek, 2001). Pairing orders in which the more highly associated concepts and attributes are paired first generally yield larger IAT effects. Pairing order was not counter balanced in this study because it was deemed desirable to maximize IAT effects in order to facilitate discrimination between males and females on the basis of their implicit gender stereotypes.

that they had finished the experiment after completing the IATs, their data do not include explicit measures.

Error Rates and Latencies: Subjects' Ability to Categorize Auditory Stimuli

Error rates. The ease with which subjects categorized the auditory stimuli is best indicated by the speed and accuracy of the practice blocks in which they completed single-category distinctions between male and female stimuli. The mean error rates during those blocks were low and did not differ statistically. Error rates were 2.1 percent for sounds, 2.7 percent for text, and 5.5 percent for speech.

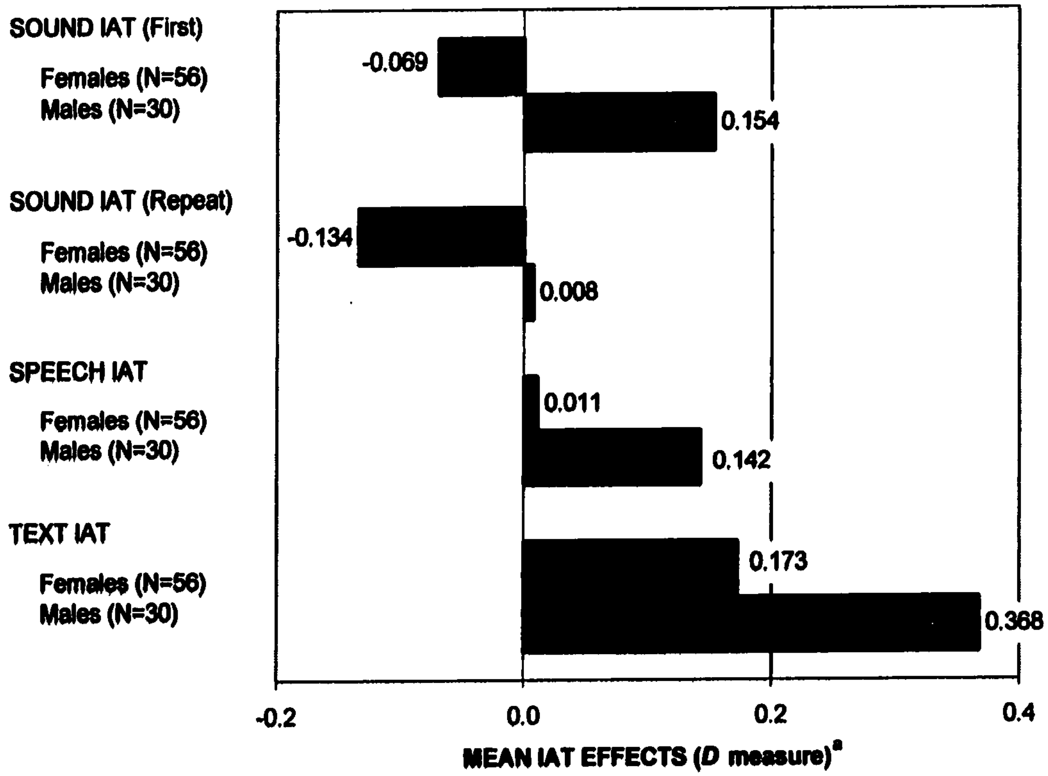
Latencies. The single-category distinctions between male and female stimuli took longer for auditory stimuli than for text. Latencies averaged 573 ms for text, 697 ms for sounds, and 811 ms for speech. Latencies for all three modalities differed significantly. The smallest statistical difference was observed between sound and speech stimuli, $t(80) = 5.00, p = 10^{-5}; d = .56$.

Both the pattern of error and latency results were similar to Study 1. The data suggest that categorizing the auditory stimuli was sufficiently easy to allow measurement of automatic associations using auditory IATs.

Implicit and Explicit Measures of Gender Stereotypes Concerning Power

Implicit measures. As in Study 1, the *D* measure proposed by Greenwald, Nosek, et al., (2002) was calculated and used to represent the IAT effect for the IATs used in Study 2.

Figure 3 presents Study 2's results separately for male and female subjects. The sound IATs and text IAT showed that men exhibited stronger implicit gender stereotypes than women (i.e., superior performance for male-or-high power vs. female-or-low power response-key pairings than for female-or-high power vs. male-or-low power combinations). The sex difference was only marginally significant for the speech IAT, but a repeated measures ANOVA found no evidence that the sex difference observed in the speech IAT differed from that observed in the other IATs, $F(3, 252) = .481, p = .70$. All four effect sizes were small (see Figure 3).



	IAT Effect[a]			Sex Difference	
	SD	p^b	Cohen's d^c	p^d	Cohen's d^c
SOUND IAT (First) Females (N=56)	.388	.130	.18	.005	.32
SOUND IAT (First) Males (N=30)	.338	.019	.46		
SOUND IAT (Repeat) Females (N=56)	.266	10^{-4}	.50	.031	.24
SOUND IAT (Repeat) Males (N=30)	.319	.898	.03		
SPEECH IAT Females (N=56)	.341	.810	.03	.078	.19
SPEECH IAT Males (N=30)	.293	.013	.48		
TEXT IAT Females (N=56)	.317	10^{-4}	.55	.005	.31
TEXT IAT Males (N=30)	.264	10^{-7}	1.39		

Figure 3. Summary Statistics for Implicit Gender Stereotype Measures (Study 2). [a] Mean IAT effects are represented in terms of the *D* measure (Greenwald, Nosek, & Banaji, 2002). All implicit measures were scored so more positive scores indicate greater association of high-power with males than high-power with females. [b] *p*-values correspond to one-sample *t*-tests of the IAT measure against 0. [c] Effect sizes, *d*, were computed by dividing mean group difference scores by their pooled standard deviations. Conventional small, medium, and large effects for *d* are .2, .5, and .8, respectively. [d] *p*-values correspond to *t*-tests of the differences between group means.

Contrary to expectations, a repeated measures ANOVA found significant differences between the magnitude of the four gender-stereotype IAT effects averaged across all subjects, $F(3, 261) = 21.25, p = 10^{-11}$. All but the first sound IAT and speech IAT (mean sound IAT effect = .010; mean speech IAT effect = .055; $t(87) = 1.10, p = .27; d = .12$) differed significantly. The smallest statistical difference was observed between the first and repeat sound IATs (.010, -.091), $t(87) = 2.17, p = .03; d = .12$. The largest was observed between the text IAT (.237) and the repeat sound IAT (-.091), $t(87) = 8.02, p = 10^{-10}; d = .86$.

The implications of the difference in IAT effects across the four IATs are clarified by comparing whether men and women exhibited gender stereotypes in each IAT. The text IAT results suggest that both men and women held stereotypic automatic associations between gender and power (i.e., they exhibited significantly better performance for the gender stereotypic response-key pairing; see Figure 3), while the results of the repeat sound IAT (completed only a few minutes later) suggest that men were stereotypically neutral while women exhibited counter-stereotypic automatic associations (see Figure3). This pattern of results could have resulted if the name stimuli representing the male and female concept categories in the text and speech IATs elicited associations with power for males and females in general -- associations that were stereotypic in nature. In contrast, the fact that only one male and one female computer-generated voice (each speaking four different words) were categorized in the sound IAT may have led the sound IATs to measure associations with those specific individuals. Subjectively assessed, the female computer-generated voice featured clear enunciation and a strong tone of voice that sounded rather confident and assertive, while the male voice featured a relatively relaxed delivery. Although these aspects of the stimuli were independent of the target category distinction (i.e., male vs. female), they may have nonetheless lead the sound IATs to measure different associations than were measured in the speech and text IATs.

Procedural variables in the IATs. Completion order of the speech IAT and text IAT had no effect on the results of the speech or sound IAT results. The text IAT effect was larger when it was completed third than when completed second (.303 vs. .174), although the difference fell just short of conventional statistical significance, $t(86) = 1.97$; $p = .052$; $d = .21$. This effect was opposite the usual effect of IAT order in which the latter of multiple IATs show diminished effects. However, it is logically consistent with the above argument that the female voice used in the sound IAT was perceived as unusually high in power. When the text IAT (the most abstract of the IATs) was completed immediately after the sound IAT, the context of the high-power female voice may have lingered. In contrast, when completion of the speech IAT intervened, conventional stereotypes may have been more strongly elicited by the text stimuli.

Explicit measures. Subjects reported strong stereotypic associations between gender and power. Stereotype thermometer ratings were 73.2 for MEN and 57.7 for WOMEN, $t(86) = 9.64$; $p = 10^{-14}$; $d = 1.03$. The semantic differential scales averaged 5.3 for MEN and 4.2 for WOMEN, $t(86) = 9.98$; $p = 10^{-15}$; $d = 1.07$.
Men reported stronger stereotypic associations on the semantic differential scales (5.5 for MEN and 3.9 for WOMEN) than did women (5.3 for MEN and 4.4 for WOMEN), $t(85) = 3.26$; $p = .002$; $d = .35$, but significant sex differences were not observed for the stereotype thermometers (MEN – WOMEN ratings differences were 18.3 points for male subjects and 13.9 for female subjects) , $t(85) = 1.27$; $p = .208$; $d = .14$. Thus, sex differences for the explicit and implicit measures were roughly comparable – they had small effect sizes, some of which did not differ significantly.

<u>Convergent Validity of IATs and Explicit Measures</u>

Intercorrelations of IATs and explicit measures. Four inferences are suggested by the correlations among the various attitude measures. The correlations in Table 2 are divided into four boxes corresponding to those inferences. First, the test-retest correlation for the sound IAT failed to reach statistical significance (see box A in Table 2). This low correlation was unexpected, particularly because individual variability was expected (and observed) in the gender stereotypic automatic associations that were measured (in contrast to Study 1 in which the limited variability in implicit attitudes was expected to limit correlations). It is unclear why the sound IAT should show lower test-retest reliability than the correlations greater than .6 commonly observed in other IATs (Greenwald & Nosek, 2001). Perhaps the completion of the speech and text IATs in the interval separating the two sound IATs introduced contextual factors that altered the results of the second IAT -- recall that the results of the two IATs differed significantly.

Table 2. Correlations between IATs and explicit measures of gender stereotypes.

		SOUND IAT	REPEAT SOUND IAT	SPEECH IAT	TEXT IAT	THERMOMETER	SEMANTIC DIFF.
SOUND IAT	r	1					
	N	88					
REPEAT SOUND IAT	r	.10 A	1				
	N	88	88				
SPEECH IAT	r	.38**	.09 B	1			
	N	88	88	88			
TEXT IAT	r	.30**	.19	.43**	1		
	N	88	88	88	88		
THERM.	r	.07	.08	.04	-.01 C	1	
	N	86	86	86	86	86	
SEM. DIFF.	r	.13	.14	.06	.16	.68** D	1
	N	86	86	86	86	86	86

* Correlation is significant at $p < .05$.
** Correlation is significant at $p < .01$.

The low test-retest correlation might also be attributed, at least in part, to high levels of measurement error. Cunningham et al. (2001) had subjects repeat an implicit prejudice IAT in four sessions at two-week intervals and found an average test-retest correlation of only .27. Nonetheless, when analyzed using a latent variable approach (Tisak & Tisak, 2000) the proportion of variance in the IATs that remained stable over time was calculated to be .68. The *D* measure used as a scoring procedure in this study can be broken into two measures of implicit gender stereotypes for each sound IAT (one from practice blocks and one from test blocks). A principal components factor analysis of those four measures revealed two factors with eigenvalues greater than 1. After varimax rotation, the two measures from the first sound IAT had strong loadings on the first factor and the two measures calculated from the second sound IAT had strong loadings on the second factor (see Table 3). Although the interpretation of the second factor is unclear, its presence suggests that the low test-retest correlation arises from more than measurement error.

The second inference suggested by the correlations in Table 2 was that the speech, text, and first sound IAT were moderately correlated, while the repeat sound IAT was not correlated with either the text or sound IAT (see box B in Table 2). Although the correlations were not as large as some of those observed between the attitude IATs in Study 1, the relationships among the speech, text, and first sound IAT support the convergent validity of the auditory IATs. The very low correlations between the repeat sound IAT and the speech or text IATs suggest that the results of the repeat sound IAT may be responsible for the low test-retest correlation discussed above. As in Study 1, the strongest correlation was observed between the speech and text IATs. This was expected but notable because in this study all the stimuli used in the speech IAT were auditory, whereas only the target category stimuli were auditory in Study 1.

Table 3. Rotated factor matrix from principal components factor analysis of four sound IAT measures in Study 2 (varimax rotation). Eigenvalues were 1.57 and 1.15.

	FACTOR	
	1	2
FIRST SOUND IAT - PRACTICE BLOCKS	.858	-.040
FIRST SOUND IAT - TEST BLOCKS	.819	.158
SECOND SOUND IAT - PRACTICE BLOCKS	-.110	.832
SECOND SOUND IAT - TEST BLOCKS	.239	.722

Finally, the correlations between explicit and implicit measures were very weak (see Table 2, box C), while the explicit measures were correlated strongly with each other (see Table 2, box D). One might expect such low implicit/explicit correlations if presentation biases led subjects to respond to the explicit measures so as to minimize socially undesirable stereotypic beliefs, but the effect sizes showed stronger stereotypic response in the explicit measures (the smallest effect was $d = 1.03$ for the stereotype thermometer) than in the implicit measures (the largest effect was $d = .94$ for the text IAT).

The very weak explicit/implicit correlations in Study 2 raise the possibility that the explicit and implicit measures were not measuring the same constructs (i.e., they do not support convergent validity). This is somewhat unexpected. Rudman et al (2002) observed a moderate correlation ($r = .29$) between their Gender-Potency IAT and a semantic differential scale much like that used in Study 2 (the analogous correlation in Table 2 was .16). The earlier-reported findings that the auditory IATs and explicit measures showed comparable sensitivity to male vs. female differences in this known-groups comparison suggest that more information would be necessary to judge the relative merits of the explicit and implicit measures used in Study 2.

<u>Summary of Study 2</u>

The results of Study 2 both replicated and diverged from those reported by Rudman et al. (2002). The results of the text IAT (the IAT most similar to that used by Rudman et al.) replicated the sex difference found in their Gender-Potency IAT, although contrary to their findings, female subjects in Study 2 showed significant gender stereotypes. Also consistent with Rudman et al., the results demonstrated that the sound IAT could discriminate differences in gender stereotypes held by male and female subjects.

Unexpectedly, the sound IAT showed low test-retest reliability (see Table 2), and correlations between all implicit and explicit measures were very weak. Also unexpectedly, the magnitude of implicit gender stereotypes measured by the sound, speech, and text IATs differed significantly. At their most extreme, these differences meant that within a few minutes, female subjects displayed significant implicit gender stereotypes as measured by the text IAT and significant contradictory counter-stereotypic associations between gender and power as measured by the repeat sound IAT. These findings suggest that characteristics of IAT stimuli other than their category membership may have acted within, and even across IATs to affect test results.

STUDY 3

Study 3 examined whether an auditory IAT could measure personally and/or socially undesirable implicit racial attitudes that were elicited by acoustic cues in the spoken words of African Americans and European Americans. Such attitudes might not be readily detected through explicit self-report measures, and could be elicited in any verbal interaction. Subjects completed two IATs; an accent IAT in which spoken phrases were categorized based on the race of the speaker, and a face IAT in which pictures of faces were categorized based on race. In both IATs the race categorization task was used in combination with a task requiring categorization of pleasant and unpleasant attributes (presented in text). Consistent with prior research showing that implicit racial attitudes are not always consciously expressed (Crosby, Bromley, & Saxe, 1980; Devine, 1989; Fazio, Jackson, Dunton, & Williams, 1995; Gaertner & McLaughlin, 1983; Greenwald & Banaji, 1995; Greenwald et al., 1998; Wittenbrink, Judd, & Park, 1997), the study was expected to show that both the accent and face IATs would measure automatic expressions of race-related attitudes among subjects who consciously disavowed such attitudes on explicit measures.

Method

Subjects

Eighty University of Washington undergraduates received course credit for participating in this study. All reported that their vision and hearing were normal or corrected-to-normal and that they were fluent English speakers. Fifty-five subjects (68.8%) reported that they were White, 18 (22.5%) were Asian, two (2.5%) were African-American, and five (6.3%) reported other ethnicities.

Stimuli

The study's three categorization tasks used text, face pictures, and spoken words as stimuli. Text stimuli included six pleasant-meaning words (gift, health, joy, laughter, rainbow, warmth) and six unpleasant-meaning words (agony, cancer, failure, poison, sickness, vomit) selected from stimulus sets used by Greenwald, et al., (1998). Face pictures were selected from those used in Cunningham, et al. (2001). They included three African American (AA) and three European American (EA) men. Face pictures showed only a rectangular area including the eyes, nose, and mouth, excluding secondary features such as hair and clothing (see Appendix A for stimuli).

Spoken stimuli were greetings (e.g., "Hey, how you doin'?" see Appendix A). Voices included four AA and four EA men. Digital files (.wav format) of AA voices were recorded from radio talk-show files available on the internet. The sampled program focused on African-American issues in Chicago (http://www.wvon.com/home.htm). EA voices were recorded by telephoning male friends of the experimenter, each of whom voluntarily repeated the same greetings as the African Americans with similar levels of warmth and enthusiasm. For each speaker, one greeting was selected that matched a corresponding AA greeting in phrasing and warmth while ensuring that four different EA speakers were represented.

Explicit measures were presented using MediaLab software (Empirisoft, 2001). Measures included the Modern Racism Scale (McConahay, Hardee, & Batts, 1981), two scales (diversity and discrimination) developed by Wittenbrink et al. (1997), as well as attitude thermometers and semantic-differential measures (see Appendix B). The attitude thermometers were identical to those used in Study 1 except subjects were asked to enter a number from the scale indicating "...how you feel about African-Americans in general" (the same type of rating for European-Americans was entered for the second thermometer). On the semantic differential measure, subjects rated African-Americans and European-Americans on five seven-point scales (Ugly/Beautiful, Bad/Good, Unpleasant/Pleasant, Harmful/Harmless, Awful/Nice).

Apparatus and Program

Study 3 was administered using the same experimental cubicles and apparatus described for Studies 1 and 2.

Procedure

Experimental procedures prior to the computer-administered portion of the study were the same as in Study 1 except that explicit measures were not completed at that time.

The first computer-administered portion of the study consisted of two IATs (the order of which was balanced across subjects). Subjects first did 24 practice trials of single-category distinctions for the pleasant and unpleasant text stimuli. If the accent IAT was to be administered first, all the sound stimuli were then played in sequence through the headphones while the "European American" and "African American" category labels, as well as item labels (e.g., "AA Voice 1" and "EA Voice 1") were shown on the monitor. If the accent IAT was second, all the sound stimuli were presented after completion of the face IAT. Photo stimuli for the face IAT were not presented prior to the single-category practice trials using those stimuli. Subjects completed each IAT in the following sequence: 1) a block of 24 single-category distinctions for EA and AA stimuli (i.e., faces or voices); 2) a block of 24 trials to practice the mixed-categorization task; 3) a block of 40 test trials; 4) a block of 24 single-category distinctions for AA and EA stimuli with the categories on the opposite sides of the screen (i.e., switching the associated response keys); 5) a block of 24 trials to practice the mixed-categorization task with the new category configuration; and 6) a block of 40 test trials in the new configuration. Initial placement of the EA and AA categories on the right or left was balanced across subjects.

Presentation of stimuli and category labels within trials, as well as recording of responses were the same as in Study 1 and 2, except that a 150 ms inter-trial pause was used.

In all mixed-categorization blocks, trials alternated between the EA/AA and Pleasant/Unpleasant tasks. Otherwise, trial orders in practice and test blocks were randomized.

In the final portion of the study, subjects completed computer-administered explicit measures in the following order: 1) AA Semantic Differential Scales; 2) EA Semantic Differential Scales; 3) Modern Racism Scale; 4) Diversity Scale; 5) Discrimination Scale; and 6) the AA and EA attitude thermometers.

Results and Discussion

Data Preparation

The improved method of scoring the IAT used in Study 1 and 2 was employed in this study (see Greenwald, Nosek, et al., 2002). As before, trials with response latencies longer than 10,000 ms were excluded from analyses .

Explicit measures are missing for one subject who incorrectly reported finishing the experiment after completing the IATs but before completing the full experiment.

Error Rates and Latencies: Subjects' Ability to Categorize Auditory Stimuli

Error rates. The ease with which subjects categorized the auditory stimuli is best indicated by the speed and accuracy of the practice blocks in which they completed single-category distinctions between AA and EA stimuli. The mean error rates during those blocks were significantly higher for voices (8.1%) than for faces (2.6%), $t(76) = 7.57, p = 10^{-10}; d = .86$.

Closer examination of the voice and face practice blocks showed that some voices were significantly more difficult to classify than others. Error rates among the EA voices were 1 = 3.4%, 2 = 2.2%, 3 = 25.1%, and 4 = 3.1% (3 had higher error rates than each of the other voices, all $ts(76) > 7.18, ps < 10^{-9}; ds > .81$). Error rates among the corresponding AA voices were 1 = 8.8%, 2 = 13.2%, 3 = 5.9%, and 4 = 3.2% (2 was higher than 3 and 4 $ts(76) > 3.58, ps < .002; ds > .40$, and 1 was higher than 4, $t(76) = 2.85, p = .006; d = .32$). Error rates were particularly high for trials involving the third EA and second AA voice stimuli ("How ya doin', man?" and "Hey, good morning."). However, analyses of the mixed-categorization trials (i.e., the IAT results reported below) that included or excluded those trials showed no differences that would alter the interpretation of the results. Therefore, the results reported below represent all trials.[10]

Latencies. The single-category distinctions between AA and EA stimuli took longer for voices than for faces. Latencies averaged 1052 ms for voices and 618 ms for faces, $t(76) = 26.25, p = 10^{-40}; d = 2.99$. Because latencies were measured until subjects gave the correct response, the mean difference partially reflects the higher error rates for

---

[10] Analyses excluding the third EA and AA voice stimuli ("How ya doin', man?") were also consistent with analyses of the complete dataset.

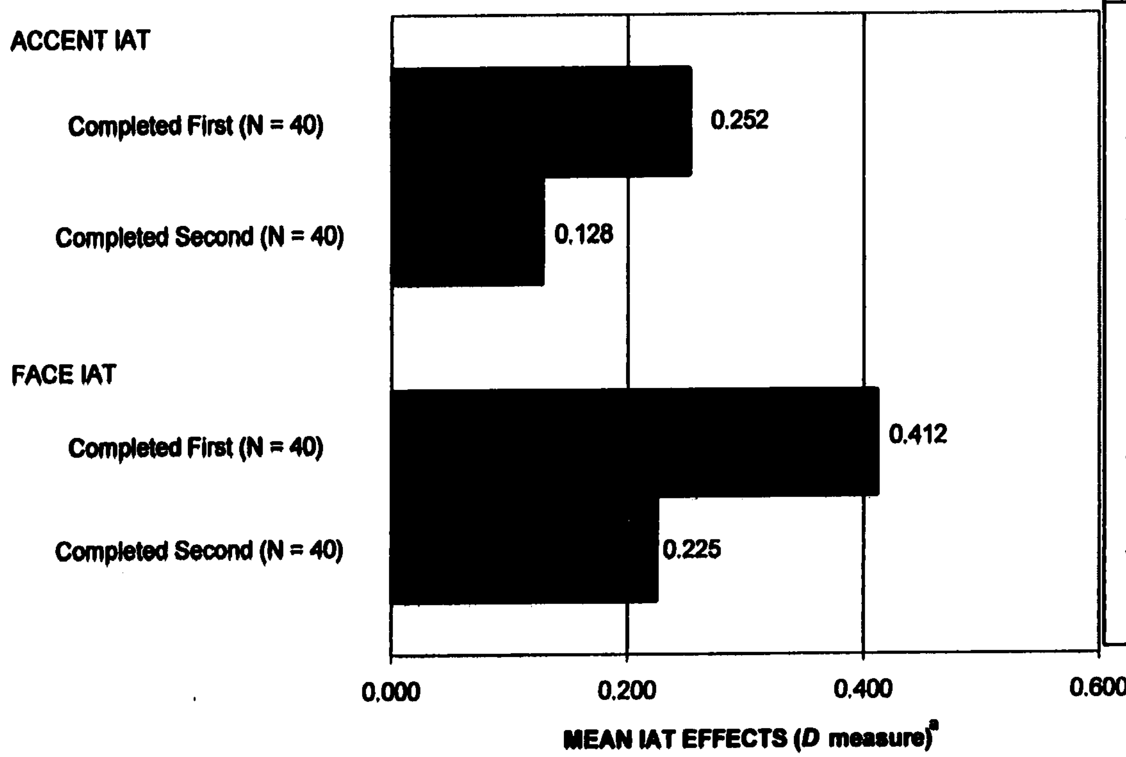
voices. When only correct trials are considered, mean latencies were still longer for voices (979 ms) than for faces (602 ms), $t(76) = 25.82, p = 10^{-39}, d = 2.98$.

The voice stimuli were clearly more difficult to categorize than the face stimuli, probably because the EA and AA voices in the accent IAT spoke the same greetings, forcing a distinction based on vocal accent alone. EA and AA dialects may also differ in other respects. Spontaneous contextual speech is imbued with specific phonological, morphological, lexical, and/or syntactic information that can be used to identify racial dialect (Walton & Orlikoff, 1994). However, the finding that exclusion from IAT analyses of the stimuli that were most difficult to categorize did not alter interpretation of the study results suggests that the observed level of difficulty in categorizing the auditory stimuli did not interfere with the IATs' ability to detect automatic associations.

Implicit and Explicit Measures of Racial Attitudes

Implicit measures. The same method of calculating IAT effects used in Study 1 and 2 was used in Study 3 (see Greenwald, Nosek, et al., 2002). Results are reported below in terms of the *D* measure resulting from that scoring procedure.

Figure 4 presents Study 3's results. The expected IAT effects were observed. More specifically, performance in both implicit measures was better in the Pleasant-or-EA vs. Unpleasant-or-AA categorization task than in the Unpleasant-or- EA vs. Pleasant-or- AA task. Effect sizes were large for the face IAT ($d = .82$) and medium for the accent IAT ($d = .52$). Thus, both IATs detected differences in association of the concepts African American and European American with valence (i.e., differences in implicit racial attitudes).



	IAT EFFECT [a]			COMPLETION ORDER	
	SD	p^b	Cohen's d^c	p^d	Cohen's d^c
ACCENT IAT: Completed First (N = 40)	.377	10^{-3}	.67	.129	.17
ACCENT IAT: Completed Second (N = 40)	.346	.025	.37		
FACE IAT: Completed First (N = 40)	.405	10^{-6}	1.02	.030	.25
FACE IAT: Completed Second (N = 40)	.347	10^{-3}	.65		

Figure 4. Summary Statistics for Implicit Racial Attitude Measures (Study 3). [a] Mean IAT effects are represented in terms of the *D* measure (Greenwald, Nosek, & Banaji, 2002). All implicit measures were scored so more positive scores indicate greater association of pleasant with European Americans than pleasant with African Americans. [b] *p*-values correspond to one-sample *t*-tests of the IAT metric against 0. [c] Effect sizes, *d*, were computed by dividing mean completion order differences by the pooled standard deviations. Conventional small, medium, and large effects for *d* are .2, .5, and .8, respectively. [d] *p*-values correspond to *t*-tests of the IAT differences by completion order.

Statistical comparisons between the IAT effects showed that the difference in critical block latencies measured by the accent IAT was significantly smaller than the face IAT, $t(79) = 2.34$, $p = .022$, $d = .26$. This result suggests that the auditory stimuli used in the accent IAT elicited slightly weaker stereotypic evaluative associations than did the picture stimuli. Although stimuli for the accent IAT were selected to be similar in emotional warmth, all the greetings were warm and friendly in tone. This characteristic may have been somewhat inconsistent with a negative stereotype for AA targets, and thus may have decreased the magnitude of the effects observed by the accent IAT.

Procedural variables in the IATs. ANOVAs were used to test whether IAT completion order or pairing order affected the results of the accent or face IATs. Accent IAT effects were not significantly affected by completion order. In contrast, face IAT effects were stronger when the face IAT was completed first than when completed second (see Figure 4; $F(1, 78) = 8.92$, $p = .004$, for the interaction of IAT type and completion order). Three of the IATs completed in Study 1 also showed such a reduction in the IAT effect with later completion, as have other studies using multiple IATs (see Swanson, 2001a). Such completion-order effects might be explained by factors ranging from subject fatigue to practice effects that increase subjects' skill. However, a possibility suggested by other analyses of Study 3 data is that when the face IAT was completed second, the context of *friendly* AA and EA targets represented in the accent IAT may have remained salient, decreasing the difference between the evaluative associations with each class of targets and thus decreasing the size of the observed IAT effect.

Accent IAT effects were significantly larger when the Pleasant-or-AA vs. Unpleasant-or-EA task was completed first rather than second (.358 vs. .022), $t(78) = 4.62$, $p = 10^{-4}$, $d = .525$. This pairing-order effect was opposite that commonly reported in prior IAT research (see Greenwald & Nosek, 2001) -- IAT effects are usually larger when the attitudinally-consistent pairing is completed first. Such a result was surprising, but similar to that observed for the speech IAT in Study 1. Face IAT effects were not significantly affected by pairing order, $F(1, 78) = 9.81$, $p = .002$, for the interaction of

IAT type and pairing order in a repeated measures ANOVA. It is not entirely clear why the use of auditory stimuli should reverse the pairing-order effect. However, with the observation of a similar result in Study 1, the possibility of type 1 error is remote.

Summarizing the effects of procedural variables, the analyses of completion order suggest that the accent and face IATs have internal validity that is comparable and acceptable, but the analyses of pairing order are more puzzling. Pairing-order effects have been observed in many IAT studies and are generally thought to constitute a minor threat to internal validity (see Greenwald & Nosek, 2001). However, the accent IAT showed an effect opposite that commonly observed. The implications of this finding for the internal validity of the accent IAT are unclear.

Explicit measures. Average racial attitudes measured by the thermometers were more favorable toward European Americans than African Americans (78 vs. 72), $t(78) = 4.08$; $p = 10^{-6}$; $d = .46$. However, the mean semantic differential ratings did not differ (EA = 2.84, AA = 2.97), and responses indicative of prejudice were rare on the Modern Racism Scale (McConahay et al., 1981), Diversity Scale, and Discrimination Scale (Wittenbrink et al., 1997). Across all subjects, responses to the three scales (scored from 1 to 5 with a neutral point at 3) indicated disagreement with statements indicative of prejudicial beliefs (MRS = 2.01, Div. = 2.70, Disc. = 2.37). Relatively few subjects (1 for the MRS, 15 for the Div., and 16 for the Disc.) had mean responses to the scales greater than 3.0 (i.e., responses indicating more agreement than disagreement with the statements). Thus, with the exception of the attitude thermometers, the explicit measures suggested that as a group, subjects in Study 3 did not favor European Americans over African Americans. These findings are inconsistent with the results of the face and accent IATs. Such discrepancies are commonly reported in studies of racial attitudes (cf. Greenwald et al., 1998; Blair et al., 2001) and are thought to arise because IATs are sensitive to automatic associations that subjects are either unaware of making or hesitant to express on explicit measures because of personal or social undesirability

Convergent Validity of IATs and Explicit Measures

Intercorrelations of IATs and explicit measures. Four inferences are suggested by the correlations between the various attitude measures. The correlations in Table 4 are divided into three boxes corresponding to those inferences. First, there was a very weak correlation between the accent and face IATs (see box A in Table 4). It was expected that the IATs would be correlated, and further examination of this relationship showed that the relationship was significantly stronger ($r = .50$) when the accent IAT was completed first than when the face IAT was completed first ($r = -.07$), $z = 2.63$; $p = .009$ (test of correlations for independent samples, see Hays, 1988).[11] Analogous comparisons of the correlations between the IATs completed in Study 1 provided a replication of this order effect. The correlation between the sound IAT and text IAT was .58 when the sound IAT was completed first, but -.20 when the text IAT was completed first, $z = 2.52$; $p = .012$. Similarly, the correlation between the sound and speech IAT was .55 when the sound IAT was completed first, but .25 when the text IAT was completed first; a difference that was not statistically significant, $z = 1.10$; $p = .271$, but shows the same pattern. Such effects of completion order might result if the stimuli used in the sound IAT had extraneous information that made a particular context salient and thus altered the automatic associations measured by later IATs using relatively generic stimuli. The general discussion expands upon this explanation of the way auditory and visual IAT results may be altered by the order in which the measures are completed. In general, however, the conditional nature of the correlation between the accent and face IATs provides equivocal evidence for the convergent validity of the measures.

Second, the correlations between the implicit measures and the explicit measures were very weak or nonexistent (see Table 4, box B), even though the explicit measures

[11] Data from one subject (number 43) had an inordinate effect on the correlation between the face and accent IATs. Number 43 had the largest face IAT effect (2.57 standard deviations above the mean) and the next-to-smallest accent IAT effect (2.11 standard deviations below the mean). When the face IAT data were regressed on the accent IAT, number 43 had a leverage 4.5 times the mean leverage. Leverage values twice as large as the mean are considered to indicate outlying cases (Neter, Wasserman, & Kutner, 1990). Because the pattern of number 43's responses did not indicate erroneous or flippant completion of the tasks, her data were not excluded from all analyses. However, computation of the correlation between the face and accent IATs without her data lifted the overall correlation to $r = .24$ and the correlation for the face IAT-first completion order to $r = .06$.

were moderately to strongly correlated with each other (see Table 4, box C). These results were expected, given evidence that implicit and explicit measures of racial attitudes are weakly correlated (Crosby et al., 1980; Devine, 1989; Fazio et al., 1995; Gaertner & McLaughlin, 1983, Greenwald & Banaji, 1995; Wittenbrink, et al., 1997) and prior reports of comparable correlation between race IATs and the same explicit measures used in Study 3 (cf. Greenwald et al., 1998).

Table 4. Correlations between IATs and explicit measures of racial attitudes.

		ACCENT IAT	FACE IAT	THERM.	SEMANTIC DIFF.	MODERN RACISM	DISCRIM.	DIVERSITY
ACCENT IAT	r	1						
	N	77						
FACE IAT	r	.15 A	1					
	N	80	80					
THERM.	r	.19	.19 B	1				
	N	79	79	79				
SEMANTIC DIFF.	r	.14	.13	.67** C	1			
	N	79	79	79	79			
MODERN RACISM	r	-.12	.04	.18	.30**	1		
	N	79	79	79	79	79		
DISCRIM.	r	.08	.08	.40**	.31**	.48**	1	
	N	79	79	79	79	79	79	
DIVERSITY	r	.17	-.16	.09	.03	.15	.25*	1
	N	79	79	79	79	79	79	79

* Correlation is significant at $p < .05$.
** Correlation is significant at $p < .01$.

Summary of Study 3

The study demonstrated that an accent IAT, like the face IAT, and in contrast to most of the administered explicit measures, measured implicit racial attitudes that were personally and/or socially undesirable.

The convergent validity of the accent and face IATs, as indicated by their correlation, was moderated by the order in which the IATs were completed. The correlation was indicative of convergent validity when the accent IAT was completed first, but the IAT results were not correlated for the other completion order. As in Study 2, these results suggested that for the accent IAT, aspects of the auditory stimuli other than their category membership elicited attitudes toward a specific subset of the target category, rather than attitudes toward the category in general.

As expected, correlations between explicit and implicit measures ranged from weak to nonexistent. Coupled with the findings that of the explicit measures, only the thermometer measure indicated that subjects favored AA over EA targets, these data show that the IATs measured implicit racial attitudes that were consciously disavowed by the subjects who displayed them.

GENERAL DISCUSSION

In the article initially describing the IAT, Greenwald et al. (1998) conclude that their three studies confirm the usefulness of the IAT as a measure of automatic associations. In a similar manner, the three studies in this research confirm that auditory IATs (either a sound IAT in which sorted stimuli were environmental sounds or a speech IAT in which stimuli were spoken words) measured automatic associations. Study 1 showed that implicit attitudes toward a pair of target attitude concepts for which subjects had relatively uniform evaluative associations (birds and insects) were measured by: a) a sound IAT using stimuli consisting of the sounds made by birds and insects, and b) by a speech IAT using stimuli consisting of the spoken names of birds and insects. Study 2 showed that a sound IAT in which male and female voices were sorted measured sex differences in implicit gender stereotypes (the corresponding results were marginally significant for a speech IAT in which spoken names of men and women were sorted). Finally, Study 3 showed that a sound IAT in which African-American and European-American voices were sorted measured personally and/or socially undesirable implicit racial attitudes that were not detected through explicit self-report measures.

Although auditory stimuli in all cases took longer to categorize than did visual stimuli, categorization was sufficiently easy to support the IAT method. Also, differences in the results of auditory IATs due to procedural variables (i.e., the order in which the response-key pairings were completed, and the completion order when multiple IATs were administered) were generally comparable to those found in prior research (see Greenwald & Nosek, 2001; Swanson, 2001a).

The development of auditory IATs opens the way for studies of automatic associations that could not otherwise be conducted. Sound IATs could be used to measure associations with constructs or attitude objects that are best represented by sounds and cannot be well represented by text or verbal description. Study 3 examined racial accents, an example of one such construct that has societal importance as a source of discrimination or prejudice. Other constructs that require the presentation of sound stimuli include aircraft or automobile noise. The effects of such environmental noises are poorly understood. Thus, measurement of their impacts and the development of techniques for mitigating those impacts could benefit from the application of more sophisticated tools such as the IAT (Staples, 1996).

Speech IATs also have the potential to widen the applicability of the IAT method. The use of speech stimuli may provide the means to translate text IATs for subjects who are blind or too young to read, thus making them eligible for current IAT studies. In addition, speech IATs open up the possibility to study hypotheses concerning the particular automatic associations held by the populations that would otherwise be unable to participate. For example, hypotheses concerning the development of automatic associations in young children might be examined.

Speech IATs will serve most effectively as translations of text IATs if the tests are functionally equivalent. Consistent with such functional equivalence, the highest correlations observed in Study 1 and 2 were found between the results of the speech and text IATs (Study 3 did not include a speech IAT). However, the correlation in Study 1 does not represent the relationship between a text IAT and completely auditory speech IAT because the attributes in that study's speech IAT were represented by text stimuli. The speech IAT in Study 2 was completely auditory, but its correlation with the text IAT was somewhat lower ($r = .371$) than correlations commonly observed in test/retest analyses (i.e., > .6; Greenwald & Nosek, 2001). Thus, the results provide limited support for the functional equivalence of the speech and text IATs. Further studies are necessary before concluding that the IATs have sufficiently strong convergent validity to be considered interchangeable.

Auditory IATs and the Interpretation of IAT Results

Sophisticated interpretation of IAT results (whether they use sound, speech, text, or picture stimuli) depends on an understanding of how the test works and the factors that affect its results. However, research on alternative theoretical interpretations has not progressed sufficiently to establish a widely accepted theoretical interpretation of the IAT effect (Greenwald & Nosek, 2001). The present research provided an opportunity for a variety of comparisons among the results of several sound, speech, picture, and text

IATs. This preliminary assessment of convergent validity suggested that IAT results were affected by, a) the stimuli selected to represent the target concepts, and b) factors outside the IAT procedure creating a context in which the test was completed. The evidence for each of these effects is discussed in more detail below.

Effects of Stimulus Selection in Sound IATs

Greenwald and Nosek (2001) concluded that existing IATs are generally insensitive to characteristics of the stimulus items as long as the items represent the categories well and their extraneous characteristics are not confounded with the category contrasts. However, several studies have found that the stimuli selected to represent categories can affect IAT results. Mitchell et al. (2002) found that subjects showed no racial preference when disliked politicians represented whites and liked athletes represented blacks but that subjects showed preferences for whites when unknown targets were sorted. Even more dramatically, De Houwer, (2002) administered an IAT in which IAT effects for individual stimuli were consistent with the stimulus valence rather than the valence of the target concept. De Houwer suggested that his unusual results occurred because the target concepts (person and animal) were weakly associated with the attributes (positive and negative) while the individual stimuli (e.g., FRIEND, LIAR, SWAN, SPIDER) had strong positive and negative characteristics that were extraneous to their category membership. Mitchell et al.'s findings could likewise have resulted because the target concepts (black and white) had relatively weak associations with the attributes compared to the individual stimuli (e.g., Michael Jordan, Jesse Helms).

Although De Houwer's explanation focuses on situations in which targets are weakly associated with attributes, it clearly specifies that IAT results can be affected by associations other than the targeted association between the attributes and the category membership of each stimulus when those extraneous associations are stronger than those that are targeted. There are two reasons why the results of sound IATs are more likely than IATs using text or pictures to reflect such effects of stimulus features. First, sounds such as voices, animal calls, or musical phrases represent the target category because they are examples of some characteristic aspect of the category's members. In contrast, speech, text, and pictures are all, to a greater or lesser degree, symbols. For instance, hearing a recorded robin's song is in almost all respects the same experience as an auditory encounter with an actual robin, while the word "robin" is a symbol that has no inherent similarity to the bird. The verisimilitude of sound stimuli is likely to increase the degree to which they elicit extraneous associations that could affect IAT results.

The second reason why sound IAT results might be particularly affected by stimulus selection is closely related to the first. Unlike text or picture stimuli, it is difficult to select or modify sound stimuli to reduce or eliminate extraneous characteristics. Consider the peacock, a visually beautiful exemplar of a generally positive class of animal. The call of a peacock is readily classifiable but few would find it pleasant, and the stimulus cannot be altered to make it more neutral without making it unrecognizable. In contrast, although picture stimuli (particularly photographs) share the problem of verisimilitude associated with sounds, many of their extraneous features are readily attenuated. A picture of a peacock could be shown in gray-scale, or reduced to a symbolic outline (e.g., the NBC television network logo). In a more relevant example, facial photographs used in Study 3 and several other IAT studies of racial attitudes (e.g., Dasgupta & Greenwald, 2000) were presented in gray-scale, carefully cropped to eliminate extraneous features such as hairstyle and clothing, and balanced for brightness.[12]

Although this research was not specifically designed to test whether the nature of sound stimuli might make the results of sound IATs particularly dependent on the stimuli selected to represent the target categories, several aspects of the observed results suggest that they are. In both Study 2 and 3, the magnitude of the sound IAT effects differed from other IATs, and post hoc examination suggested that the characteristics of the sound stimuli used in each of those studies produced the differences. Specifically, the assertiveness of the computerized female voice appeared to weaken gender stereotypes observed in Study 2, and the friendliness of all the recorded greetings in Study 3 appeared

[12] Picture IATs can use stimuli that carry extraneous information analogous to that carried by sound stimuli. Thus, sound IATs may not always be more susceptible to stimulus effects. However, extraneous features can be more readily removed from pictures than sounds, and picture IATs have used stimuli modified for that purpose.

to elicit attitudes toward friendly African-Americans and European-Americans that were less discrepant than attitudes toward AA and EA persons in general.

Patterns in the correlation results were also consistent with stimulus effects in the sound IATs. In Study 3, when the racial-attitude IAT using pictures of faces with neutral expressions (i.e., the face IAT) was completed first its results were uncorrelated with the corresponding sound IAT in which friendly voices were sorted (i.e., the accent IAT), even though the category labels and attribute stimuli were identical. Similarly, when the text IAT (using bird and insect names) was completed first in Study 1, its results were uncorrelated with the sound IAT in which the sounds made by those birds and insects were sorted. In both situations, the stimuli included in the sound IATs apparently elicited automatic associations that were different from the associations elicited by the stimuli included in the face or text IATs. Although the lack of correlation does not necessarily indicate that the sound IAT was sensitive to stimulus characteristics, the differences between the stimuli neatly explain how the IATs could produce unrelated results.

External Contextual Effects on IAT Results

We have seen that the stimuli chosen to represent the target categories in an IAT appear to act within the test to alter its results. In the discussion of Study 3, I also suggested that factors outside an IAT, such as the context in which it was completed, could affect its results. A variety of published studies have demonstrated such effects of external contextual factors on the automatic associations measured by IATs. For example, IATs measuring implicit racial attitudes showed weaker preferences for Whites when completed in the presence of a Black rather than White experimenter (Lowery, Hardin, & Sinclair, 2001). Similarly, subjects showed weaker implicit racial preferences after examining photos of admired Black individuals and disliked Whites (Dasgupta & Greenwald, 2001).

Analyses of IAT completion order in Study 3 suggested that the stimuli used in one IAT created a context that affected the results of a following IAT. How might this occur? Imagine two scenarios. First, subject A begins the experiment with the accent IAT and categorizes the friendly greetings based on the race of the speaker. She then begins the second IAT, categorizing cropped faces with blank expressions. However, the context of *friendly* AA and EA targets set by the first IAT remains salient while she categorizes these relatively simple neutral stimuli. Contrast her situation with subject B who begins the experiment with the face IAT. The cropped faces with neutral expressions were carefully manipulated to remove contextual cues, so his automatic associations are elicited based on the identification of the stimuli as AA or EA, and no other context is systematically elicited. Racial prejudice was apparently greater without the context of friendly targets because the face IAT results for participants who completed it first (like subject B) showed stronger implicit preferences for EA over AA targets than did the face IAT results of participants completing it second (like subject A; see Figure 4).

Several additional results from Study 3 also suggest that context effects were present. The accent IAT and face IAT were correlated when the accent IAT was completed first ($r = .50$), and uncorrelated when the face IAT was first ($r = .066$). It is unsurprising that the two IATs were correlated for persons like subject A, who completed both IATs in the context of friendly EA and AA targets. In contrast, subject B first completed the face IAT with no particular context, and then completed the accent IAT, in which the stimuli elicited the context of *friendly* AA and EA targets. The different contexts produced uncorrelated IAT results. Finally, because the accent IAT elicited the same context of friendly targets for both subject A and B, the accent IAT results in Study 3 did not differ by completion order (see Figure 4).

The scenarios illustrate how in Study 3 the friendly characteristics of the accent stimuli, even though they were unrelated to (and not confounded with) their racial category, may have altered the automatic associations elicited in both the accent IAT and the face IAT (depending on completion order). Study 1 also included several IATs that might be expected to show similar patterns of results. Specifically, order effects might suggest that a context created by the sound IAT affected the results of a following IAT. No such order effects were evident in the correlations between the sound and picture or sound and speech IATs. However, completion order effects similar to Study 3 were observed for the sound IAT and text IAT. First, the magnitudes of the IAT effects showed a pattern consistent with Study 3. Specifically, the sound and text IAT effects were comparable when the sound IAT preceded the text IAT (257 and 275 ms) but diverged (although not significantly, 171 and 235 ms; $t(41) = 1.58$; $p = .123$; $d = .5$) when the text IAT was completed first. Second, the correlation between the sound and text IATs was higher (although not significantly higher in this test with low statistical power) when the sound IAT was completed first ($r = .449$) than when completed second ($r = .023$).

In summary, a variety of results from this research suggest that stimuli used in one IAT created a context that affected the results of a following IAT. Although further research is necessary before concluding if or when such external contextual effects can alter IAT results, confirmation of such effects would be predicted based on a recently adopted theoretical model of automatic associations.

A Distributed Processing Model of Automatic Associations

Mitchell et al. (2002) propose that automatic associations can best be understood using a model that describes cognitive associations as flows of activation along connections between units rather than as static relationships that are retrieved (see McClelland & Rummelhart, 1985). Building on the arguments of Smith (1996), they suggest that attitudes that operate outside conscious control (i.e., the type of attitudes measured by the IAT) are best conceptualized as, "...patterns that are reconstructed within the parameters of a particular context," (p. 4), and argue that variability in attitude expression brought on by such contextual influences should not be thought of as measurement error, but as a characteristic inherent in the nature of attitudes.[13]

One piece of empirical support Mitchell et al. (2002) provide for the distributed processing model is a study in which IATs using different stimuli but the same category labels produced different results. They found that subjects showed no racial preference when disliked politicians represented whites and liked athletes represented blacks, but

[13] This is not to say that attitude expression should be random. Attitude measurement is expected to be much like defining the shoreline at an ocean beach. The position of the water's edge at every moment reflects a unique combination of factors including tides, wind, and waves, such that its exact location cannot be predicted. Still, there is consistency on useful scales of time and location.

that subjects showed preferences for whites when unknown targets were sorted. Study 2 showed a similar result in that two IATs used the same category labels (male vs. female) but different stimuli (computerized voices vs. written names) evoked different associations, possibly because the female computerized voice sounded relatively confident and assertive. It is consistent with the distributed processing model that the stimuli under consideration are a factor in the evocation of automatic associations.

The distributed processing model can also be used to explain why the context in which stimuli are evaluated alter the automatic associations that stimuli elicit. Recall that the order effects observed in Study 3 suggested that identical face IATs measured different automatic associations when a context of "friendly" targets was made salient by a preceding voice IAT. If automatic associations are reconstructed patterns, it is not surprising that the associations in Study 3 differed in different contexts and that evaluative associations for the African-American and European-American targets were less discrepant in the context of "friendly" voices.

Potential Implications For IAT Construction

The primary implication of the distributed processing model arises from the proposition that IATs measure unique patterns of cognitive association evoked by the combination of many contextual factors. Thus, the proper use of IATs in applied settings and for purposes of theoretical development depends on awareness and proper manipulation of such contextual factors, particularly within the IAT design (e.g., stimulus selection), but also outside the IAT procedure, considering factors such as experimenter effects and the order of task completion.

The accent IAT used in Study 3 provides an example of how careful stimulus selection might increase the test's utility for application. At first glance, the emotionally warm nature of the greetings might appear to be a limitation of the test. However, they may enhance its validity for diagnosing the automatic associations evoked in job interviews, apartment applications, or any other situation in which persons present themselves in a friendly manner in order to make a good impression.

Careful stimulus selection may also be critical in the use of IATs for the development of theories concerning the properties of automatic associations. Studies of change in the automatic associations measured using cropped, gray-scale photos of faces with neutral expressions (e.g., Dasgupta & Greenwald, 2001) or other context-free stimuli such as names presented in text (e.g., Blair, Ma, & Lenton, 2001) suggest that implicit racial attitudes can be altered through simple interventions. However, while it may be relatively easy to pare complex stimuli (e.g., human beings) of all associations but race then provide new associations with those simple stimuli that alter the IAT results they elicit in the laboratory context, it may be dramatically more difficult to alter implicit racial attitudes outside the lab by altering the myriad associations elicited by actual human beings in multi-dimensional contexts. IATs measuring changes in implicit racial attitudes might show smaller effects if they were to use a different set of realistically complex stimuli in the pre and post-manipulation.

Conclusion

A variety of implicit cognitive associations can be measured by IATs that use auditory stimuli. Auditory IATs include sound IATs that use environmental sounds as stimuli, and speech IATs in which spoken words are used. Both sound and speech IATs widen the applicability of the IAT in a variety of useful ways. The results of three studies demonstrating auditory IATs suggest that their results can be influenced by contextual factors such as the stimuli chosen to represent the target categories. Such sensitivity presents an opportunity for greater understanding of implicit cognitive associations and for the development of IATs that are effective in measuring the automatic associations relevant to specific applications. Auditory IATs thus promise to both expand the application of the IAT and enhance understanding of the way it measures automatic cognitive associations.

REFERENCES

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

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

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.

Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin, 87*, 546-563.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*, 163-170.

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.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36*, 316-328.

De Houwer, J. (2002). *The extrinsic affective simon task.* Manuscript submitted for publication. University of Ghent, Belgium.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental* Social *Psychology, 37*, 443-451.

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

Draine, S. (1998). Inquisit [Computer software]. Seattle, WA: Millisecond Software. Available: http://www.millisecond.com/ [version 28].

Empirisoft (2001). Medialab. [Computer software]. New York, NY: Empirisoft Software. Available: http://www.empirisoft.com/ .

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

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*, 71-99.

Gaertner, S. L. (1973). Helping behavior and racial discrimination among liberals and conservatives. *Journal of Personality and Social Psychology, 25*, 335-341.

Gaertner, S. L. & McLaughlin, J. P. (1983). Racial stereotypes: Associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly, 46*, 23-30.

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

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

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A. & Rosier, M. (2000). Prologue to a unified theory of attitudes, stereotypes, and self-concept. In J. P. Forgas (Ed), *Feeling and thinking: The role of affect in social cognition. Studies in emotion and social interaction, second series* (pp. 308-330). New York, NY: Cambridge University Press.

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

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

Greenwald, A. G., Nosek, B., & Banaji, M. R. (2002). *Scoring procedures to improve implicit association test measures.* Manuscript submitted for publication.

Haines, E. L. (2000). Elements of a social power schema: Gender standpoint, self-concept, and experience (Doctoral dissertation). *Dissertation Abstracts International, 60(9-B)*: 4964.

Hays, W. L. (1988). *Statistics.* Fort Worth, TX: Holt, Rinehart, & Winston.

Katz, I., Cohen, S., & Glass, D. (1975). Some determinants of cross-racial helping behavior. *Journal of Personality and Social Psychology, 32,* 964-970.

Kim, D-Y., & Greenwald, A. G. (2000). *Voluntary controllability of implicit cognition: Can an implicit measure (the IAT) of attitudes be faked?* Manuscript submitted for publication.

Kroemer, K. H. E. & Grandjean, E. (1997). *Fitting the task to the human : a textbook of occupational ergonomics.* London ; Bristol, PA : Taylor & Francis.

Kuhn, T. S. (1970). *The structure of scientific revolutions.* Chicago: University of Chicago Press.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81,* 842-845.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The implicit association test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 32,* 1-9.

McClelland, J. L. & Rummelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General,* 114, 159-188.

McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution, 25,* 563-579.

McConnell, A. R. & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*(5), 435-442.

McGhee, D. (2002). Measuring relative versus absolute associations with the implicit association test. Manuscript in preparation.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2002). Contextual variations in implicit evaluation. Manuscript submitted for publication.

Neter, J., Wasserman, W., & Kutner, M. H. (1990). Applied linear statistical models: regression, analysis of variance, and experimental designs. Boston, MA : Irwin.

Nosek, B. A. & Banaji, M. R. (in press). The go/no-go association task. Social Cognition.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2000). Math = Male, Me = Female, therefore Math ≠ Me. Unpublished manuscript. Yale University, New Haven, Ct.

Percodani, J., Haines, E. L., Adames, H., Fisher, D., & Greenwald, A. G. (2000). Implicit race preferences generalize to younger members of outgroups. Poster session presented at the annual meeting of the Eastern Psychological Association, Baltimore, MD.

Purnell, T., Idsardi, W., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology, 18*, 10-30.

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin. 27*, 1164-1178.

Sibley, D. A. (2000). National Audubon Society the Sibley guide to birds. New York: Alfred A. Knopf.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology, 70*, 893-912.

Staples, S. (1996). Human response to environmental noise: Psychological research and public policy. *American Psychologist, 51*(2), 143-150.

Swanson, J. E. (2001a). [Measuring implicit attitudes toward smoking]. Unpublished raw data.

Swanson, J. E. (2001b). Investigating implicit and explicit cognitions associated with smoking. Unpublished doctoral dissertation, University of Washington, Seattle.

Swanson, J. E. & Greenwald, A. G. (1997). [Measuring omnivores implicit attitudes toward different foods]. Unpublished raw data.

Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude-behavior consistency for stigmatized behavior. *Cognition and Emotion, 15,* 207-230.

Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110,* 226-235.

Tisak, J., & Tisak, M. S. (2000). Permanency and ephemerality of psychological measures with application to organizational commitment. *Psychological Methods, 5*, 175-198.

Van Gausig, D. (2000). http://Naturesongs.com.

Walton, J. H. & Orlikoff. (1994). Speaker race identification from acoustic cues in the vocal signal. *Journal of Speech and Hearing Research, 37*, 738-745.

Wegner, D. M. & Bargh, J. A. (1998). Control and automaticity in social life. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 1, pp. 446-496). New York: Oxford University Press, Inc.

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

APPENDIX A

Categories and Stimuli Used in IATs

Study 1

Attributes:

Pleasant CUDDLE, HAPPY, SMILE, JOY
Unpleasant ABUSE, CRASH, DISASTER, GRIEF

Target Concepts – text and speech IATs:

Bird BLACKBIRD, CARDINAL, WARBLER, ROBIN
Insect BEE, CICADA, LOCUST, MOSQUITO

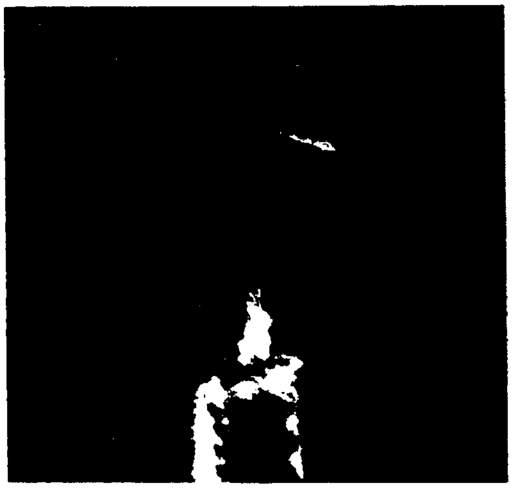
Target Concepts – sound IAT:

Bird .WAV files of calls of redwing blackbird, cardinal, warbler, robin
Insect .WAV files of sounds of bee, cicada, locust, mosquito

All from http://Naturesongs.com and used with permission

Target Concepts – picture IAT:

Bird




Insect

<u>Study 2</u>

Attributes:

high power command, confident, dominant, strong
low power follow, submissive, timid, weak

Target Concepts – text and speech IATs:

MALE ERIC, JASON, PETER, STEVE
FEMALE DIANNE, KAREN, LISA, SANDRA

Target Concepts – sound IAT:

MALE center, gallon, halfway, inch – Spoken by male computer-generated voice
FEMALE center, gallon, halfway, inch – Spoken by female computer-generated voice

<u>Study 3</u>

Attributes:

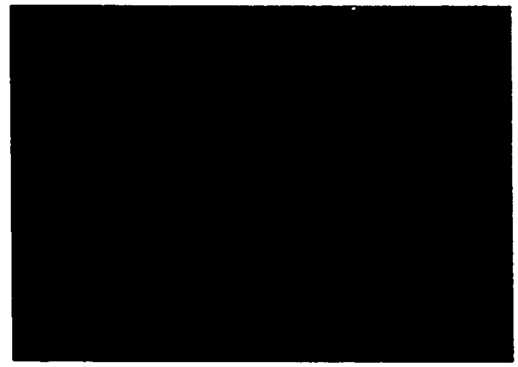
Pleasant gift, health, joy, laughter, rainbow, warmth
Unpleasant agony, cancer, failure, poison, sickness, vomit

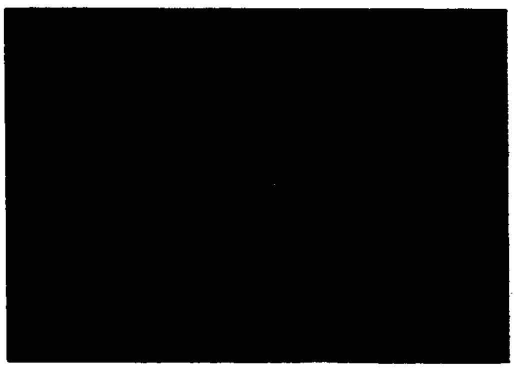
Target Concepts – accent IAT:

African American	Hey, how ya' doin? How are you? How ya' doin', man? Hey, good morning.	Spoken by male African Americans
European American	Hey, how ya' doin? How are you? How ya' doin', man? Hey, good morning.	Spoken by male European Americans

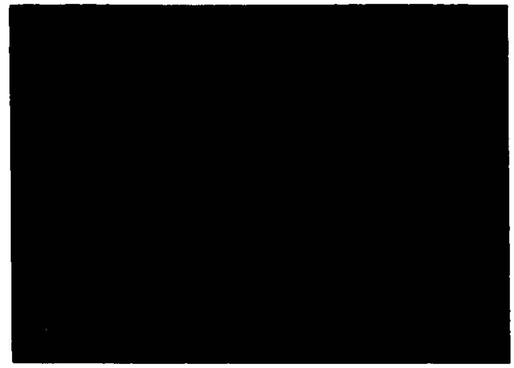
Target Concepts – face IAT:

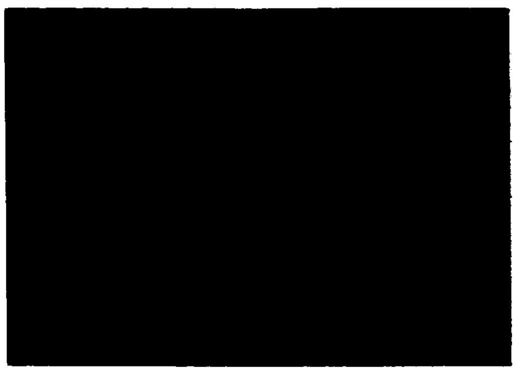
African American





European American





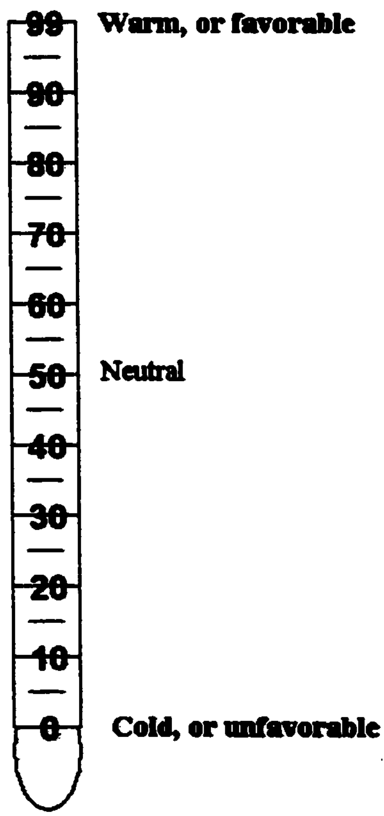
APPENDIX B

Explicit Measures

<u>Study 1</u>

Attitude thermometers (thermometer for BIRDS shown below, INSECTS used same format):

1. Please make a mark on the scale below indicating how you feel about **<u>BIRDS</u>** in general. If you mark somewhere 1° and 49° that indicates you feel cold or unfavorable toward birds. Marking 50° means that you feel neutral, and marking between 51° and 99° means that you feel warm or favorable toward birds.



Semantic differential measure. Subjects rated BIRDS and INSECTS on five seven-point scales (Ugly/Beautiful, Bad/Good, Unpleasant/Pleasant, Harmful/Harmless, Awful/Nice).

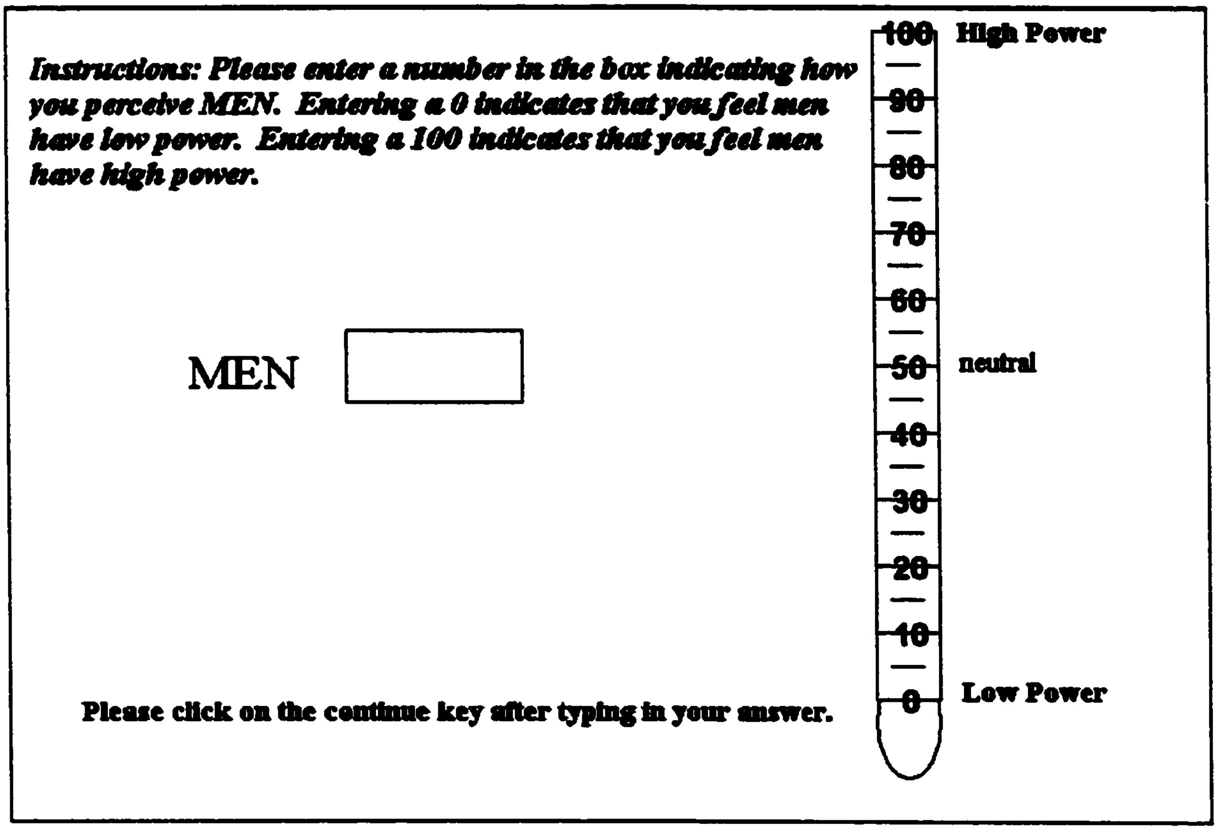
Example:

<u>BIRDS</u>

Ugly :____:____:____:____:____:____:____: **Beautiful**

Study 2

Stereotype thermometers (thermometer for MEN shown below, WOMEN used same format):



Semantic differential measure. Subjects rated MEN and WOMEN on four seven-point scales corresponding to the high-power and low-power words used in the text IAT and speech IAT (Weak/Strong, Timid/Confident, Submissive/Dominant, Follow/Command).

Example:

MEN

1	**2**	**3**	**4**	**5**	**6**	**7**
Weak						**Strong**

Study 3

Modern Racism Scale (McConahay, Hardee, & Batts, 1981).

1. Discrimination against Blacks is no longer a problem in the United States.
2. It is easy to understand the anger of Black people in America.
3. Blacks have more influence upon school desegregation plans than they ought to have.
4. Blacks are getting too demanding in their push for equal rights.
5. Blacks should not push themselves where they are not wanted.
6. Over the past few years, Blacks have gotten more economically than they deserve.
7. Over the past few years, the government and news media have shown more respect for Blacks than they deserve.

Diversity scale (Wittenbrink, Judd, & Park, 1997).

1. There is a real danger that too much emphasis on cultural diversity will tear the United States apart.
2. The desire of many ethnic minorities to maintain their cultural traditions impedes the achievement of racial equality.
3. Whites will need to learn about Black Culture if positive interethnic relations are to be achieved.
4. The establishment and maintenance of all-Black groups and coalitions prevents successful racial integration.

Discrimination scale (Wittenbrink, Judd, & Park, 1997).

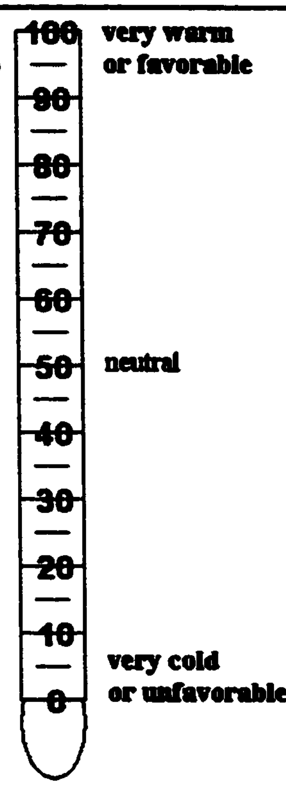
1. Members of ethnic minorities have a tendency to blame Whites too much for problems that are of their own doing.
2. Members of ethnic minorities often exaggerate the extent to which they suffer from racial inequality.
3. Black people often blame the system instead of looking at how they could improve their situation themselves.
4. These days, reverse discrimination against Whites is as much a problem as discrimination against Blacks itself.
5. More and more, Blacks use accusations of racism for their own advantage.
6. Blacks are ultimately responsible for the state of race relations in this country.
7. Discrimination against Blacks is no longer a problem in the United States.
8. A primary reason that ethnic minorities tend to stay in lower paying jobs is that they lack the motivation required for moving up.
9. Many ethnic minorities do not understand how hard one has to work to achieve success.
10. In the U.S. people are no longer judged by their skin color.

Attitude thermometers (thermometer for AFRICAN-AMERICANS shown below, EUROPEAN-AMERICANS used same format):

Instructions: Please enter a number in the box indicating how you feel about AFRICAN-AMERICANS. Entering a 0 indicates that you feel extremely unfavorable toward African-Americans. Entering a 100 indicates that you feel extremely favorable toward African-Americans.

AFRICAN-
AMERICANS [ ]

Please click on the continue key after typing in your answer.



Semantic differential measure. Subjects rated AFRICAN-AMERICANS and EUROPEAN-AMERICANS on five seven-point scales (Ugly/Beautiful, Bad/Good, Unpleasant/Pleasant, Harmful/Harmless, Awful/Nice).

Example:

AFRICAN-AMERICANS

1	2	3	4	5	6	7
Ugly						**Beautiful**

Curriculum Vitae
December 2002

Mark E. Vande Kamp

Personal Information:

Address: (Office)
National Park Service
Cooperative Ecosystem Studies Unit
College of Forest Resources
Box 352100
Seattle, Washington 98195-2100
(206) 543-0378

(Residence)
555 N 75th Street
Seattle, WA 98103
(206) 789-6455

Education:

Ph.D. University of Washington, Seattle, WA, December, 2002.
Major: Social Psychology
Minor: Quantitative Methods

M.S. University of Washington, Seattle, WA, August, 1988.
Major: Social Psychology
Minor: Quantitative Methods

B.S. Central College, Pella, IA, May 1986, *summa cum laude*.
Major: General Studies
Honors Research Director: Prof. Edmond E. Willis

Awards:

H. S. Kuyper Scholarship: Central College, Pella, IA
National Merit Scholar: Corporate Scholarship (Hormel Corporation)

Professional Experience:

July 1991 - Present	Research Associate, NPS, Cooperative Ecosystem Studies Unit, College of Forest Resources, University of Washington.
Mar. 1991 - July 1991	Research Consultant, Fred Hutchinson Cancer Research Center, Cancer Prevention Research Unit.

Professional Experience (continued):

Oct. 1990 - Mar. 1991	Research Consultant, Seattle METRO, Research and Market Strategy Division.
Jan. 1990 - Jun. 1990	Instructor: Introductory Social Psychology, Psychological Research Methods
1986 - 1990	Teaching/Research Assistant, University of Washington.

Professional Presentations:

Johnson, D. R., Swearingen, T. C., & Vande Kamp, M. E. (1994, June) *A survey of park managers' perceptions of noncompliant visitor behavior causing resource damage in the National Park System.* Paper presented at The Fifth North American Symposium On Society And Resource Management, Fort Collins, CO.

Manning, R. E., Johnson, D. R., & Vande Kamp, M. E. (1994, June) *Norm congruence among tour boat passengers to Glacier Bay National Park.* Paper presented at The Fifth North American Symposium On Society And Resource Management, Fort Collins, CO.

Vande Kamp, M. E. & Greenwald, A. G. (1990, May) *Further investigation of self-esteem effects in "Normal" behavior.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.

Vande Kamp, M. E. & Johnson, D. R. (1997, March) *Mount Rainier National Park: A case study in the application of social science to the VERP GMP planning model.* Panel discussion at the George Wright Society meetings, Albuquerque, NM.

Vande Kamp, M. E., Johnson, D. R., & Manning, R. E. (1994, June) *Encounters, disturbing encounters, expectations, preferences and trip enjoyment: Glacier Bay cruise ship passengers who see planes, boats, and other ships.* Paper presented at The Fifth North American Symposium On Society And Resource Management, Fort Collins, CO.

Vande Kamp, M. E., Johnson, D. R., & Swearingen, T. C. (1992, May) *Deterring minor acts of noncompliance.* Paper presented at The Fourth North American Symposium On Society And Resource Management, Madison, WI.

Vande Kamp, M. E., Kerr, K. K., & Greenwald, A. G. (1989, May) *Is high self-esteem a precondition of "normal" behavior?* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, IL.

Professional Presentations (continued):

Vande Kamp, M. E. & Swanson, J. E. (2002, September). *Tools for managing social impacts of recreation.* Paper presented at the National Park Service Recreational Impacts Workshop, Seattle, WA.

Book Review:

Wildlife and Recreationists: Coexistence Through Management and Research. Richard L. Knight and Kevin J. Gutzwiller (eds.) in *Northwest Science, 70*(4), 1996. pp. 367-369.

Publications and Technical Reports:

Johnson, D. R., Hunn, E., Russell, P., Vande Kamp, M. E., & Searles, N. (1998). *Subsistence Uses of Vegetal Resources in and Around Lake Clark National Park and Preserve.* Technical Report NPS-19. Field Station/ Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 487 pp.

Johnson D. R., Rugh, J. C., Vande Kamp, M. E., & Swearingen T. C. (1994). Minor violations, major damage: A survey of noncompliant visitor behavior and managerial practice. *Park Science: A Resource Management Bulletin, 14*(3): 8-10.

Johnson, D. R. & Vande Kamp, M. E. (1996). Extent and control of resource damage due to noncompliant visitor behavior: A case study from the U.S. National Parks. *Natural Areas Journal, 16*(2): 134-141.

Johnson, D. R., Vande Kamp, M. E., & Swearingen, T. C. (1992). *A survey of park managers' perceptions of noncompliant visitor behavior causing resource damage in the national park system.* Technical Report NPS/PNRUW/NRTR-92-07. United States Department of Interior, National Park Service, Pacific Northwest Region, Seattle, Washington. 131 pp.

Manning, R. E., Johnson, D. R., & Vande Kamp, M. E. (1996). Norm congruence among tour boat passengers to Glacier Bay National Park. *Leisure Sciences, 18*, 2: 225-241.

Pergola, T., Johnson, D. R., Paschel, J. M., & Vande Kamp, M. E. (1997). *Ebey's Landing National Historical Reserve 1995 Visitor Survey.* Technical Report NPS/PNRUW/NRTR-97-06. Field Station/ Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 176 pp.

Publications and Technical Reports (continued):

Porte-Gendron, R. W., Simpson, T., Carlson, K. K., & Vande Kamp, M. E. (1997). Baccalaureate nurse educators and critical care nurse managers' perceptions of clinical competencies necessary for new graduate baccalaureate critical care nurses. *American Journal of Critical Care, 16,* 2: 147-158.

Swanson, J. E., Vande Kamp, M. E., Johnson, D. R., Manning, R. E., & Lawson, S. R. (2002). A Survey of Overnight Backcountry Visitors to Denali National Park and Preserve. Technical Report NPS/CCSOUW/NRTR-2002-04. Cascadia Field Station USGS/BRD/FRESC, College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 408 pp.

Vande Kamp, M. E. (1995). *Public input concerning proposed changes to visitor facilities in the Sunrise area of Mount Rainier National Park: A summary of public meetings held October 19, 1994 in Yakima, Washington and October 20, 1994 in Tacoma, Washington.* Unpublished Report. NBS/CPSU, College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 58 pp.

Vande Kamp, M. E. (1997). *How can we best use existing information to set social carrying capacity standards in the wilderness areas of Zion National Park?* Unpublished Report. Field Station/ Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 34 pp.

Vande Kamp, M. E. (1998). *The uses of existing information in the process of setting social standards for proposed wilderness areas of Zion National Park.* Technical Report NPS/PNRUW/NRTR-98-07. Field Station For Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 84 pp.

Vande Kamp, M. E. & Johnson, D. R. (1998). *A survey of visitors who planned their party's trip to the White River and Sunrise areas of Mount Rainier National Park.* Technical Report NPS/PNRUW/NRTR-98-05. Field Station/ Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 190 pp.

Vande Kamp, M. E. & Johnson, D. R. (1998). *Visitor Density in Facilities: 1995 Mt. Rainier National Park Facilities Survey.* Technical Report NPS/PNRUW/NRTR-98-11. Field Station/ Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 346 pp.

Publications and Technical Reports (continued):

Vande Kamp M. E., Johnson, D. R., Kucera, A. & Young, Y. (1997). *Describing and estimating the system of visitor distribution in Mt. Rainier National Park: 1995 Visitor Distribution Survey.* Technical Report NPS/PNRUW/NRTR-97-07. Field Station/ Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 431 pp.

Vande Kamp, M. E., Johnson, D. R., & Manning, R. E. (2001). *Application of Visitor Experience And Resource Protection (VERP) to Alaskan National Park Wilderness.* Technical Report NPS/CCSOUW/NRTR-2001-01. Cascadia Field Station USGS/BRD/FRESC, College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100.

Vande Kamp, M. E. Johnson, D. R., Paschel, J. M., & Pergola, T. (1998). *A Survey of Visitors to Mowich Lake in Mt. Rainier National Park.* Technical Report NPS/CCSOUW/NRTR-99-01. University of Washington Field Station/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 123 pp.

Vande Kamp M. E., Johnson, D. R., & Swanson, J. E. (1998). *Mt. Rainier National Park 1993 Spray Park Visitor Survey.* Technical Report NPS/PNRUW/NRTR-98-04. Field Station/ Protected Area Research/USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 294 pp.

Vande Kamp, M. E., Johnson D. R., & Swearingen T. C. (1992). *Deterring Minor Acts of Noncompliance: A Literature Review.* Technical Report NPS/PNRUW/NRTR-92-08. United States Department of Interior, National Park Service, Pacific Northwest Region, Seattle, Washington. 92 pp.

Vande Kamp, M. E., Johnson D. R., & Swearingen T. C. (1994). Preventing visitor-caused damage to National Park resources: What do we know? What should be done? *Park Science: A Resource Management Bulletin, 14*(3): 8-10.

Vande Kamp, M. E. & Schulz, S. L. (1993). *The Hanford Reach of the Columbia River: Public input regarding the draft comprehensive river conservation study and environmental impact statement.* Unpublished Report. NPS/CPSU, College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 62 pp.

Publications and Technical Reports (continued):

Vande Kamp, M. E., Swanson, J. E., & Johnson, D. R. (1999). *A Survey of Wilderness Trail Users in Mount Rainier National Park.* Technical Report NPS/CCSOUW/NRTR-2000-01. Cascadia Field Station USGS/BRD/FRESC. College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 234 pp.

Vande Kamp, M. E., Swanson, J. E., & Johnson, D. R. (2002). *A Survey of Visitors to the Nisqually River-Stevens Canyon Corridor in Mount Rainier National Park.* Technical Report NPS/CCSOUW/NRTR-2002-01. Cascadia Field Station USGS/BRD/FRESC, College of Forest Resources, University of Washington, Box 352100, Seattle, WA, 98195-2100. 233 pp.