Social Cognition, Vol. 19, No. 6, 2001, pp. 625-664

THE GO/NO-GO ASSOCIATION TASK

Brian A. Nosek and Mahzarin R. Banaji
Yale University

Theory is constrained by the quality and versatility of measurement tools. As such, the development of techniques for measurement is critical to the successful development of theory. This paper presents a technique — the Go/No-go Association Task (GNAT) — that joins a family of existing techniques for measuring implicit social cognition generally, with a focus on attitude (evaluation). To expand the measurement potential supplied by its closest cousin, the Implicit Association Test (IAT), the GNAT can be used to examine automatic social cognition toward a single target category. That is, the GNAT obtains a measure of implicit social cognition without requiring the direct involvement of complementary or contrasting objects. Also, by implementing a response deadline in the procedure, this version of the GNAT trades off response latency for sensitivity as the dependent variable measure. We illustrate the technique through a series of experiments (1-5) using simple attitude objects (bugs and fruit). In Experiment 6, the GNAT is used to investigate attitudes toward race (black and white) and gender (male and female). To explore the theoretical leverage offered by this tool, Experiment 6 puts to test a recurring question concerning automatic in-group favoritism versus out-group derogation. Results demonstrate the dual presence of both out-group derogation (e.g., negativity toward black Americans) and in-group favoritism (positivity toward white Americans), a finding that emerges because the GNAT offers the potential for separable measures of attitude toward the two groups. Through these experiments, the GNAT is shown to be an effective tool for assessing automatic preferences as well as resolving persistent questions that require measures of individual attitude objects while maintaining the advantages of response competition tasks.

The authors thank Mark Honbo for assistance with stimuli collection, David McNellis for assistance with data collection, and Eric Uhlmann and Thierry Devos for helpful comments on a previous draft. Research reported in this paper was supported by grants from the National Institute of Mental Health (MH-57672) and from the National Science Foundation (SBR-9709924) to Mahzarin R. Banaji. Brian Nosek was supported by a graduate fellowship from Yale University.

Address correspondence to Brian Nosek, Department of Psychology, William James Hall, 33 Kirkland St., Cambridge, MA 02138.

> To say that a theory should deal with consequences is to say that it should directly address issues of measurement.
> — Thomas. M. Ostrom (1989, p. 11)

Like any area of psychological research, investigations of implicit social cognition are dependent on tools of measurement. In recent years, the tool kit for such research has been significantly expanded by introducing techniques such as evaluative priming (Fazio, Jackson, Dunton, & Williams, 1995; Fazio, Sanbonmatsu, Powell, & Kardes, 1986), unconscious priming (Draine & Greenwald, 1998), the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), semantic priming (see Neely, 1991), the Stroop task (see MacLeod, 1991), and the Simon task (DeHouwer & Eelen, 1998; see DeHouwer, in press, for a review of multiple measures). Each of these tasks was designed to assess judgments that reflect automatically activated evaluation or knowledge, without the performer's conscious awareness or control (see Banaji, 2001, for a discussion).

Each such tool in use today attempts to obtain a rough indication of the contents of cognition and the processes that reveal their operation. Among those interested in the connection between the mental world and the social world, the work has often focused on attitudes and preferences as well as knowledge and stereotypes. Yet every tool, no matter how robust, is likely to only partially reflect the underlying construct along with some error associated with the measurement, sometimes introduced by the properties of the tool itself. As Ostrom (1989) commented about attitude measurement: "As new measurement techniques emerge, new questions about the nature of attitudes arise. Each method embodies its own view of attitude and alerts the researcher to phenomena that emerge within the context of that view" (p. 20). Moreover, his comment with which we begin this paper, was written specifically about attitude theory to suggest that any theory of attitude that expects to speak to the implications or consequences for thought and action must necessarily and directly deal with questions of measurement.

The proliferation of techniques at this early stage promises many rewards because the known properties of each can reveal unique aspects of social cognition that may otherwise remain hidden. In the spirit of promoting a diversity of methods, we introduce a technique called the Go/No-go Association Task (GNAT) as a measure of implicit social cognition.[1]

1. Sample GNAT scripts and analysis programs are available at www.briannosek.com

Using simple attitude objects (insects versus fruit), the purpose of the first five experiments is purely tool development, with an emphasis on the technical components of the task and methods for data analysis. The final experiment extends GNAT to social attitudes and shows its advantage in detecting group favoritism versus derogation that has been difficult to tease apart with some previous measures.

DESIGN OF THE GNAT

Like priming tasks and the IAT, the GNAT indexes an implicit attitude or belief by assessing the strength of association between a target category and two poles of an attribute dimension (for all experiments in this paper the attribute dimension is evaluation, that is, good-bad). In the GNAT, strength of association is assessed by the degree to which items belonging to the target category and attribute (e.g., fruit and good) can be discriminated from distracter items that do not belong to those concepts. One condition requires simultaneous identification of stimuli that represent the target category (fruit) and an attribute (good). A second condition requires simultaneous identification of stimuli that represent the same target category and an alternative attribute (bad). The extent to which fruit is associated with good versus bad ought to be reflected in the relative ease of discriminating fruit with one versus the other evaluative attribute. In the present example, fruit is generally regarded positively because they are a tasty and nourishing food source, compared to insects that are generally regarded negatively as annoying pests. If this is so, accuracy in discriminating fruit and good items from distracters ought to be higher than accuracy in discriminating fruit and bad items from distracters. The difference in accuracy ("sensitivity" in signal detection terms) between these conditions is taken as a measure of automatic attitude.

The GNAT works by presenting target (signal) and distracter (noise) stimuli for brief periods of time. Unlike other tasks in this family, the GNAT requires the same response - "go" (press the space bar) — to items that belong to instances of a category (e.g., fruit) and a particular evaluative attribute (e.g., good) both of which, for this purpose, serving as the signal. No response "no-go" (do not press any key) is called for when items appear that do not belong to the target category and attribute (noise). The extent to which the target category and attribute comprising the signal are associated should determine sensitivity, or the discriminability of signal from noise (indexed as d' in signal detection theory [SDT]). Thus, differences in sensitivity between pairing conditions (e.g., *fruit + good* vs. *fruit + bad*) reflect the association between the

concept and evaluation, and association is taken to be a measure of automatic attitude.

In any task requiring speeded classification of stimulus items, task performance is affected by the relative emphasis on speed versus accuracy. Increasing the speed of responding increases the potential for errors. Likewise, making certain that responses are "correct" results in a decline in overall speed. Because of the speed-accuracy tradeoff, meaningful information about task performance can be found in both average response times and error rates. Indeed, Greenwald et al. (1998) noted that IAT effects could be observed using either response latency or error rates as the operational dependent variable (see also Nosek, Banaji, & Greenwald, in press). However, most implicit measures use response latency exclusively as the dependent variable and therefore may lose relevant information contained in error rates. In Signal Detection Theory, response strategies in the speed-accuracy tradeoff are indexed separately as bias (the statistic, ß) from sensitivity (d'), which is calculated via error rates. Because bias is conceptually independent from sensitivity, subjects' response strategy does not obscure the measure of interest.

The GNAT shares many features with other measures of implicit social cognition while containing some unique ones. Although the typical dependent variable for measures of automatic social cognition is response latency, the GNAT procedure builds in a measure of accuracy through the statistic of sensitivity (Green & Swets, 1966) and that is demonstrated in Experiments 1-4. It can equally be implemented using response latency as the dependent variable as shown in Experiment 5. More important, the GNAT is flexible in the use of distracters, and in this measure, these distracters form the critical context of measurement. Assessment of implicit social cognition, as any psychological assessment, is likely heavily dependent on the context in which the measure is obtained. Just as the measurement of explicit attitudes are influenced by numerous variables including the format of the scale, the options for response, the manner in which the item is framed, and the order of the items (Judd & McClelland, 1998; Schwarz, Groves, & Schuman, 1998), implicit measures are not likely to be impervious to influence. Indeed, recent papers suggest that relatively minor changes in context can have substantial effects on implicit preferences (Lowery, Hardin, & Sinclair, 2001; Mitchell, Nosek, & Banaji, 2001).

Among the most obvious determinants of measures of implicit social cognition is the context set by the contrasting category in reaction to which the attitude is obtained. In procedures like evaluative priming and the IAT, attitudes and stereotypes toward black Americans are obtained in relation to white Americans (Fazio et al., 1995; Greenwald et al., 1998; Nosek et al., in press), toward women in relation to men (Banaji & Hardin, 1996;

Blair & Banaji, 1996; Carpenter & Banaji, 2000; Rudman & Kilianski, 2000), and toward the elderly in relation to the young (Levy, 1996; Nosek et al., in press; Perdue & Gurtman, 1990). In both the IAT and priming measures, the presence of such polarities have been the norm, but with a difference. Evaluative and semantic priming procedures allow separable measures of response latencies toward each attitude object (e.g., black versus white), whereas the IAT requires the joint, and hence more starkly relative, assessment of the attitude. The IAT requirement of the presence of a second attitude object directly in the measurement context constrains the interpretation of the effect to the particular comparison object. Attitudes toward black Americans may differ when the context object is white versus Hispanic versus Asian Americans.

In addition, the data obtained in IAT procedures cannot be meaningfully broken down trial by trial. That is, the data for trials associating white with good are collected simultaneously with trials associating black with bad and cannot be analyzed separately. Because responses to items belonging to one pairing (*white* + *good*)[2] occur at the same time as responses to the other pairing (*black* + *bad*), faster response times are a product of both associative pairs. That is, what may appear to be an association between white and good could also be a function of a strong association in the opposing pairing (*black* + *bad*) or even inhibition from the categories requiring opposite responses (e.g., black and good).

For a variety of concepts in the social world, evaluation of one category may naturally occur relative to, or with reference to, a second category (e.g., male and female, old and young, Coke and Pepsi). Such natural pairs are critical to a measure like the IAT because it relies on dual and competing categorizations. There are, however, some types of research questions in which this feature of the IAT is limiting. Even in domains in which the relative comparison may be critical, assessments of single categories may still be of interest. For example, in-group liking and out-group derogation may be reciprocal in the sense that a decrease in one comes at the expense of an increase in the other. On the other hand, they may be independent of each other such that one exists without corresponding to the other (Hinkle & Brown, 1990; see Brewer & Brown, 1998, for a review). These are features to which we remained alert in designing the GNAT.

A feature of the GNAT is its adaptability in selecting a context for measuring preferences. Preferences for a category (e.g., trout) might be as-

2. In this example, and for the rest of the paper, use of a "+" joiner between two concepts means that the same behavioral response (a key press) is required for members of both categories.

sessed in the context of (1) a single other concept (e.g., salmon; single category context), (2) a superordinate category (e.g., fish; superordinate context), (3) a generic category (generic context), or (4) no categories at all, but just evaluative or other attributes (attribute-only context). This flexibility of the GNAT allows for assessment of preferences for specific categories to be isolated from relative comparisons with a singular category allowing for more direct investigations of a specific attitude object (see Blair, Ma, & Lenton, 2001, and Mitchell et al., 2001, for examples).[3]

In six experiments the primary features of the GNAT were examined. In the first five, the main purpose was to identify the parameters under which the GNAT performs robustly. In Experiment 1, we tested whether the GNAT is able to assess automatic attitudes in a format parallel to the IAT, with the exception that sensitivity was adopted as the dependent variable to test its viability as an alternative to response latency. In addition, we tested variations in the time to respond to the stimuli (response deadline) and its influence on the magnitude of the obtained automatic evaluation effect. In Experiments 2-4 we systematically varied the distracters that served as the context and its consequence on automatic evaluation. In particular, we examined automatic evaluations of target categories in (1) single category, (2) generic category, (3) superordinate category, and (4) attribute-only (no distracter category) contexts. In Experiment 5, we sought additional evidence for the automaticity of the measured preferences by changing the dependent variable to a traditionally used measure of implicit social cognition - response latency. The attitude objects selected were fruit and bugs because they are familiar and likely to produce strong, opposing evaluations. Also, use of fruit and bugs as target categories allowed tests of the technique with attitude objects known to be simple and likely to produce consistent evaluations across subjects irrespective of their own social group membership. Finally, in Experiment 6 we applied the GNAT to examine automatic race and gender evaluation with a focus on favoritism versus derogation of in- and out-groups. See Table 1.

3. Though the ostensible purpose of many measures of preference is to assess absolute preference for an attitude object, an absolute measure of preference in the truest sense may be a difficult, if not impossible, goal. Schwarz et al. (1998) outlined a number of contextual effects on the reporting of attitudes suggesting that any attitude report is likely to be filtered through the contextual features of the reporting situation. For example, the order of questions can have dramatic effects the obtained results (see Bradburn, 1983, for a review). The GNAT allows direct measure of a single attitude object, but that does not mean that evaluation of that attitude object is independent (i.e., not affected by contextual features of the evaluation scenario). Experiments 2-5 address this point directly.

TABLE 1. Summary of Procedural Variations and GNAT Designs for Experiments 1–6

	Target (Signal)	Distracters (Noise)	Response deadline	Ratio of signal to noise	ISI	Critical Trials
Experiment 1	Fruit	Bugs	1000, 833, 666, 500	1:1	150	40
	Bugs	Fruit				
Experiment 2a	Fruit	Bugs	833, 666	1:1	150	40
	Bugs	Fruit				
	Fruit	General				
	Bugs	General				
Experiment 2b	Fruit	Bugs	833, 666	4:3	150	49
	Bugs	Fruit				
	Fruit	Food				
	Bugs	Animals				
Experiment 3	Fruit	Bugs	750, 666, 550	4:3	300	49
(single–category)	Bugs	Fruit				
Experiment 3	Fruit	Food	750, 666, 550	4:3	300	49
(superordinate)	Bugs	Animals				
Experiment 3	Fruit	none	750, 666, 550	4:3	300	49
(attribute–only)	Bugs	none				
Experiment 4	Fruit	General (+)	833, 666	1:1	550	40
	Bugs	General (+)				
	Fruit	General (–)				
	Bugs	General (–)				
Experiment 5	Fruit	Bugs	1000, 833, 666, 500 (2× for signal)	1:1	150	40
	Bugs	Fruit				
Experiment 6	Black	Humans	600, 500	1:1	300	60
	White	Humans				
	Males	Females				
	Females	Males				
	Black Females	Humans				
	Black Males	Humans				

EXPERIMENT 1: FIRST TEST

The goal of Experiment 1 was to show that the GNAT could measure automatic evaluation for categories with known preferences (fruit and bugs). This initial demonstration was designed to be a conceptual analogue of the IAT in which evaluations of one category are assessed in relative contrast to a second category. While the presence of only two categories in the GNAT is conceptually similar to the IAT, the measurement of automatic evaluation differs. In the GNAT, participants actively respond to only one category and one evaluative attribute (e.g., *fruit* + *good*) while members of other categories and evaluation are ignored. In Experiment 1, the distracters in any given block (e.g., when *fruit* + *good* were the target categories) were members of the opposing category or evaluation (e.g., bugs or bad). Unlike the typical measure of automatic evaluation the dependent variable was not latency to respond but the ability to respond within a response deadline. The final score was a measure of the accuracy with which categorizations were made in the two pairing conditions.

Use of a response deadline has been shown to effectively capture unconscious cognition even when the underlying effects are small as in subliminal priming (Draine & Greenwald, 1998). The length of the response deadline is certain to influence overall accuracy and sensitivity, but sensitivity differences between *category* + *evaluation* pairings (e.g., *fruit* + *good* vs. *fruit* + *bad*) are of interest here. If sensitivity as a measure of automatic evaluation differs with the stringency of the response deadline, the optimal response deadlines need to be explored, and this variation was included in Experiment 1.

METHOD

SUBJECTS

Subjects were 12 undergraduate students enrolled in introductory psychology who participated for partial course credit. Subjects completed the experiment individually in experimental rooms equipped with Pentium-based Windows computers. One subject was removed for making excessive errors ($d' < 0$); leaving 11 subjects for data analysis.

MATERIALS

The categorization tasks used 96 stimulus words: 24 names of fruit, 24 names of bugs, 24 terms connoting pleasant meaning, 24 terms connoting unpleasant meaning. Pleasant and unpleasant terms were selected from

norms provided by Bellezza, Greenwald, and Banaji (1986) and synonyms of those terms. Fruit and bugs names were selected from category lists provided by Battig and Montague (1969) and supplemented by the experimenters to represent easily recognizable instances of each category. (A complete list of stimuli for all experiments is in the Appendix.)

PROCEDURE

Trial Blocks. Each GNAT consisted of two blocks. In one block, the target category (e.g., fruit) was paired with an attribute, good. In the other block, the target category was paired with the opposing attribute, bad. Each block consisted of 56 trials. The first 16 trials were practice trials and were followed by a reminder screen before the subject completed the 40 critical trials. Trials began with the appearance of a single stimulus item from one of the four categories (fruit, bugs, good, bad). Target category labels appeared and remained on the screen in the upper left and right quadrants as reminders of the target category and/or target attribute for that block. Subjects were instructed to either (1) press the space bar as quickly as possible for items belonging to either of the labeled categories (go), or (2) do nothing for items that did not belong (no-go). A subsequent trial began when the subject hit the space bar, or the response deadline was reached, whichever came first.

In Experiment 1, the distracter trials (noise) were items from the alternate category (i.e., when fruit was signal, bugs was noise) and the alternate attribute (e.g., when good was signal, bad was noise).

Response Deadline. Subjects could categorize an item at any time during the brief duration it remained on the screen. The response deadline was manipulated as a within subjects variable. All subjects performed four blocks with deadlines of 1000 milliseconds (ms), 833 ms, 666 ms, and 500 ms. A 150 ms interstimulus interval (ISI) separated the end of a trial (and disappearance of the stimulus item) and the beginning of the next trial.

Trials where noise items were incorrectly categorized as signal (false alarms) or signal items were not categorized (misses) were scored as errors. For those trials, a red "X" appeared below the stimulus item during the interitem interval for 100 ms to provide continuing feedback about performance accuracy. Trials where signal items were correctly responded to as signal (hits) or noise items were not correctly responded to (correct rejections) were noted as correct responses with a green "O" appearing instead of the red "X."

GNAT Procedure. Two practice blocks to acquaint the subject with the task were presented first. In these practice blocks subjects learned to dis-

criminate between two categories - fruit from bugs and good from bad. These practice blocks consisted of 30 trials (1/2 targets, 1/2 distracters) and used a 1000 ms response deadline.

After the practice blocks, subjects completed eight GNATs, with each GNAT consisting of two blocks. Four GNATs measuring automatic attitudes toward fruit and an additional four measuring automatic attitudes toward bugs were completed, each using a different response deadline (1000 ms, 833 ms, 666 ms, and 500 ms). The GNATs were presented in a partially randomized block design. Subjects completed all blocks (i.e., *fruit + good*, *fruit + bad*, *bugs + good*, *bugs + bad*) with a longer deadline (e.g., 1000 ms), before proceeding to the blocks at the next shortest deadline (i.e., 833 ms) so as to increase task difficulty set by response deadline over time. Within each deadline, the order of the blocks was randomized.[4] After completing the GNATs, subjects completed a short questionnaire assessing explicit preferences for fruit and bugs on a temperature scale (0-100 rating of feelings of coldness or warmth toward the attitude object) and provided minimal demographic information. Finally, subjects were debriefed and thanked.

Stimuli. Words were selected randomly and without replacement until the list items available for a task were exhausted. For each block, equal numbers of items were selected from the four concepts (fruit, bugs, good, bad). The label and stimuli for the target and distracter categories (fruit, bugs) were presented in blue (on a black background), and the label and stimuli for the target and distracter attributes (good, bad) were presented in white. Stimuli appeared in the center of the screen and remained on the screen until the response deadline was reached.

Analysis of SDT. Calculation of d-prime (d') proceeded according to the approach defined by Green and Swets (1966). Empty cells (i.e., no false alarms or misses) present a problem for calculation of sensitivity effects and require a correction. We applied the approach recommended by Banaji & Greenwald (1995), resulting in a correction of 0.35, divided by the number of trials, to empty cells. In signal detection, sensitivity is calculated with the following algorithm: (1) the proportion of hits (correct "go" response for signal items) and false alarms (incorrect "go" response for noise items) are each converted to *z*-scores; (2) a difference

4. We randomized blocks assuming that, like it does in the IAT, the order of blocks would impact performance on the GNAT (e.g., performing *fruit + good* before *fruit + bad* may worsen performance on the latter block; Greenwald et al., 1998). However, as a reviewer pointed out, we have not tested this assumption. Blocks for all target category + attribute pairs were randomized within response deadlines in every experiment. Testing this assumption will require an experiment that counterbalances the order of blocks for a single target category.

between the *z*-score values for hits and false alarms is d'. D-prime values of 0 or below indicate that subjects were either unable to discriminate any signal from noise or were not performing the task as instructed. As such, blocks with sensitivity scores of 0 or below (0 is chance responding) were removed from the analysis.[5]

RESULTS

AUTOMATIC ATTITUDES TOWARD FRUIT AND BUGS USING SENSITIVITY

Sensitivity, indexed by d-prime (d') in SDT, indicates the ability to discriminate targets (signal) from distracters (noise). The assumption underlying the use of d' in the present experiments is that subjects ought to be more sensitive (i.e., discriminate signal from noise more easily) when the two components of the signal are positively associated relative to when they are not associated (or are negatively associated). To the extent that fruit is automatically associated with good and not with bad, performance ought to be better when jointly discriminating fruit and good from distracters than jointly discriminating fruit and bad. Likewise, to the extent that bugs are automatically associated with bad, performance ought to be better when jointly discriminating bugs and bad from distracters, than bugs and good from distracters. Greater sensitivity indicates a stronger association between the target category and attribute. This association is defined to be a measure of automatic attitude toward the target category.

Sensitivity (d') was calculated by combining data for each pairing (e.g., *fruit + good*) over the four response deadlines. As expected, subjects showed greater sensitivity when fruit and good were signal (d' = 2.77) than when fruit and bad were signal (d' = 1.65; $t(10) = 8.51$, $p < .0001$, Cohen's $d = 2.69$) suggesting that fruit and good are more strongly associated than fruit and bad. In other words, subjects showed a positive automatic attitude toward fruit. When the target concept (signal) was bugs, the opposite pattern was observed. Subjects showed greater sensitivity to *bugs + bad* (d' = 2.69) than to *bugs + good* (d' = 1.8; $t(10) = -5.8$, $p = .0002$, $d = -1.83$) suggesting a negative automatic attitude toward bugs. The GNAT effectively discriminated automatic positive preferences for fruit and negative preferences for bugs. Importantly, separable sensitivity measures were taken for each attitude object (fruit and bugs).

5. More information about the effects of bias (ß) in the GNAT is available from the first author.

In addition, sensitivity did not vary by the target concept (fruit or bugs; $F(1,10) = .37$, $p = .55$) or by the target evaluative category (good or bad; $F(1,10) = 1.2$, $p = .30$) indicating that the sensitivity score was uniquely a reflection of the association between category and evaluation.

Variation in Automatic Attitudes Across Response Deadlines. With more time to make a decision and initiate a response, better discrimination between signal and noise should result. As such we should observe greater overall sensitivity as a function of the response deadline. More importantly, we tested whether variation in overall sensitivity due to differences in response deadline would alter the observed automatic attitude as assessed by the difference in sensitivity between blocks (e.g., *bugs + good* versus *bugs + bad*).

As expected, longer response deadlines were associated with fewer errors and greater average sensitivity than shorter response deadlines (Error rates: 1000 ms = 6%, 833 ms = 9%, 666ms = 16%, 500 ms = 29%; Average sensitivity (d'): 1000 ms = 3.30, 833ms = 2.97, 666 ms = 2.34, 500 ms = 1.30). Despite the large differences in average sensitivity across response deadline conditions, response deadline did not systematically affect sensitivity between target pairings (see Table 2). These data suggest that the GNAT allows detection of automatic attitude effects at a variety of response deadlines.

Although an effect was observable with the 1000 ms response deadline, the overall error rate was just 6%, and many subjects were able to respond without error on one or more blocks. Ceiling effects of this kind are generally undesirable because they can mask smaller effects that might be detected with a deadline that elicits greater variability in accuracy. Therefore, in the remaining experiments we typically employed deadlines of 500-850 milliseconds to minimize ceiling effects in error rates (and also avoided response deadlines so short that performance could not exceed chance level accuracy).

Recommendation. The GNAT offers a measure of automatic attitudes toward target categories using a single contrasting category. The most effective response deadlines for measuring automatic cognition are those fast enough to eliminate perfect responding but not so fast as to lower accuracy substantially, with an appropriate range of response deadlines falling between 500 and 850 milliseconds.

OVERVIEW OF EXPERIMENTS 2A, 2B, 3 AND 4: TESTS OF CONTEXT

A unique feature of the GNAT is the flexibility of the contextual frame in which the target category is evaluated. The distracter items comprise

TABLE 2. Sensitivity Calculations for Experiment 1 Separated by Response Deadline and Target Concept. Larger (Positive Or Negative) D–Prime Values Indicate Greater Sensitivity to Category Pairing. Negative Values for Cohen's *d* Indicate Negative Evaluations of the Target Category. Standard Deviations are in Parentheses.

Deadline	Target	Good	Bad	Cohen's *d*	Significance test
1000	Fruit	3.64 (.55)	2.9 (.73)	1.03	$F(1,10) = 10.51$, $p = .009$
1000	Bugs	3.06 (.51)	3.61 (.58)	–0.88	$F(1,10) = 7.73$, $p = .019$
833	Fruit	3.44 (.63)	2.41 (.92)	1.05	$F(1,10) = 11.00$, $p = .008$
833	Bugs	2.43 (.85)	3.6 (.54)	–2.26	$F(1,10) = 51.01$, $p < .0001$
666	Fruit	2.83 (.74)	1.61 (.64)	1.16	$F(1,10) = 13.43$, $p = .004$
666	Bugs	1.97 (1.03)	2.95 (.89)	–1.17	$F(1,10) = 13.73$, $p = .004$
500	Fruit	2.04 (.92)	0.69 (.63)	1.45	$F(1,10) = 21.07$, $p = .001$
500	Bugs	0.87 (.75)	1.61 (.80)	–0.69	$F(1,10) = 4.81$, $p = .053$
combined	Fruit	2.7 (.40)	1.64 (.49)	2.78	$F(1,10) = 77.15$, $p < .0001$
combined	Bugs	1.82 (.52)	2.61 (.51)	–1.79	$F(1,10) = 31.91$, $p = .0002$

this contextual frame and can consist of a single comparison category (e.g., bugs for the target category fruit and vice versa, as in Experiment 1), which is comparable to the form of an IAT or include multiple categories as the context. Because the flexibility of what constitutes the context is an important feature of the GNAT, we conducted four experiments to test this feature.

The procedure for each of these experiments holds closely to the method described in Experiment 1 with specific changes to some of the critical features of the task. Table 1 presents a summary of these procedural variables and their differences across experiments. In Experiment 2a, automatic attitudes for fruit and bugs were tested when the distracters were a set of generic items (e.g., bookshelf, coffee, gem) that together did not constitute a clearly identifiable category (generic context). In Experiment 2b, we tested whether automatic evaluations can be assessed against a context more relevant to the target category - members of a common superordinate category (superordinate context). For fruit, items denoting other types of food (e.g., beef, broccoli, butter) represented the superordinate context. For bugs, items denoting other types of animals (e.g., salamander, shark, sparrow) represented the superordinate context. In Experiment 3, we investigated whether automatic attitude effects could be observed even when only evaluative attribute items (e.g., bad), but no concept items, served as distracters for the target pairing (i.e., *fruit* + *good*) to approximate a pristine measure of a single attitude object without the presence of any other attitude object (attribute-only context). This is a critical experiment because, if successful, it shows that no contrasting category need be present in the measurement context. To our knowledge, this is the first time that implicit evaluations toward a category have been assessed without the presence of any other category in the measurement context. In Experiment 4, we tested whether having a distracter set that is distinctly valenced (either positively or negatively) plays a role in evaluation of the target concept (positive-category or negative-category context).

EXPERIMENT 2A: SINGLE AND GENERIC CONTEXT

In Experiment 1, (N = 26) we demonstrated that automatic attitudes could be measured when the context was a single contrasting category (single-category context). In Experiment 2a, we sought to replicate and extend this finding with a more general distracter set to demonstrate one of the primary features of the GNAT - the flexibility in creating the contextual category. In this case, distracter items were selected to be a set of ge-

neric items (generic context) rather than a single category as is typical in the IAT.

METHOD

PROCEDURE

The procedure was the same as in Experiment 1 with two changes. First, two response windows were used (833, 666 ms) instead of four to simplify the design. Second, the distracter items were varied to test the flexibility of the GNAT. Half of the blocks were identical to Experiment 1 (single-category context) - half of the distracter items were the attribute opposite to the target attribute (good or bad); the other half of the items represented a single comparison category (e.g., bugs when fruit was the signal). The other half of the blocks used a generic context. Distracter items (e.g., table, potato, car) were taken from a variety of categories and did not share an obvious common category, and were not systematically positively or negatively valenced (from Battig & Montague, 1969).

Subjects completed eight GNATs for a total of 16 blocks. The four GNATs with a response window of 833 milliseconds were presented first in random order. Two of those had the single-category context; the other two had the generic context. Testing of both contexts allowed a direct comparison of the single opposing category form of measurement (similar to the form of the IAT) with a more general context for automatic attitude measurement. The four GNATs with a response window of 666 milliseconds were presented second in random order. Two of those were implemented with the single-category context; the other two with the generic context.

RESULTS

Comparing Automatic Attitudes in Single-Category and Generic Contexts. Data were collapsed across the two response deadlines to compare the effect of changing the distracter items from a single category (of opposing valence) to a generic category (of neutral valence). We submitted these data to a 2 (single-category vs. generic context) × 2 (bugs or fruit target category) × 2 (good or bad target attribute) ANOVA, with the last two variables serving as within-subjects variables.

The primary finding from Experiment 1 was replicated; the target category by target attribute interaction was significant indicating that sensitivity was stronger for *bugs + bad* and *fruit + good* than *bugs + good* and *fruit + bad* ($F(1,24) = 76.6$, $p < .0001$, $d = 1.79$). That is, fruit elicited

positive automatic attitudes, bugs elicited negative automatic attitudes. However, type of context did have an impact on the observed sensitivity effects. Subjects showed greater overall sensitivity with a generic context than with a single-category context ($F(1,24) = 21.1$, $p = .0001$, $d = 0.94$). A three-way interaction qualified this effect indicating that sensitivity differences between target evaluations were stronger for both target categories when the context was a single category compared to generic ($F(1,24) = 13.2$, $p = .002$, $d = 0.74$). In sum, the strength of sensitivity effects for the GNAT are moderated when the context is changed from one that includes a semantically and evaluatively opposing set of stimuli (single-category context) to a set that includes a semantically and evaluatively mixed set of stimuli (generic context). The impact of the evaluative quality of the distracter set on GNAT effects will be directly investigated in Experiment 4.

EXPERIMENT 2B: SUPERORDINATE CONTEXT

Experiment 2b ($N = 21$) introduced yet another variation in context, this time, one representing a superordinate category (food for the category fruit, and animals for the category bugs). Such a specification diverges from the single category context and the generic context by creating a contrast between the attitude object and distinct instances that also belong within the superordinate category.

In addition, a methodological issue concerning the ratio of signal-to-noise was considered. In the previous experiments, half the trials contained signal and the other half noise stimuli. Increasing the proportion of signal-to-noise could affect evaluation by increasing the focus of attention on target items. As a variation, the ratio of signal-to-noise was increased to emphasize the target category and attribute and to observe its effects on the magnitude of the automatic attitude effect. To test this variation we compared results of the single-category context in this experiment to the results of the single-category context in Experiment 2a that were identical with the exception of the signal-to-noise ratio.

METHOD

PROCEDURE

Experiment 2b was identical to Experiment 2a with two changes. First, instead of a generic context, a superordinate context was used. For fruit, all distracter items were foods. For bugs, all distracter items were animals.

Second, in the critical blocks, proportionally more items from the target category and attribute were presented than from the distracter category and attribute to observe its role in the magnitude of the attitude effect. For example, in the critical block pairing *fruit + good*, 2/7 of the presented items were fruit, 2/7 were from the category good, 2/7 were from the category bad, and 1/7 was from the category bugs. In all cases, target items were presented more frequently the distracter items by a 4:3 ratio, and evaluative distracters were presented twice as frequently as category distracters.

RESULTS

Comparing Automatic Attitudes in Single-Category and Superordinate Contexts. Results for Experiment 2b replicated the effects observed in Experiment 2a (see Table 3). Even more robustly than obtained before, an interaction between target category (fruit vs. bugs) and target attribute (good vs. bad) indicated that subjects were more sensitive to *fruit + good* and *bugs + bad* pairs compared to the opposite pairs ($F(1,20) = 149.1$, $p < .0001$, $d = 2.73$). That is, fruit elicited positive automatic attitudes and bugs elicited negative automatic attitudes. Greater overall sensitivity was observed in the superordinate context compared to the single-category context ($F(1,20) = 11.6$, $p = .003$, $d = 0.76$). Finally, these effects were qualified by a three-way interaction of context (superordinate vs. single-category), target category (fruit vs. bugs), and target attribute (good vs. bad). That is, positive attitudes toward fruit and negative attitudes toward bugs were more evident in a single-category context than in a superordinate context ($F(1,20) = 21.4$, $p = .0002$, $d = 1.03$). The differences between conditions in these data, like those in Experiment 2a, may be due to the manipulation of context conditions within subjects. We will address this possibility in Experiment 3.

Ratio of Targets to Distracter Stimuli. A secondary goal of Experiment 2b was to test whether variation of the signal-to-noise ratio would affect the magnitude of automatic attitude activation. The procedural details of measuring automatic attitudes toward fruit and bugs in the single-category context were identical between Experiments 2a and 2b except for the difference in proportion of signal to noise (50% signal in Experiment 2a; 58% signal in Experiment 2b). A comparison between

6. The effect sizes in Table 3 suggest that greater signal to noise ratios will magnify GNAT effects. It appears, however, that those effect sizes are due to small standard deviations in the higher signal-to-noise ratio data (Experiment 3), a notable effect in its own right. Additional investigations will be needed to determine whether variation in signal-to-noise ratios has a direct impact on variability in GNAT effects.

TABLE 3. Sensitivity Calculations for Experiments 2–5 Separated by Target Concept and Distracter Set. Larger (Positive or Negative) D–Prime Values Indicate Greater Sensitivity to Category Pairing. Negative Values for Cohen's *d* Indicate Negative Evaluations of the Target Category. Standard Deviations are in Parentheses.

Target	Distracters	Good	Bad	Cohen's *d*	Significance Test
Experiment 2a					
Fruit	Bugs	2.46 (.88)	1.47 (.60)	1.20	$F(1,24) = 34.61$, $p < .0001$
Bugs	Fruit	1.48 (.64)	2.36 (.86)	–1.39	$F(1,24) = 46.06$, $p < .0001$
Fruit	Generic	2.50 (.81)	2.14 (.78)	0.62	$F(1,24) = 9.30$, $p = .0055$
Bugs	Generic	1.98 (.73)	2.46 (.78)	–0.78	$F(1,24) = 14.46$, $p = .0009$
Experiment 2b					
Fruit	Bugs	2.57 (.66)	1.45 (.48)	2.32	$F(1,20) = 107.42$, $p < .0001$
Bugs	Fruit	1.41 (.55)	2.64 (.85)	–1.71	$F(1,20) = 58.59$, $p < .0001$
Fruit	Food	2.44 (.67)	1.99 (.61)	0.77	$F(1,20) = 11.85$, $p = .0026$
Bugs	Animals	1.98 (.63)	2.59 (.77)	–1.17	$F(1,20) = 27.29$, $p = .0001$
Experiment 3					
Fruit	Bugs	2.39 (.63)	1.24 (.40)	2.45	$F(1,13) = 78.25$, $p < .0001$
Bugs	Fruit	1.46 (.43)	2.39 (.67)	–2.23	$F(1,13) = 64.55$, $p < .0001$
Fruit	Generic	2.11 (.78)	1.36 (.67)	2.76	$F(1,14) = 106.52$, $p < .0001$
Bugs	Generic	1.39 (.54)	2.10 (.80)	–1.51	$F(1,14) = 32.00$, $p < .0001$
Fruit	none	2.02 (.48)	1.62 (.60)	0.85	$F(1,14) = 10.09$, $p = .007$
Bugs	none	1.67 (.57)	2.18 (.64)	–0.79	$F(1,14) = 8.81$, $p = .01$
Experiment 4					
Fruit	Generic+	2.85 (.83)	2.68 (.75)	0.37	$F(1,14) = 1.95$, $p = .18$
Bugs	Generic+	2.46 (.70)	2.79 (.90)	–0.71	$F(1,14) = 7.07$, $p = .02$
Fruit	Generic–	3.06 (.85)	2.18 (1.06)	1.15	$F(1,14) = 18.36$, $p = .0008$
Bugs	Generic–	2.40 (1.00)	2.71 (.86)	–0.51	$F(1,14) = 3.64$, $p = .08$

Fruit and Bug GNATs in the single-category context between Experiments 2a and 2b was conducted with a mixed ANOVA. Although effects appeared to be somewhat stronger with the larger proportion of signal items, no significant differences were observed to indicate that changing the signal-to-noise ratio affects automatic attitude activation $F(1,44) = 2.2$, $p = .15$.[6] While dramatic alterations to the signal-to-noise ratio may affect automatic attitude activation, this small alteration (4:3 vs. 1:1) had no effect.

EXPERIMENT 3: REMOVING THE ATTITUDE OBJECT FROM THE COMPARISON CONTEXT

In each previous experiment, some type of contrasting attitude object was always present in the measurement context as background. Both IAT and priming techniques share this quality where another category is present during evaluation of a target category as either its relative comparison (IAT) or as a different set of primes (priming). Can a meaningful measure of automatic attitude be obtained in the absence of an alternative attitude object in the measurement context? In Experiment 3 ($N = 44$ after four subjects were removed for excessive errors on the GNATs), we tested whether automatic attitudes can be observed with the GNAT when the context contained no other attitude object, but rather, contained only the evaluative attribute opposing the one that served as signal (attribute-only context). The ability to detect an effect here will demonstrate that automatic attitude effects do not require some type of contextual attitude category for assessment, and open up possibilities for attitude measurement that are exclusively focused on the target category.

A second important goal in Experiment 3 was to compare the potency of the GNAT between single-category, superordinate, and attribute-only contexts. Direct comparison in the magnitude of effects between the contexts manipulated in Experiments 2a and 2b were not possible because these were varied across experiment. In Experiment 3, all variations in context were manipulated between subjects to directly compare the relative magnitudes of their attitude effects.

6. The effect sizes in Table 3 suggest that greater signal to noise ratios will magnify GNAT effects. It appears, however, that those effect sizes are due to small standard deviations in the higher signal-to-noise ratio data (Experiment 3), a notable effect in its own right. Additional investigations will be needed to determine whether variation in signal-to-noise ratios has a direct impact on variability in GNAT effects.

METHOD

PROCEDURE

The procedure for Experiment 3 was the same as Experiment 2b with a few procedural changes summarized in Table 1.[7]

In three between-subject conditions, the distracter items were manipulated. Consistent with the previous studies, all conditions used the opposing attribute items as part of the context. In the first condition, the rest of the distracters were a specific-category (single-category context) as a replication of previous experiments. In the second condition, a superordinate context identical to that used in Experiment 2b was used. In the third condition, no distracters other than the items representing the opposing attribute were used (attribute-only context). Therefore, the proportion of opposing attribute items was increased from 2/7 to 3/7 of the total items presented. Also, items from all categories in the third condition were presented in white lettering. This was done to ensure that categorization of target categories occurred based on semantic meaning rather than stimulus color.

In summary, Experiment 3 was a mixed-design study with three within-subjects factors, two target category (fruit vs. bugs) × 2 target attribute (good vs. bad) × 3 response deadlines (750 ms, 666 ms, 550 ms), and one between-subjects factor, three contexts (single-category, superordinate, and attribute-only).

RESULTS AND DISCUSSION

Comparing Automatic Attitudes in Single-Category, Superordinate, and Attribute-only Contexts. Whether the context was a single category, a superordinate category, or just the opposing attribute, fruit elicited positive automatic attitudes and bugs elicited negative automatic attitudes (see Table 3). Yet, some of the contexts did vary in the magnitude of the elicited effects ($F(2,41) = 10.2$, $p = .0003$). Follow-up tests showed that the strongest effects were observed in the single-category context, with the superordinate context showing slightly smaller overall effects ($p = .04$). The smallest effects were apparent in the attribute-only context, which were significantly smaller than both the single-category ($p = .0002$) and the superordinate ($p = .02$) contexts. Nonetheless, the ef-

7. Some subjects in the previous experiments complained that stimulus items appeared too rapidly after the previous item had disappeared. To counter this distraction, we extended the interstimulus interval from 150 ms to 300 ms.

fects were still in the range considered strong by conventional effect size standards (for Cohen's *d* .2 = small, .5 = medium, .8 = large; Cohen, 1988). The nature of the context does have an impact on the strength of the observed effect - evaluations in the attribute-only context were smaller than either the single-category or superordinate category contexts.

Recommendations from Experiments 2 and 3. Automatic attitudes toward fruit and bugs were reliably and consistently measured with single-category, generic, superordinate, and attribute-only contexts. Results in Experiments 2 and 3 do indicate that the magnitude of automatic attitude effects may be smaller for some contexts than others, but the direction of effects was consistent across contexts that varied in semantic qualities. Decisions about the type of context to use for assessment can then be made with regard to the qualities of the attitude object. That is, attitude measurement for objects with a well-defined opposing category (e.g., men and women) can use a single-category context; attitude measurement for objects that are members of a larger family of objects can use a superordinate context (e.g., all types of cars as the context for Hondas); and, attitude measurement for objects with no obvious comparison categories can use either a generic context or an attribute-only context (e.g., attitudes toward smoking). Experiments 2a, 2b, and 3 confirm that each of these context conditions is a viable method for measuring automatic attitudes.

EXPERIMENT 4: VARYING THE VALENCE OF THE CONTEXT CATEGORY

In this experiment we focus on a particular aspect of varying the evaluative quality of the context. When the context is evaluatively closer to the foreground attitude, the attitude effects should be smaller than when the evaluative match between context stimuli and foreground are less close. The greater evaluative discrepancy between context and foreground should heighten the disparity between them and elicit stronger attitude effects.

In Experiment 4 ($N = 15$ after removing one subject for excessive errors), we examined the effect of varying the valence (positive or negative) of the concept items in the distracter set (generic context) on the activation of automatic attitudes for the target concept (fruit or bugs). When category distracter items are of an opposing valence to the target category (negative, but semantically unrelated concept items such as vinegar and platypus when fruit is the target category) the distinction between signal and noise ought to be more polarized in the compatible (*fruit + good*) condition and diminished in the incompatible

(*fruit + bad*) condition. The result should be a magnification of the attitude effect. Likewise, when category distracters are of the same valence as the target category (positive, but semantically unrelated, concept items such as diamond and neighbor when fruit is the target category), the distinction between signal and noise ought to be diminished in the compatible condition and polarized in the incompatible condition. The result should be a relative reduction of the attitude effect.

METHOD

PROCEDURE

The procedure was identical to Experiments 2b and 3 with minor procedural changes (see Table 1). Experimenters created two lists of 18 items where the items were things with positive connotations (e.g., gold, cinnamon) or negative connotations (e.g., coal, anchovy). Following administration of the GNATs, subjects rated the positivity-negativity of each of the items as a manipulation check. In summary, this Experiment had four within-subjects factors, 2 target category (fruit vs. bugs) × 2 target attribute (good vs. bad) × 2 response deadline (750 ms, 600 ms) × 2 context (positive-category vs. negative-category).

RESULTS AND DISCUSSION

Manipulation Check. As expected, subjects' conscious attitudes about the items in the positive- and negative-category contexts indicated that positive-category distracters were evaluated more positively than the negative-category distracters ($t(14) = 8.3$, $p < .0001$).

Comparing Automatic Attitudes Between Negative- and Positive-Category Contexts. The basic effects replicated Experiments 1-3, fruit was evaluated positively and bugs were evaluated negatively (see Table 3). For fruit, however, this effect was moderated by the valence of the generic items. Positive attitudes toward fruit were stronger in the negative-category context compared to the positive-category context ($F(1,14) = 12.5$, $p = .003$, $d = 0.94$). Though the magnitude of negativity toward bugs was somewhat stronger in the positive-category than the negative-category context, the difference was not significant. This suggests that the magnitude of automatic attitude activation is decreased, but not reversed, when the valence of the concepts in the contextual category matches the valence of the target concept.

Recommendation. Valence of the context does play a role in the magnitude of the evaluations observed with the GNAT. If the concepts in the

context are selected to match the valence of the target category, a smaller GNAT effect will be observed compared to selection of an opposing valence set of context items. In general, experimenters will want to use evaluatively mixed or neutral concepts as the context to avoid missing effects because of evaluative matching between target and context, or inflating effects because of evaluative mismatching between target and context.

EXPERIMENT 5: RESPONSE LATENCY

In Experiments 1-4, signal detection analysis capitalized on errant responses (or nonresponses) to assess automatic attitudes. However, most other measures of automatic association such as the IAT and evaluative priming use response latency as the dependent variable. In Experiment 5 ($N = 12$), we modified the GNAT to conceptually replicate earlier findings with response latency to provide a dependent variable measure that is comparable to existing measures and to test the adaptation of the GNAT to this feature.

METHOD

PROCEDURE

Experiment 5 was identical to Experiment 1 with a single substantive change. To maximize the range available for response latency measurement (and avoid a restricted range of response latency), the response deadline for target trials was twice that for distracter trials. For example, in the 1000 ms response deadline condition, target trials actually had a 2000 ms response deadline. The response deadline for target items was extended to minimize errors and maximize the range of possible response times. The deadline was not extended for distracter items because the task requires some pressure to respond quickly. If the items all appear for an extended period of time, subjects could intentionally slow down and decrease the automaticity of their responses. Instructions were identical to those in Experiment 1 and no subject noticed the difference in deadline between the target and distracter trials.

ANALYSIS

Data for all errors and distracter items were removed before analysis. Only correct responses to target items were used. Average response latency was calculated on the remaining items for each block. Stronger asso-

TABLE 4. Average Response Latencies for Category Pairings for Experiment 5 Separated by Response Deadline and Target Concept. Lower (Faster) Response Latencies Indicate Stronger Association Between Target Concept and Attribute. Negative Values for Cohen's *d* Indicate Negative Evaluations of Target Categories. Standard Deviations are in Parentheses.

Timeout	Signal	Good	Bad	Cohen's *d*	Significance Test
1000	Fruit	627 (109)	684 (118)	0.89	$F(1,11) = 8.77$, $p = .013$
1000	Bugs	673 (102)	626 (117)	–0.47	$F(1,11) = 2.42$, $p = .15$
833	Fruit	619 (125)	697 (103)	0.96	$F(1,11) = 10.19$, $p = .009$
833	Bugs	677 (125)	600 (104)	–0.81	$F(1,11) = 7.24$, $p = .021$
666	Fruit	565 (79)	661 (98)	1.20	$F(1,11) = 15.83$, $p = .002$
666	Bugs	640 (112)	588 (94)	–0.87	$F(1,11) = 8.41$, $p = .014$
500	Fruit	593 (94)	668 (82)	1.20	$F(1,11) = 15.93$, $p = .002$
500	Bugs	660 (64)	605 (80)	–1.00	$F(1,11) = 10.95$, $p = .007$
overall	Fruit	601 (82)	678 (88)	2.09	$F(1,11) = 47.95$, $p < .0001$
overall	Bugs	662 (75)	605 (84)	–1.55	$F(1,11) = 26.39$, $p = .0003$

ciations between target concept and evaluation are indexed by faster average response latency in that block. For example, if subjects have a positive automatic attitude toward fruit then they should be able to respond more quickly to *fruit* + *good* target pairings versus *fruit* + *bad* pairings.

RESULTS

MEASURING AUTOMATIC ATTITUDES TOWARD FRUIT AND BUGS WITH RESPONSE LATENCY

Experiment 5 replicated all major findings of Experiment 1 using response latency rather than sensitivity as the dependent variable. Overall effects were calculated by average response times across blocks with the four different response deadlines. Subjects were faster at responding to target words when the categories were *fruit* + *good* ($M = 601$) than when the categories were *fruit* + *bad* ($M = 678$; $F(1, 11) = 47.95$, $p < .0001$, $d = 2.09$; see Table 4) indicating a positive automatic attitude toward fruit. Also, subjects were faster at responding to target words when the categories were *bugs* + *bad* ($M = 605$) than when the categories were *bugs* + *good* ($M = 662$; $F(1, 11) = 26.39$, $p = .0003$. $d = 1.55$) indicating a negative automatic attitude toward bugs. These effects were consistent across the various response deadlines (Table 4). As with signal detection, when using response latency as a dependent variable the GNAT effectively measured automatic attitudes toward fruit and bugs.

Recommendation. At least two dependent variables are available to the researcher using the GNAT. Both sensitivity and response latency are effective in assessing automatic cognition. These data do not suggest whether one is more effective than the other. At this point, decisions regarding which measure to use should be based on practical concerns in individual experimental designs. Since each trial in the GNAT provides a dichotomous data point (hit or miss, false alarm or correct rejection) in sensitivity, but a continuous data point (milliseconds) in response latency, it is likely that use of response latency will result in greater internal reliability (see Discussion). Since signal detection derives from analysis of errors, the signal detection GNAT can be frustrating for perfectionists (or even pretty-goodists) to perform. Using response latency as the operational dependent variable may have a pragmatic advantage of not irritating the subjects, as long as the response deadline is sufficiently long to enable accurate responding. The advantages of one dependent variable over the other require additional investigation.

EXPERIMENT 6: RACE AND GENDER ATTITUDES

Thus far we have examined automatic attitudes toward rather mundane objects with limited implications for social life. In Experiment 6, we applied this technique to attitude objects that carry greater social weight — race and gender attitudes. Since other existing procedures have predominately examined social groups, this test also provides a more direct comparison to those techniques.

Race and Gender Attitudes. Investigations of implicit attitudes toward racial groups consistently demonstrate negativity toward black targets (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio et al., 1986, 1995; Greenwald et al., 1998; Nosek et al, in press). The GNAT affords an opportunity to examine attitudes toward specific attitude objects (e.g., black American) without requiring a direct relative comparison to other groups. This opportunity allows a test of whether observed differences in relative evaluation are primarily a function of in-group preference, out-group derogation, or both, and this question is of some theoretical importance. Brewer (2001) for example has pointed out that contrary to previous thinking (Sumner, 1906) a positive attitude toward the in-group need not imply a negative attitude toward the out-group. That is, strong positivity toward white Americans need not result in opposite and commensurate negativity toward black Americans.

We also examined the in-group and out-group components of gender attitudes. Despite the disadvantages women face in social, political, and economic life, research on implicit and explicit attitudes

toward the group show positivity (Carpenter & Banaji, 2000; Eagly & Mladnic, 1994; Lemm & Banaji, 1999). We can further the investigation of implicit gender attitudes by using the GNAT to look at attitudes toward males and females separately to determine whether the relative comparison is due to female positivity, male negativity, or both. This domain also makes it relatively easy to compare assessments by both group members, given the strong gender difference observed in implicit gender attitudes (Carpenter & Banaji, 2000; Lemm & Banaji, 1999).

Conjoint Groups. Little research in implicit social cognition has investigated groups that are the conjoint of two separate categories such as black females and black males. Yet, there is no reason to assume that subgroups of larger social groups (e.g., race/ethnicity) are necessarily evaluated similarly to the larger group. Evaluations of white Americans in general may produce a particular attitude that is dissimilar to evaluations of Italian Americans and Polish Americans in particular, just as a baseball fan may show positive evaluations of baseball players as a group, but negative evaluations of the Los Angeles Dodgers in particular.

Also, evaluations of conjoint groups need not necessarily reflect the average of the constituent groups. For example, babies are attitudinally positive, corpses are negative, but baby corpses are not expected to be evaluatively neutral. Unanswered by existing research is the question of how members of conjoint social groups are evaluated. In this experiment, we begin an investigation of conjoint groups by using the GNAT to measure automatic attitudes toward black males and black females with the hope that such investigations will be taken up in their own right in the future.

In addition to the conceptual issues advanced in this experiment, a number of methodological features of the GNAT were also addressed. First, a direct comparison to the IAT was made to test if they produce comparable results. Second, we used pictorial stimuli (faces) instead of words as stimuli. For social groups these have been regarded as the ecologically more valid exemplars and they do not suffer from confounds contained in names (i.e., names denoting race also contain socioeconomic class information, not just race). Because the previous experiments used verbal stimuli, the use of novel stimuli in this experiment also allowed a test of the generality of the GNAT with novel stimuli. Finally, we compare GNAT and IAT effects with explicit preferences to see whether they show a consistent relationship (or lack of relationship) with self-reported attitudes.

METHOD

SUBJECTS

Subjects were 53 undergraduate students enrolled in introductory psychology, who participated for partial course credit. Subjects completed the experiment individually in experimental rooms equipped with a Pentium-based Windows computer. Three subjects were removed for not following instructions or making excessive errors on the implicit measures, leaving 50 subjects for data analysis. Of the 50 remaining subjects, 27 were female and 23 were male. Also, 33 were white, 2 black, 10 Asian, 3 Hispanic, and 2 chose not to report their ethnicity.

MATERIALS

Faces. Black, white, and Asian faces (94 total; 21 males and 21 females of both blacks and whites, and 5 Asian males and 5 Asian females) were selected from two sources - smiling game portraits (head shots) of current or former NBA and WNBA players and coaches and media guide portraits of members of a collegiate athletic program at a California university. To minimize the likelihood that subjects would recognize faces presented in the Experiment, we selected NBA and WNBA players who were relatively unknown. During debriefing, subjects were systematically interviewed to probe for recognition of the presented faces. Nine subjects recognized one or more of the faces as an athlete or were able to give a specific name. Results were unaffected by whether or not subjects recognized one or more faces, so all subjects were retained.

PROCEDURE

Subjects completed three different types of tasks: two IATs (black/white attitude and male/female attitude), six GNATs, and explicit measures. Tasks were presented in a counterbalanced order across subjects. Half of the subjects completed explicit measures (temperature ratings) before the implicit measures (IATs and GNATs), the other half performed the implicit measures before the explicit measures. Likewise, half of the subjects performed the IATs before the GNATs, the other half performed the GNATs before the IATs.

GNAT. Subjects performed six GNATs - black attitude, white attitude, black male attitude, black female attitude, male attitude, and female atti-

tude. All tasks used the same set of faces but varied which faces were target items and which were distracter items.

Pretesting indicated that subjects were able to process faces more quickly than words. To ensure that subjects made enough errors to allow effective use of signal detection techniques, the deadlines were set to 600 ms and 500 ms. Also, to increase the stability of the GNAT effect, more trials were included in each task. Subjects performed 16 practice trials for each pairing and 60 critical trials. In addition, subjects had four practice blocks of 20 trials (850 ms deadline) to warm up with the task. In those four blocks subjects discriminated good from bad, bad from good, black from white, and white from black.

IAT. Subjects also performed two IATs: black/white attitude and male/female attitude. These IATs used the same stimuli that were used in the GNATs but without the Asian faces. Otherwise, the IAT procedure was identical to that employed at the IAT website (Nosek et al., in press).

RESULTS AND DISCUSSION

IAT

Replicating previous research on black-white attitudes and the IAT, subjects showed an automatic preference for white over black faces ($F(1, 49) = 24.4$, $p < .0001$, $d = 0.71$). Also, as observed before, men and women did not differ in such preference. Replicating previous work on gender attitudes, subjects showed an automatic preference for females over males ($F(1, 49) = 10.9$, $p = .002$, $d = 0.47$). Also as observed before, women showed a strong preference for females over males ($d = .98$), men showed only a slight preference for their group ($d = -.26$), and the subject gender difference was significant and strong ($F(1, 49) = 23.1$, $p < .0001$, $d = 1.37$). Reassuringly, these results replicate previous work using the IAT procedure to test race and gender attitudes (see Carpenter & Banaji, 2000; Greenwald et al., 1998; Mitchell et al., 2001; Nosek et al., in press).

GNAT

Race. Turning our attention to the GNAT for comparison, we found race attitudes to be in line with those obtained on the IAT, with positive automatic attitudes toward the group *white* and negative automatic attitudes toward the group *black*. Subjects evaluated white faces more positively than black faces ($F(1, 49) = 42.4$, $p < .0001$, $d = 0.93$). Specifically, sensitivity was greater for *black* + *bad* ($d' = 2.34$) than for *black* + *good* ($d' = 1.98$; $F(1, 49) = 12.8$, $p = .0008$, $d = -0.51$). Likewise, sensitivity was greater

for *white + good* (d′ = 2.26) than for *white + bad* (d′ = 1.67; $F(1, 49) = 28.7$, $p < .0001$, $d = 0.77$).[8] This finding indicates that the effects obtained in the IAT experiments are not a function of the polarity (black-white) that is in the forefront of the measurement context. The use of GNAT suggests that each group does elicit a separate and evaluatively opposing automatic attitude.[9]

Gender. Replicating the IAT results, subjects held more positive automatic attitudes toward females than toward males on the GNAT as well ($F(1, 49) = 41.4$, $p < .0001$, $d = 0.92$). Subjects showed greater sensitivity for *female + good* (d′ = 2.30) than for *female + bad* (d′ = 1.89; $F(1, 49) = 17.9$, $p = .0001$, $d = 0.60$). Likewise, subjects showed greater sensitivity for *male + bad* (d′ = 2.39) than *male + good* (d′ = 1.77; $F(1, 49) = 32.1$, $p < .0001$, $d = -0.81$). The overall preference for female has raised the question about the relative nature of the measure. Is the gender attitude detecting positive association toward female or negative association toward male? The IAT procedure had not allowed an easy separation of these attitudes. The GNAT procedure used here suggests that both attitudes are present. There is indeed a positive attitude toward females and a negative attitude toward males. However, the strong moderation of these effects by subject gender in the IAT leads us to further dissect this effect.

Examining subject gender differences in the overall attitude effects, we see that women had much stronger in-group liking and out-group derogation ($F(1, 48) = 14.4$, $p = .0004$, $d = 1.09$; see Figure 1). That is, women showed both strong positive associations to the category female ($F(1, 26) = 27.3$, $p < .0001$, $d = 1.02$) and strong negative attitudes toward males ($F(1, 26) = 29.9$, $p < .0001$, $d = -1.07$). Unlike women, men did not show strong liking for the in-group. In fact they showed a negative automatic attitude toward the category male ($F(1, 22) = 6.9$, $p = .02$, $d = -0.56$). Nevertheless, this effect was noticeably weaker than women's ($F(1, 48) = 8.4$, $p = .006$, $d = .84$). Unlike women, who showed a strong negative attitude toward the

8. A main effect was observed such that subjects were more accurate, regardless of attribute pairing, in identifying black faces than white faces ($F(1, 49) = 9.9$, $p = .003$, $d = 0.45$).

9. Women and men were equally positive in their evaluations of whites but women showed somewhat less negativity toward blacks than did men ($F(1, 48) = 3.4$, $p = .07$, $d = 0.54$). Differences that may be masked by the relative nature of the IAT measure may be observable with the GNAT. Although specific similarities and differences between evaluations resulting from techniques like the IAT and the GNAT are left to future research, it is these sorts of differences in the obtained effects that will be of greatest interpretational interest. In part, such differences are of interest because they raise doubts about the conclusion from any single measure and they point to the importance of recognizing the degree to which theory is constrained by the measurement tool. However, the marginal significance of this effect makes us cautious to interpret this specific difference until replicated.

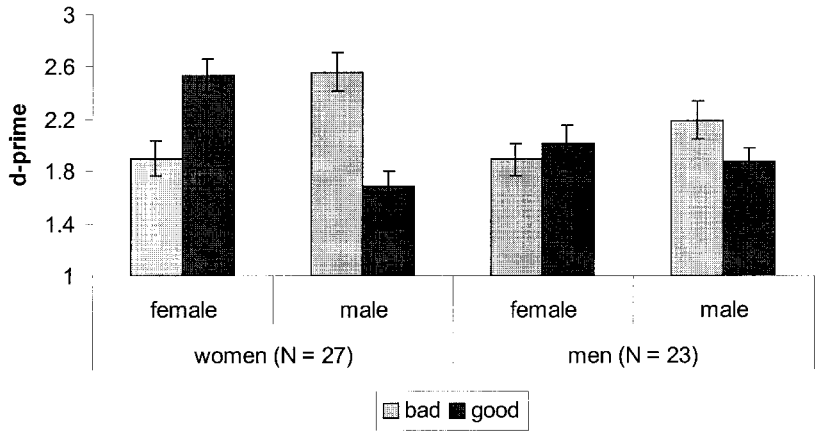

FIGURE 1. Implicit attitudes (GNATs) toward females and males separated by women and men (Experiment 6). Higher values indicate greater sensitivity to target group/attribute pairing.

out-group males, men were indifferent toward females ($F(1, 22) = 0.94, p = .34, d = 0.21$), with the mean in the direction of positivity toward female. Although such a conclusion is not easily derived from the IAT, the GNAT data suggest that men show automatic negativity toward males but remain attitudinally indifferent toward females.

This gender difference in attitudes toward female and male is surprising as it is inconsistent with gender attitudes measured via self-report (Eagly & Mladnic, 1994) in which both men and women reported positive attitudes toward females. Also, men's dislike of males is inconsistent with the expectation that being a member of a group would produce liking for that group (Tajfel, 1978, 1981). One possible explanation for this effect concerns the high proportion of black male and female faces in the sample. Although recategorization theory (Gaertner, Mann, Murrell, & Dovidio, 1989) is typically applied to the reduction of bias against out-group members by recategorizing them as in-group members (black males as male), we might also expect a reduction in positivity toward one's in-group (males) that now includes out-group members (blacks). The white male sample in this experiment may have produced more positive attitudes toward their gender group if the group was made up of white faces rather than faces of mixed ethnicity. Future research will be necessary to investigate this possibility.

Black Females and Black Males. Targets, like black females, are members of groups that have opposing evaluations (females are positive, blacks are negative; see Mitchell et al. 2001). Little research has investi-

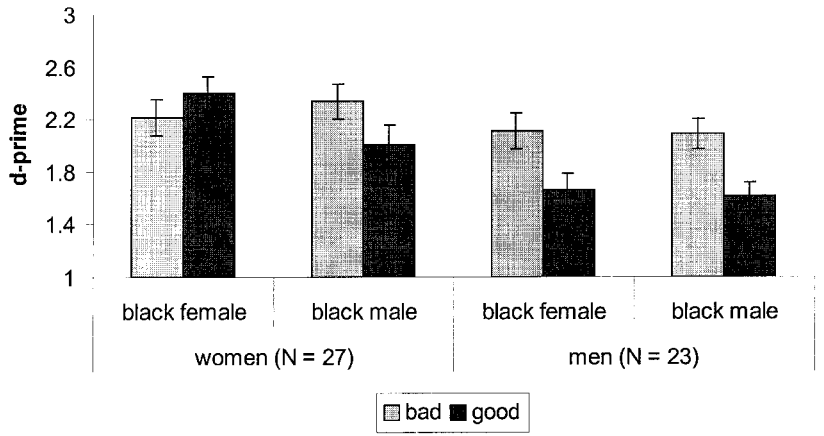

FIGURE 2. Implicit attitudes (GNATs) toward Black females and Black males separated by women and men (Experiment 6). Higher values indicate greater sensitivity to target group/attribute pairing.

gated evaluations of conjoint groups (black females) and, in particular, conjoint groups composed of categories with opposing evaluations. Are black females evaluated positively because of their gender, or negatively because of their race, or do they have a unique evaluation?

Subjects showed a slight but nonsignificant tendency to associate black females with good over bad. Black males are members of two categories that are each evaluated negatively. Consistent with evaluations of both black and males, subjects were more sensitive to *black male + bad* (d' = 2.23) than *black male + good* (d' = 1.82; $F(1, 49) = 16.8, p = .0002, d = -0.59$) indicating negative attitudes toward black males. Though both are members of a category that is viewed negatively, black females were evaluated more positively than black males ($F(1, 49) = 5.9, p = .02; d = 0.35$). These data point out that conjoint groups elicit an evaluation unique from the composing categories. Black females are neither liked as much as the superordinate category females nor disliked as much as the superordinate category black. The nature of the relationship between evaluations of conjoint groups versus the categories that constitute them is a question for future research.

To examine these data in greater detail, Figure 2 presents data for the black female and black male GNATs separated by subject sex. Men showed greater overall negativity toward black females and males than women did ($F(1, 48) = 7.4, p = .009, d = 0.79$). Also a three-way interaction emerged wherein both men and women held negative attitudes about black males, but only men held negative attitudes about black females

TABLE 5. Zero–Order Correlations between Implicit Attitudes Measured by the IAT and Implicit Attitudes Measured by the GNAT. GNAT scores reflect the difference of d′ scores between the category + good and the category + bad conditions (e.g., GNAT black effect = [d′ for *black + good*] – [d′ for *black + bad*]). Black–white and male–female GNAT scores were calculated to maximize the correspondence to the race and gender IATs. The black–white GNAT score reflects the difference between the GNAT black and GNAT white effects; the male–female GNAT score reflects the difference between the GNAT male and the GNAT female effects.

	black	white	black–white
Black–white (IAT)	.12	–.24	.27*
	male	**female**	**male–female**
Male–female (IAT)	.17	–.20	.24

*$p < .07$.

($F(1, 48) = 4.4, p = .04, d = 0.61$). This finding suggests that when the target is a member of one's own group (female) and that group is viewed positively, it could protect against negative evaluations of that group, which is consistent with recategorization theory (Gaertner et al., 1989). The GNAT will be useful for investigating the consequences of shared group membership by using multiple-category targets in which the number of features shared with the evaluator is varied.

CORRESPONDENCE AMONG IMPLICIT MEASURES

Table 5 presents relationships between IATs and GNATs designed to test similar constructs. Bosson, Swann, and Pennebaker (2000) noted that measures of implicit social cognition tend to show little correspondence with each other. Zero-order correlations in Table 5 reflect this tendency. Though all correlations are in the expected directions, their magnitudes are rather small. Cunningham, Preacher, and Banaji (2001) argued that part of the lack of correspondence among implicit measures is due to low to moderate reliability of implicit measures, and that correcting for attenuation due to unreliability will reveal correspondence that otherwise would not be observed. This point is taken up in further detail in the General Discussion.

CORRESPONDENCE BETWEEN IMPLICIT AND EXPLICIT MEASURES

A subject of intense interest in implicit social cognition research is the relationship (or lack thereof) between implicit and explicit measures of

TABLE 6. Zero–Order Correlations Between Implicit and Explicit Attitudes. Correlations are Presented for Both Preferences for a Single Target Category and as a Relative Comparison between Categories (See Table 5).

	Explicit Attitudes		
Implicit Attitudes	**Black**	**White**	**Black–White**
Black (GNAT)	.15	.34**	
White (GNAT)	–.01	.19	
Black–White (GNAT)			.02
Black–White (IAT)			.08
	Male	**Female**	**Male–Female**
Male (GNAT)	.15	–.09	
Female (GNAT)	.06	.07	
Male–Female (GNAT)			.13
Male–Female (IAT)			–.09

**$p < .05$.

preference. While early theory and data suggested that these two modes of measurement were not related (Greenwald & Banaji, 1995), recent evidence suggests that, under some conditions and for some attitude objects, implicit-explicit correspondence may be observed (Cunningham et al., 2001; Nosek et al., in press). Table 6 presents zero-order correlations between implicit measures and explicit measures of preference for racial and gender groups. No systematic correspondence was observed between implicit and explicit attitudes for either the IAT or the GNAT. While these relationships may also be obscured due to measurement error (Cunningham et al., 2001), the effects are largely consistent with observations that, at least for racial preferences, implicit-explicit correspondence is absent to small (Fazio et al., 1995; Greenwald & Banaji, 1995; Nosek et al., in press).

Recommendation. The structure of the IAT constrains evaluations to be relative comparisons between two opposing categories. The GNAT relaxes this requirement allowing separable assessment of categories. This procedural variation may be an important consideration for experimental design. As demonstrated above, the different techniques can reveal or obscure effects due to the specific constraints of the task. The variation between techniques also points out that methods of measurement constrain the development and testing of theory. Researchers interested in conjoint groups or in distinguishing in-group/out-group evaluations may find the GNAT to be useful.

GENERAL DISCUSSION

In the six preceding experiments we presented a technique (the GNAT) designed to enhance flexibility in measuring automatic cognition. In the first five experiments, we outlined some features of the GNAT showing that it could effectively measure automatic preferences toward a concept when the contextual category (distracter set) was (1) a single category, (2) generic items, (3) a superordinate category, or (4) absent. In the final experiment, we applied the task to assess implicit racial and gender attitudes and showed both implicit liking of whites and implicit derogation of blacks. Also, while the men in our sample showed equal disliking of black male and black female targets, women differentiated the groups showing negativity toward black males and slight positivity toward black females. The flexibility of this measurement technique allows evaluation of an attitude object without requiring a direct comparison to another attitude object.

Potential Applications of the GNAT. In most measurement circumstances, evaluations of target concepts are assessed without consideration of the evaluative context. In the GNAT, that context is definable and controllable by the experimenter. This feature allows for versatility in the application of the task to measure evaluations of specific attitude objects in terms of the context most relevant to their evaluation. For example, while attitudes toward groups distinguished on a single dimension can often be measured relative to one another (e.g., male vs. female, black vs. white, gay vs. straight), attitudes toward groups that are the conjoint of two features generally do not have an obvious comparison category (e.g., black females, gay men). In these cases, the evaluative context may be more appropriately the category "humans" rather than a specific subgroup. In addition, there are many concepts without an obvious comparison category that would most appropriately be assessed in a more general context (e.g., attitudes toward the United States) or with no context at all (e.g., smoking). The GNAT can be adapted to serve all of these purposes.

A related feature of the GNAT is the ability to separate natural dichotomies into their component parts. In Experiment 6, both positive evaluations of the racial in-group (whites) as well as negative evaluations of the racial out-group (blacks) were observed. Notably, while men showed equally strong derogation of blacks and liking of whites, women showed liking of whites but very little derogation of blacks. Such results, with replication, can change the current interpretation of automatic race attitudes. So far, based on techniques like evaluative priming and the IAT, the widely cited conclusion is that of a strong and equally negative race attitude among men and women. The GNAT may, in detecting dif-

ferences among groups in such an attitude, also prove to be useful in future studies in which the correlations between such effects and other behaviors are of central interest. For example, different predictions may be offered of situations in which women and men will exhibit prejudicial behavior toward blacks. Indeed, evidence from other programs of research suggests that in-group favoritism and out-group derogation may have unique histories, courses, and consequences (see Brewer & Brown, 1998; Brewer, 2001).

Blair, Ma, and Lenton (2001) recently took advantage of this feature of the GNAT to show that an imagery exercise in which participants thought about strong female leaders led to a stronger association of the concept female with strong (compared to weak) among women. Using the GNAT, Mitchell et al. (2001) manipulated the evaluative context to alter the salience of gender or race in evaluation of black female, black male, white female, and white male targets. They observed changes in the evaluation of black females and white males depending on the salience of gender or race. Black females were more positively evaluated when gender was salient compared to when race was salient. White males, on the other hand, were more negatively evaluated when gender was salient compared to when race was salient. Mitchell et al. (2001) argued that multiple attitudes exist toward targets and the nature of the evoked attitude is largely dependent on the features that are most salient during evaluation.

In the preceding sections we demonstrated the utility of the GNAT for assessing automatic evaluations for target concepts in a variety of contextual situations. That is, we showed that the GNAT could effectively assess the strength of association between a concept and the poles of an attribute dimension. The GNAT might also be used for applications in which the strength of association between two different target concepts (e.g., in-group and out-group) and a single evaluation (e.g., good) is compared. For example, Mummendey, Otten and colleagues have demonstrated that discrimination toward outgroups is more likely to be observed in allocation of positive but not negative outcomes (Mummendey, Otten, Berger, & Kessler, 2000; Otten & Mummendey, 1996). That is, the variable of interest was a relative comparison of target groups on a single attribute (positive allocations). In such circumstances, the GNAT could be used to investigate whether implicit biases are more likely to emerge in the relative association between groups to positive concepts than to negative concepts. Data from Experiment 6 suggest that implicit biases against blacks relative to whites are a function of both easier association of white with good (than black with good) and black with bad (than white with bad).

Reliability of Implicit Measures. A significant challenge for assessment of implicit social cognition is designing reliable measurement tools. As pointed out by Bosson et al. (2000), the reliability of implicit measures is far below typical standards for their explicit counterparts. Recent work shows a proliferation of examples of the context sensitivity of implicit associations, even with very mild changes to the context (Blair et al., 2001; Dasgupta, McGhee, Greenwald, & Banaji, 2000; Lowery et al., 2001; Mitchell et al., 2001). As such, implicit social cognition may be more influenced by current state than previously expected. As such, high test-retest reliability with implicit measurement may not be very likely. That does not mean, however, that higher internal reliability cannot be achieved. While the number of factors that can detract from reliable measurement with response latency, or signal detection, are numerous (e.g., distraction, reading speed, ability to respond quickly, word length, task performance strategies, attention, motivation), striving for higher internal reliability can only help to enhance the predictive utility of implicit measurement. The GNAT, as described in this paper, does not ameliorate the concerns about the reliability of implicit measures. The six tasks of Experiment 6 (the only study with a sufficient number of subjects to test internal reliability) revealed an average split-half reliability of $r = .20$, which puts the signal-detection version of the GNAT decidedly in the middle range of the modest reliability of implicit measures.[10] With approaches such as structural equation modeling to correct for attenuation of relationships due to measurement error, some correction may be made for the effects of low reliability.

There are a number of obvious ways that the reliability of the GNAT may be improved. First, the number of trials for a given GNAT can be increased providing more data points for the calculation of sensitivity.[11] Second, using a variable response deadline, instead of the fixed deadlines in these studies, can eliminate irrelevant variation due to individual differences in ability to discriminate signal from noise irrespective of the category pairings. New versions of some experimental software packages allow the experimenter to idiosyncratically define response deadlines to standardize average accuracy across subjects and potentially eliminate a large amount of irrelevant variance. Third, each data point in a signal-detection analysis is a dichotomous value - correct or in-

10. Split-half reliabilities were calculated between blocks with different response deadlines that may artificially lower reliability.

11. By increasing the signal-to-noise ratio, the number of usable trials for the response latency approach to the GNAT can be increased without increasing the overall number of trials because distracters are not used to calculate those effects.

correct. Using the response latency approach for the GNAT may be much more reliable simply because each data point is a continuous value (milliseconds). Greater experience with the GNAT will help to identify factors that will maximize its internal reliability and thus maximize its utility for psychological research.

Relationships Between Various Implicit Measures. An issue raised by Bosson et al. (2000) concerns the lack of relationship among various measures of implicit social cognition (e.g., priming, IAT). Bosson et al. point out that the various implicit measures may assess different aspects of a complicated network of associations and therefore show little overlap. In addition, implicit measures tend to show low internal reliability, which can attenuate the actual relationships between implicit measures and imply lower commonality than the measures actually share (Cunningham et al., 2001). In Experiment 6, the GNAT and the IAT were found to be related but that relationship was not strong. While this is, in part, due to the low reliability of the individual measures, there may also be other more psychologically interesting reasons for the lack of correspondence.[12] The assumption that these measures tap the same cognition, based on some similarities between the techniques, may be erroneous. The IAT and GNAT may elicit substantively different aspects of implicit social cognition because of specific aspects of their designs. For example, the IAT is a clearly relative measure and may tap evaluations in terms of their relative standing. The GNAT on the other hand, deemphasizes relative comparison and, instead, frames evaluation of a target concept in a context of other concepts. This difference can affect the representation that is in play with each technique. A deeper understanding of the differences between the evaluations elicited by these, and other, techniques will follow from process analyses of the various techniques (see DeHouwer, in press) as well as comparisons of differential effects on various techniques from identical experimental manipulations. Because of the uniqueness of each implicit measure, experimental reports that replicate implicit effects across techniques provide extra confidence that the effects are not due to a particular procedural aspect of any single tool.

12. We informally tested the relationship between the IAT and the GNAT after correcting for attenuation with a structural equation model. The GNAT and IAT were each parceled to create two subscores. A difference score between the black GNAT subscores and the white GNAT subscores were taken to mirror the relative quality of the IAT. The structural model tested the correlation between latent GNAT and IAT scores (i.e., the correlation between measures correcting for attenuation due to unreliability). The model fit the data well ($\chi^2(3) = 2.02, p = .57$, RMSEA = .000) and the GNAT-IAT correlation was .55. More information on this analysis is available from the first author.

Using the GNAT and Other Measures of Implicit Social Cognition. The flexibility of the GNAT makes it amenable to a wide range of potential applications. Even so, the GNAT is only one of a variety of effective measures of implicit social cognition. No one measure will, or can, serve the demands of all research questions. The important lesson is that the question of which tool, even within a family of largely similar tools, is one that should be made self-consciously. Research in social cognition has reached a stage where it is possible to do so. In selecting a measure of implicit attitudes, researchers should choose a task that best maps on to the theoretical questions of interest. For example, an IAT may be the most effective measure for assessing preferences for natural dichotomies (e.g., male/female, fat/thin), or concepts that we tend to think about in terms of relative comparison (e.g., Coke/Pepsi). The GNAT offers flexibility in the contextual characteristics of the evaluative situation as well as the relatively more independent assessment of an attitude object. For those attitude objects, like smoking, that have no clear comparison, the GNAT may be the most appropriate tool. When using the GNAT, it is important to select a context appropriate to the attitude object given the theoretical question. In addition, the ease of categorization of stimuli will affect the choice of response deadline to avoid ceiling (perfect responding) and floor effects (random responding). Future investigations with the GNAT may profit from using idiosyncratically defined windows calibrated to an individual subject's ability to discriminate signal from noise. In addition, caution should be exhibited in selecting stimulus items such that they belong to one, and only one, category and to avoid systematic variations in evaluative intensity (strongly positive or negative).

With effective measures that accompany advances in theory, psychological research will be in a better position to predict how and when preferences and beliefs that run unchecked and outside conscious control can affect judgments and behavior. In the coming years, the refinement of measurement techniques will enhance the theoretical development of such constructs and improve understanding of the structure and function of implicit social cognition. There is no doubt that the GNAT itself will be swatted with evidence that shows its own constraints. It is offered here as but one procedure with features that may address the limits of some existing techniques.

APPENDIX: TEXT STIMULI USED IN EXPERIMENTS 1–6

Bugs (Experiments 1–5). aphid, ants, bees, beetle, bugs, caterpillar, centipede, cockroach, cricket, dragonfly, flea, gnat, grasshopper, hornet, insect, maggot, mosquito, moth, roach, spider, stinkbug, termite, tick, wasp.

Fruit (Experiments 1–5). apple, apricot, banana, berry, blueberry, cantaloupe, cherry, fruit, grape, grapefruit, lemon, lime, mango, melon, nectarine, orange, peach, pear, pineapple, plum, raspberry, strawberry, tangerine, watermelon.

Good (Experiments 1–6). beautiful, celebrating, cheerful, excellent, excitement, fabulous, friendly, glad, glee, good, happy, joyful, laughing, likable, loving, marvelous, pleasure, smiling, splendid, superb, paradise, terrific, triumph, wonderful.

Bad (Experiments 1–6). angry, bad, brutal, destroy, dirty, disaster, disgusting, dislike, evil, gross, hate, horrible, humiliate, nasty, noxious, painful, revolting, sickening, terrible, tragic, ugly, unpleasant, yucky.

Distracter set for fruit and bugs GNATs (Experiment 2a). antelope, bookshelf, coffee, copper, dog, flannel, gem, horse, meter, monkey, month, pasta, periodical, pizza, potato, pudding, rabbit, raccoon, rug, spatula, square, steak, table, tulips.

Distracter set for fruit GNAT (Experiments 2b and 3–super-ordinate condition). almonds, beef, broccoli, butter, cheesecake, chicken, gravy, herbs, lasagna, mozzarella, oregano, paprika, parsley, peanuts, peas, pizza, potato, pudding, rice, salt, spaghetti, steak, turnip, yams.

Distracter set for bugs GNAT (Experiments 2b and 3–super-ordinate condition). alligator, antelope, dog, eagle, fish, giraffe, groundhog, horse, lizard, monkey, mouse, octopus, ostrich, penguin, pig, salamander, shark, snake, sparrow, rabbit, raccoon, trout, whale, wolf.

Positively valenced distracter set for fruit and bugs GNATs (Experiment 4). cinnamon, deer, diamond, dolphin, doughnut, eagle, gold, horse, house, kitchen, mountain, neighbor, rabbit, rose, sailboat, silk, summer, violin.

Negatively valenced distracter set for fruit and bugs GNATs (Experiment 4). accordion, anchovy, attic, burlap, coal, lead, mule, platypus, rat, rowboat, shack, snake, stranger, vinegar, volcano, vulture, weed, winter.

Faces used in Experiment 6 are available, upon request, from the first author.

REFERENCES

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder (pp. 117-150).* Washington DC: American Psychological Association.

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68,* 181-198.

Banaji, M. R., & Hardin, C. (1996). Automatic Stereotyping. *Psychological Science, 7,* 136-141.

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology, 80,* 1-46.

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. *Behavior Research Methods, Instruments, and Computers, 18,* 299-303.

Blair, I.V. & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70,* 1142-1163.

Blair, I. V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of automatic stereotypes through mental imagery. *Journal of Personality and Social Psychology, 81*(5) 828-841.

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79(4),* 631-643.

Bradburn, N. M. (1983). Response effects. In P. H. Rossi, J. D. Wright, & A. B. Anderson (Eds.), *Handbook of survey research* (pp. 289-328). New York: Academic Press.

Brewer, M. B. (2001). Ingroup identification and intergroup conflict: When does ingroup love become outgroup hate? In R. Ashmore, L. Jussim, & D. Wilder (Eds.), *Social Identity, Intergroup conflict, and conflict reduction* (pp. 17-41). New York: Oxford University Press.

Brewer, M. B., & Brown, R. J. (1998). Intergroup relations. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 554-594). New York: Oxford University Press.

Carpenter, S., & Banaji, M. R. (2000). *Implicit gender attitudes: Group membership, cultural construal, and malleability.* Unpublished manuscript, Yale University, New Haven.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

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., 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*(3), 316-328.

DeHouwer, J. (in press). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klaner (Eds.) *The Psychology of Evaluation: Affective Processes in Cognition and Emotion.* Mahwah, NJ: Lawrence Erlbaum.

DeHouwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 12,* 45-61.

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

Draine, S. C, & Greenwald, A. G. (1998). Replicable unconscious semantic priming. *Journal of Experimental Psychology: General, 127,* 286-303.

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answer from research on attitudes, gender stereotypes, and judgments of competence. In

W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 5, pp. 1-35). Chichester, England: John Wiley.

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.

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

Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology, 57,* 239-249.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

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

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

Hinkle, S., & Brown, R. (1990). Intergroup comparisons and social identity: Some links and lacunae. In D. Abrams and M. Hogg (Eds.), *Social identity theory: Constructive and critical advances* (pp. 48-70). London: Harvester Wheatsheaf.

Judd, C. M., & McClelland, G. H. (1998). Measurement. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 180 - 232). New York: Oxford University Press.

Lemm, K., & Banaji, M.R. (1999). Unconscious attitudes and beliefs about women and men. In U. Pasero & F. Braun (Eds.), *Wahrnehmung und Herstellung von Geschlecht [Perceiving and performing gender]* (pg. 215-233). Opladen: Westdutscher Verlag.

Levy, B. (1996). Improving memory in old age through implicit self-stereotyping. *Journal of Personality and Social Psychology, 71,* 1092-1107.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*(5), 842-855.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163-203.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2001). *Contextual variations in implicit evaluation.* Manuscript submitted for publication.

Mummendey, A., Otten, S., Berger, U., & Kessler, T. (2000). Positive-negative asymmetry in social discrimination: Valence of evaluation and salience of categorization. *Personality and Social Psychology Bulletin, 26,* 1258-1261.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264-336). Hillsdale, NJ: Erlbaum.

Nosek, B. A., Banaji, M. R., Greenwald, A. G. (in press). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics.*

Ostrom, T. M. (1989). Interdependence of attitude theory and measurement. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald, *Attitude structure and function* (pp. 11-36). Hillsdale, NJ: Erlbaum.

Otten, S., & Mummendey, A. (1996). Intergroup discrimination in positive and negative outcome allocations: Impact of stimulus valence, relative group status, and relative group size. *Personality and Social Psychology Bulletin, 22,* 568-571.

Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology, 26,* 199-216.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin, 26, 1315-1328.*

Schwarz, N., Groves, R. M., & Schuman, H. (1998). Survey methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 143 - 179). New York: Oxford University Press.

Sumner, W. G. (1906). *Folkways*. New York: Ginn.

Tajfel, H. (Ed.) (1978). *Differentiation between social groups: Studies in the social psychology of intergroup relations*. London: Academic Press.

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge, England: Cambridge University Press.