10

Implicit Self-esteem

SHELLY D. FARNHAM, ANTHONY G. GREENWALD, AND MAHZARIN R. BANAJI

Implicit Self-esteem and Social Identity

People tend to be biased in favor of their ingroup even when ingroups are minimally defined (Tajfel and Turner, 1979). Tajfel and Turner (1986) argued that such an ingroup bias arises out of people's motivation to achieve a satisfactory image of the self through a positive social identity, leading to behaviors that enhance the ingroup and derogate the outgroup. As such, how one evaluates the self, and how one evaluates important social identities, should be related. In self-report measures positivity of the self-concept, or personal self-esteem, is correlated with evaluations of social identity, or collective self-esteem (Crocker, Luhtanen, Blaine, and Broadnax, 1994; Luhtanen and Crocker, 1992). However, the level of stigmatization of one's social identity has no effect on personal self-esteem (Crocker and Major, 1989), and the relationship between self-esteem and ingroup favoritism is unclear (Abrams and Hogg, 1988). A source of such ambiguity may be the unreliability of self-report measures of self-esteem. While self-report measures of personal self-esteem seek to assess affective self-regard, they also manage to capture constructs such as impression management and self-deception. We propose that an *indirect* measure of self-esteem, similar to indirect measures developed in attitude research, has the potential of shedding light on the relationship between personal self-esteem and ingroup favoritism.

A potential indirect measure is provided by the Implicit Association Test (IAT; Greenwald, McGhee, and Schwartz, 1998), which assesses automatic (and not necessarily consciously reportable) concept–attribute associations. The IAT has already been used to measure automatic associations of ethnic groups with evaluation (implicit prejudice, or implicit ingroup favoritism), and of gender

This research was supported by grants from National Science Foundation, SBR-9422242, SBR-9710172, SBR-9422241, and SBR-9709924, and from National Institute of Mental Health, MH-41328 and MH-001533.

Farnham, S.D., Greenwald, A.G., & Banaji, M.R. (1999). Implicit self-esteem. In D. Abrams & M.A. Hogg (Eds.), Social identity and social cognition (pp. 230-248). Oxford, UK: Blackwell.

with traits (implicit stereotypes; Rudman, Greenwald, and McGhee, 1998). This chapter (1) reviews the construct validity of self-report measures of self-esteem; (2) describes how the IAT can provide an indirect measure of self-esteem; and (3) discusses how the IAT may be used to further understanding between personal self-esteem and evaluation of social identity.

Self-esteem and questions of construct validity

William James (1890) defined self-esteem as a self-feeling that is determined by a comparison between the actual self and the ideal self. Following James's definition of self-esteem, standard self-report measures of self-esteem ask respondents either to rate themselves on a variety of specific traits (Marsh, 1986; Pelham and Swann, 1989; Wells and Marwell, 1976), or to indicate how they feel about themselves globally (Rosenberg, 1979). However, research has not supported James's formulation because self-esteem does not appear to be the product of honest appraisal of one's traits and abilities (Rosenberg, 1979) or one's social identity (Crocker and Major, 1989). Rather, research indicates that the higher one's self-esteem, the greater the self-enhancing bias (see Brown, 1991, for review). Consequently, psychologists have debated extensively whether self-esteem causes self-appraisals or vice versa (Brown, 1993; Pelham and Swann, 1989), whether self-esteem leads to discriminatory behavior or vice versa (Abrams and Hogg, 1988), whether people are motivated towards accuracy or positivity in their self-concepts (Brown, 1991; Shrauger, 1975; Swann, 1990), and why, if having high self-esteem is not based on accurate self-appraisals, anyone would have low self-esteem (Baumeister, 1993).

What psychologists have only recently considered is that the correspondence between self-esteem measures and self-enhancing behaviors suggests that self-esteem measures may be capturing the wrong construct (Baumeister, Tice, and Hutton, 1989): the motive to present a positive attitude toward self rather than genuine self-esteem.

A *positivity bias* provides no threat to the construct validity of self-esteem measures (i.e., their ability to measure the self-esteem construct). Whether such biases arise from positive feelings toward the self (Brown, 1993) or cognitive beliefs about the self (Markus and Wurf, 1986), they are a reflection of the level of positive self-regard. Such an automatic positivity bias can be interpreted as a manifestation of *implicit self-esteem*. Greenwald and Banaji defined implicit self-esteem as "the introspectively unidentified (or inaccurately identified) effect of the self-attitude on evaluation of self-associated and self-dissociated objects" (1995, p. 11). This tendency to overestimate one's traits and abilities is understood as a spillover of positive affect from the self to objects associated with the self. Because most people have positive self-affect (Banaji and Prentice, 1994; Greenwald, 1980; Taylor and Brown, 1988), implicit self-esteem effects usually

involve a positivity bias in processing information about the self (see Greenwald and Banaji, 1995, for review). In the realm of social identity, an individual's tendency to exalt any group by virtue of its association with self is an implicit self-esteem effect (Greenwald and Banaji, 1995).

Whereas a positivity bias provides no threat to the construct validity of self-report measures of self-esteem, *self-enhancing self-presentation strategies* provide a great threat to construct validity (Paulhus, 1986). As a consequence of self-presentation strategies, explicit measures may assess strategies of associating self with positive traits and dissociating self from negative traits. These self-presentation strategies are not necessarily to be identified with the construct of self-esteem (i.e., affective self-regard).

Paulhus (1986) defined the self-presentation strategies of impression management and self-deception as follows:

> I will use *impression management* to refer to conscious dissimilation of test responses designed to create a favorable impression in some audience. In contrast, the term *self-deception* will refer to any positively biased response that the respondent actually believes to be true. (p. 144)

Impression management and self-deception can be conceived as two ends of a continuum of self-presentation, ranging from self-presentation to others to self-presentation to self (Greenwald and Breckler, 1985). Impression management is directed toward an outward audience (Goffman, 1959; Schlenker, 1980) and self-deception is directed inwardly.

Paulhus (1986) argues that while researchers should control for impression management in self-report self-esteem measures, self-deception should be allowed to emerge. Self-deception indicates high self-esteem, and thus psychologists *want* self-esteem scales to capture self-deception. However, this assertion is debatable. According to Sackeim and Gur (1978), self-deception could involve holding positive explicit beliefs and negative implicit beliefs simultaneously. Explicit measures may therefore not distinguish self-deception (explicit positivity with implicit negativity) from genuine high self-esteem (positivity at both explicit and implicit levels).

Self-report self-esteem measures have questionable construct validity

We propose that explicit (self-report) measures of self-esteem capture self-presentation in addition to affective self-regard. The support for this assertion follows.

The discriminant validity of self-reported self-esteem. In order for a measure to have high construct validity, it should discriminate its target construct from other constructs. Therefore, it is disconcerting that self-report measures of self-esteem correlate highly with measures of self-presentation style, suggesting a discriminant validity problem (Wells and Marwell, 1976). A correlation between a tendency toward self-presentation, as a personality trait (Crowne and Marlowe, 1964), and self-esteem indicates that self-esteem measures are biased by self-presentation. Both self-deception and impression management measures correlate with self-esteem measures (Lindeman and Verkasalo, 1995), with self-deception having a higher correlation (around .6) than impression management (around .3; Raskin, Novacek, and Hogan, 1991). That people with high self-deception and impression management scores also have high self-reported self-esteem suggests they are denying or defending against threatening negative information in the items of the self-esteem questionnaires (Cohen, 1959; Coopersmith, 1959; Schneider and Turkat, 1975). Self-report measures of self-esteem do not appear to discriminate well between self-presentation and self-esteem.

The convergent validity of self-reported self-esteem. In order for a measure to have high construct validity, it should correlate with theoretically related constructs. Although self-report measures of self-esteem tend to correlate highly with each other and other related self-reported constructs such as anxiety and depression (Blascovich and Tomaka, 1991; Fleming and Courtney, 1984; Wells and Marwell, 1976), they do not correlate as well with peer or observer reports of self-esteem (Demo, 1985). To some extent the low correlations between self-report measures and peer measures may be due to differences in kind of measures. However, using confirmatory factor analyses, Demo (1985) found that self-reported self-esteem and observer ratings of self-esteem are best considered two distinct, moderately correlated factors. Whereas Demo assumes the self-reported self-esteem more accurately represents genuine, experienced self-esteem, one might as easily argue that the observer ratings are more accurate representations, given people's tendencies toward self-presentation.

The predictive validity of self-reported self-esteem. In order for a measure to have high construct validity, it should predict the behaviors that are theoretically related to the construct in question. However, self-esteem measures are low in predictive validity for the following reasons:

1 *Self-esteem measures inconsistently predict sensitivity to feedback.* One of James's (1890) assumptions is that a person with genuine high self-esteem should be able to receive negative feedback without finding it too painful. However, some people with self-reported high self-esteem are highly sensitive to negative feedback (Baumeister, Heatherton, and Tice, 1993). In particular, people classified as having "defensive self-esteem" (high self-reported self-esteem and high need for approval) appear to find negative feedback painful: they do not like others who give them negative feedback (Hewitt and Goldman, 1974); they increase in their need for approval following failure (Schneider and Turkat, 1975); they cheat to do well on a task (Lobel and

Levanon, 1988); and they lower their levels of aspiration following failure (Lobel and Teiber, 1994).

2 *Self-esteem scores inconsistently predict quality of relationship with parents.* Contrary to predictions of developmental psychologists, some people who have high self-esteem scores have histories of negative interactions with parents. Developmental psychologists generally assume that self-esteem is acquired through parent–child relationships, with positive self-regard being a reflection, or internal model, of the parents' regard for the child (Bretherton, 1985; Cassidy, 1988). However, on occasion a child is placed in an extremely distressing situation when the parent has negative regard for the child, and the child responds to the situation by distancing from the parent (Bretherton, 1985). Both Mikulincer (1995) and Cassidy (1988) found that persons who showed such distant, negative relations with their parents had idealized, perfectly positive self-images. Such an effect, Mikulincer (1995) argues, "may imply that their self-esteem is so low and fragile that they cannot tolerate discovery of the slightest flaw. This idealization of the self seems to be a defense against the experience of rejection by others on the recognition of one's imperfections" (p. 1213).

3 *Self-esteem measures predict behaviors that are more theoretically related to self-presentation strategies than self-esteem.* Self-reported high self-esteem scores predict a wide variety of self-enhancing behaviors (Brown, 1991) that involve self-deceptive or impression management strategies. For example, when faced with negative feedback, the high self-esteem scorer turns attention to other positive traits (Baumeister, 1982; Baumeister and Jones, 1978), inflates the self by deflating others (Brown, Collins, and Schmidt, 1988), and exaggerates estimates of how many others share a negative trait (Campbell, 1986). Baumeister et al. (1989; Tice, 1991) suggest that self-esteem scales measure differences in self-presentational styles: either self-enhancing, or self-protecting. That self-esteem measures predict the tendency to use self-presentation strategies suggests they may measure the construct "favorable self-presentation" rather than "positive self-regard."

4 *Self-esteem measurements predict self-enhancing behaviors mainly in situations where self-presentational demands are high.* The self-enhancing behaviors of those with self-reported high self-esteem become augmented in public situations, and the self-protective behaviors of those with low self-esteem increase in public situations (see Baumeister et al., 1989, for review). That self-esteem scores have a greater probability of predicting self-enhancing behaviors when they occur in public again suggests that self-esteem scores measure self-presentational tendencies.

5 *Self-esteem measurements do not predict behaviors considered implicit self-esteem effects.* Many of the behaviors considered implicit self-esteem effects, resulting from a positivity bias, are not predicted by self-esteem measures. For example, the degree to which an ingroup bias occurs in a minimal group paradigm (where self becomes associated with a group formed at random) is unrelated to level of self-reported self-esteem (Crocker and Schwartz, 1985; Crocker et al., 1987).

6 *Self-esteem measures' ability to predict mental health may be due to its association with self-enhancing behaviors.* Self-esteem measures do a fair job of predicting mental health (Kaplan, 1975; Rosenberg, 1965). However, self-reported self-esteem's relation to self-enhancing behaviors has led a few to theorize that self-esteem leads to mental health because it plays a buffering role against the stressors of life (Greenberg, Solomon, Pyszczynski, Rosenblatt, Burling, Lyon, Simon, and Pinel, 1992; Taylor and Brown, 1988). According to this view, people with high self-esteem have a proclivity towards self-deception and self-enhancement, and are thus able to respond to stressful situations with a minimum of anxiety. In other words, self-deception and self-enhancement buffer the self against anxiety, rather than level of self-esteem.

In conclusion, self-report measures of self-esteem have questionable discriminant, convergent, and predictive validity. In particular, self-report measures of self-esteem correlate with measures of self-presentation, and predict self-presentational behaviors, suggesting that these measures capture a construct of self-presentation more than affective self-regard. In order to measure genuine self-esteem, self-presentation must be avoided altogether through indirect measures of self-esteem. Another area of measurement in social cognition that has been beleaguered by self-presentation biases is that of socially sensitive attitudes related to prejudice and discrimination. Recent developments in measures that indirectly assess attitudes (Dovidio and Fazio, 1992; Greenwald et al., 1998) provide the necessary methodology allowing the indirect measure of self-regard.

A Different Approach – Indirect Measure of Self-esteem

Recent developments in the indirect measurement of attitudes borrow extensively from a neural network model of the brain developed in cognitive psychology (see Schneider and Shiffrin, 1977; Shiffrin and Schneider, 1977, for review), where information is conceived as stored at sites in a vast tangle of neural links (Collins and Loftus, 1975) that are organized hierarchically according to semantic relationships. In essence, the relationship between any two concepts can be measured by determining how far one must travel to get from one to the other across such links.

Indirect measures of attitude use two cognitive phenomena to its advantage: the automatic activation of attitudes effect, and spreading activation. Research has shown that the evaluative as well as the semantic content of words are processed automatically upon sight (Greenwald, Klinger, and Liu, 1989; Murphy and Zajonc, 1993). In other words, affective reactions such as liking, disliking, preference, and evaluations are processed instantaneously, or *automatically*

activated (Fazio, Sanbonmatsu, Powell, and Kardes, 1986; Zajonc, 1980). Research has also shown that any piece of information, once activated, makes it easier to process subsequent, similar information because of the *spread of activation* that crosses the short distance between two neighboring links (Collins and Loftus, 1975; Neely, 1977).

Using these two cognitive phenomena, Fazio et al. (1986) argued that the strength of an attitude can be measured by the ease with which a person judges the valence of one concept after being presented with another concept. For example, if a person found it very easy to judge the word "sunshine" as pleasant immediately after seeing the word "democrat," then he or she has a positive attitude towards democrats. What has occurred is that the affective information in the word "democrat" has been automatically activated, making it easier to recognize the affective information in "sunshine" through spreading activation. Fazio, Jackson, Dunton, and Williams (1995) found that images of black faces facilitated categorization of negative words for white subjects, and that images of white faces facilitated categorization of negative words for black subjects.

Whether the automatic activation of attitudes effect can measure the self-concept and self-esteem depends on whether the self is an attitude object that is automatically processed (see Greenwald and Pratkanis, 1984, for discussion of the self as an attitude object). Psychologists have argued that the self-concept is a schema, a rich and highly organized cluster of ideas surrounding the central concept of self (see Kihlstrom and Cantor, 1984; or Markus and Wurf, 1986, for review). Like other schemas, the self-schema affects information processing. For example, self-consistent information is more efficiently processed than inconsistent information, and self-relevant information is easily recalled and recognized (see Markus and Wurf, 1986, for review). Most importantly for our discussion, research shows that the content of the self-concept influences how quickly self-relevant information is processed. For example, Markus (1977) found that in categorizing adjectives as "me" or "not me," people were able to make faster judgments for words for which they had well-developed self-schemas.

If the self-schema is a well-integrated whole, it ought to have an affective, *attitudinal* component that influence processing of self-relevant information (Fiske and Pavelchak, 1986; Greenwald and Pratkanis, 1984). Research suggests that affective information about the self is automatically processed just as is the affective information of any attitude object. A number of studies have found that people are quicker to judge words as "me" or "not me" if they are positive or negative than if they are neutral (Markus, 1977; Mueller and Grove, 1991; Ross, Jurek, and Oliver, 1996). More importantly, automatic processing of affective information may be used to examine individual differences in self-concept. For example, Bargh and Tota (1988) showed that individual differences may be measured by assessing how much increasing cognitive load affects reaction times. They found that nondepressed subjects had a harder time categorizing negative self-related concepts under conditions of increased cognitive load than did depressed subjects.

In sum, individual differences in level of self-esteem may be assessed using the automatic activation of attitudes effect, because the self-concept is a well-integrated schema with an affective, attitudinal component that influences how self-relevant information is processed. In particular, individual differences in levels of self-esteem may be assessed using a procedure in which individuals perform a task that they can complete more efficiently using automatic processing if they have high implicit self-esteem. The Implicit Association Test, adapted for the self-concept, is just such a procedure.

The Implicit Association Test (IAT)

Greenwald et al. (1998) developed the Implicit Association Test (IAT) to measure automatic concept–attribute associations. An assumption of the test is that strongly associated (compatible) attribute–concept pairs should be easier to classify together than are weakly associated or opposed (incompatible) attribute–concept pairs. Ease of classifying is measured by the response times and errors in performing such categorizations.

To get an idea of the IAT procedure, imagine sorting a deck of cards. Your first task is to judge if a card is a spade or a heart. You put the spades in a pile on the left, and the hearts in a pile on the right. Next, diamonds and clubs are added to the deck, and you are asked to put both spades and clubs on the left, and diamonds and hearts on the right. If spades and clubs can be cognitively grouped according to some shared feature, the task will be relatively simple. For example, by keeping the simple instruction "only black to left" in mind, the cards can easily be sorted into the left and right piles.

What happens if color cannot be used as a grouping cue? If one's task is to place clubs and hearts on the left and spades and diamonds on the right, the simple black–left, red–right strategy does not work, necessitating a more cognitively demanding strategy.

Greenwald et al. (1998) presented subjects with a series of words on a computer screen and had them categorize each word as quickly as possible by pressing a left or right key on a keyboard (see figure 10.1). The automatic association between a concept (for example, *flower*) and the attribute of evaluative pleasantness is measured by the difference in speed between the condition in which *flower* and *pleasant* were mapped together and the condition in which *flower* and *unpleasant* were mapped together (see Compatible and Incompatible screens in figure 10.1).

Using the IAT to measure implicit ingroup bias

Applying the IAT to social attitudes, Greenwald et al. (1998) assessed Korean and Japanese subjects' attitudes towards one another. Both groups found it easier

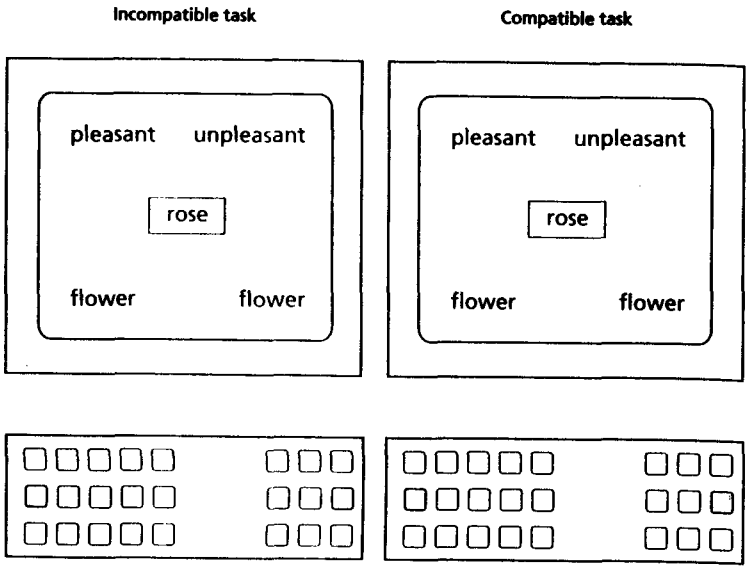


Figure 10.1 The display screen for the computer Implicit Association Test (IAT). Individuals are presented with a word in the middle of the screen ("rose"), and must categorize it into one of the categories displayed to the left and right of the word by pressing the left or right key. The categorization task is compatible when the categories are easily associated (e.g., "pleasant" and "flower"), and incompatible when the categories are associated with difficulty (e.g., "pleasant" and "insect").

to associate pleasant words with names from their own group than pleasant words with names from the other group. The IAT had a larger effect size for the difference in attitudes towards Japanese and Koreans than did the explicit measures of attitudes towards these groups. In addition, Greenwald et al. (1997) found that white subjects more easily associated white names with pleasant words and black names with unpleasant words than vice versa. Again, the effect size of the difference in attitude towards blacks and whites proved to be greater for the IAT than for the explicit measures of attitudes. Importantly, the Japanese and Korean subjects were much more willing to explicitly show ingroup bias in their attitudes towards one another than white subjects were willing to show negative attitudes towards blacks. As a consequence, the correlations between the implicit and explicit measures were higher for the Japanese/Korean experiment than those for the white/black experiment.

Using the IAT to measure implicit self-esteem

A measure of implicit self-esteem using the IAT examines the extent to which people are faster at categorizing self words and pleasant words together than categorizing self words and unpleasant words together. To apply the IAT to measuring self-esteem, we developed a computer program that allowed subjects to provide idiosyncratic information such as first and last names, home town, and telephone number. After providing such me-objects, subjects then chose from lists of similar, not-me objects. During the IAT subjects were presented with a series of words to categorize. We measured how long it took the subject to categorize each word from the moment it appeared on the screen to the moment the correct key was pressed.

For a demonstration of the steps of the self-esteem IAT adapted to pen and paper, see figure 10.2. First subjects practiced categorizing words as being either

Step 1		
unpleasant		pleasant
•	joy	•
•	vomit	•
•	agony	•
•	peace	•
•	death	•
•	sunrise	•
•	warmth	•
•	corpse	•
•	gold	•
•	slime	•

Step 2		
not-me		me
•	self	•
•	other	•
•	they	•
•	them	•
•	I	•
•	mine	•
•	it	•
•	me	•
•	their	•
•	myself	•

Step 3		
unpleasant or not-me		pleasant or me
•	self	•
•	joy	•
•	them	•
•	death	•
•	other	•
•	sunrise	•
•	my	•
•	warmth	•
•	me	•
•	corpse	•

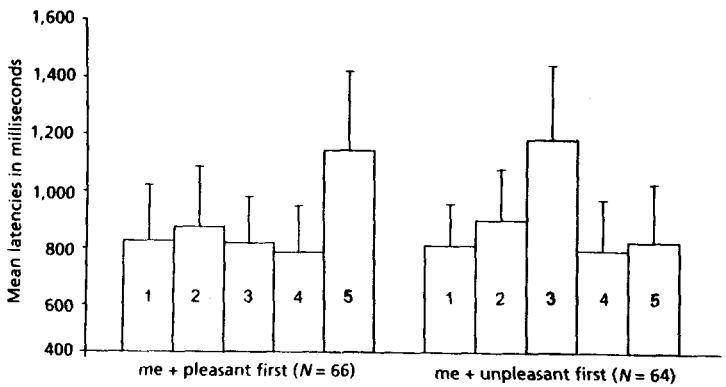
Step 4		
me		not-me
•	myself	•
•	their	•
•	me	•
•	self	•
•	them	•
•	I	•
•	they	•
•	other	•
•	mine	•
•	it	•

Step 5		
unpleasant or me		pleasant or not-me
•	they	•
•	vomit	•
•	I	•
•	agony	•
•	their	•
•	gold	•
•	it	•
•	peace	•
•	mine	•
•	slime	•

Figure 10.2 Demonstration of the self-esteem IAT. Read each word in the list, and with the back of a pen tap the black circle under the category to which the word belongs. Complete each list of words as quickly as possible, without skipping any words. You will probably find Step 5 more challenging than Step 3.

pleasant or unpleasant (figure 10.2, Step 1). Second, subjects practiced categorizing words as being me or not-me. Third, half the subjects categorized words as either me/pleasant or not-me/unpleasant, while the other half categorized words as either me/unpleasant or not-me/pleasant. Fourth, the me and not-me subcategories were switched and practiced. Finally, if initially me and pleasant were together, me and unpleasant were together, and vice versa. Implicit self-esteem was calculated by measuring the difference in reaction times, or latencies, between the two conditions where me words were categorized with pleasant words, and me words were categorized with unpleasant words (i.e., the difference in mean latency between Steps 3 and 5).

Our results indicated that people made faster judgments when me words were categorized with pleasant words than when me words were categorized with unpleasant words (see figure 10.3, which shows the latencies for each step of the IAT). The pattern of data indicates that combining the me/not-me and pleasant/unpleasant discriminations was about as easy for subjects to do as was performing either of these discriminations alone, *when* the me and pleasant items shared the same response. By contrast, the combination of two discriminations added an



1 unpleasantm *or* pleasant	1 unpleasant *or* pleasant
2 not-me *or* me	2 me *or* not-me
3 unpleasant/not-me *or* pleasant/me	**3 unpleasant/me *or* pleasant/not-me**
4 me *or* not-me	4 not-me *or* me
5 Impleasant/me *or* pleasant/not-me	**5 unpleasant/not-me or pleasant/me3**

Figure 10.3 The five steps of the Implicit Association Test, counterbalanced for order.

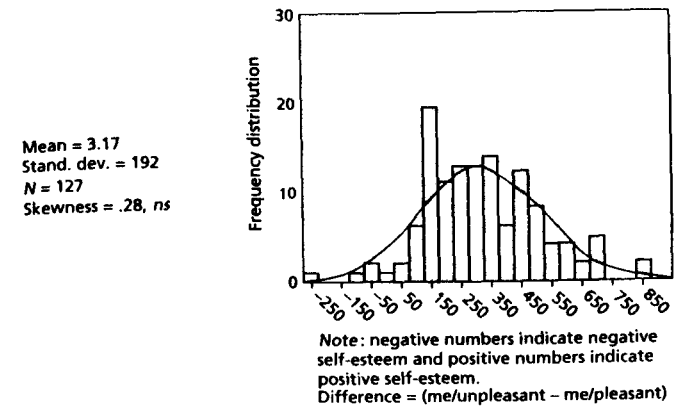


Figure 10.4 IAT effect, with raw latency score, is almost normally distributed.

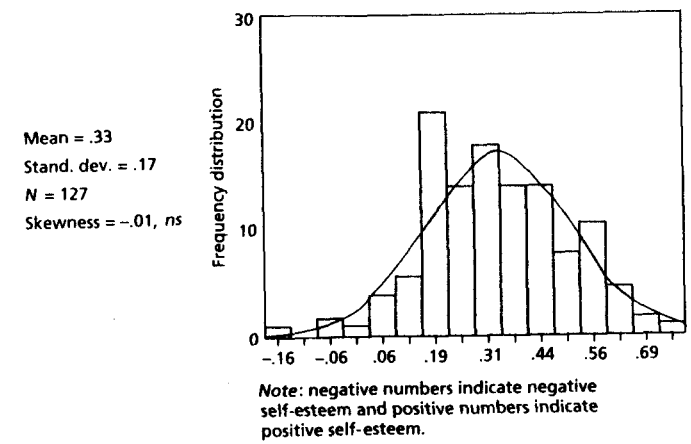


Figure 10.5 IAT effect, with log transformation, is normally distributed.

average of about 327 ms to mean latencies when, instead, the me and unpleasant items shared the same response. This mean difference in difficulty between the two combinations, called the IAT effect, corresponded to almost two times the standard deviation of the me + pleasant condition (i.e., the effect-size measure of Cohen's $d = 1.86$, and $F(1, 119) = 439$, $p \cong 10^{41}$). Analyses were performed on log transformations of the latencies, because raw scores tend to be positively skewed (although not as skewed as those of explicit measures). See figures 10.4, 10.5 and 10.6.

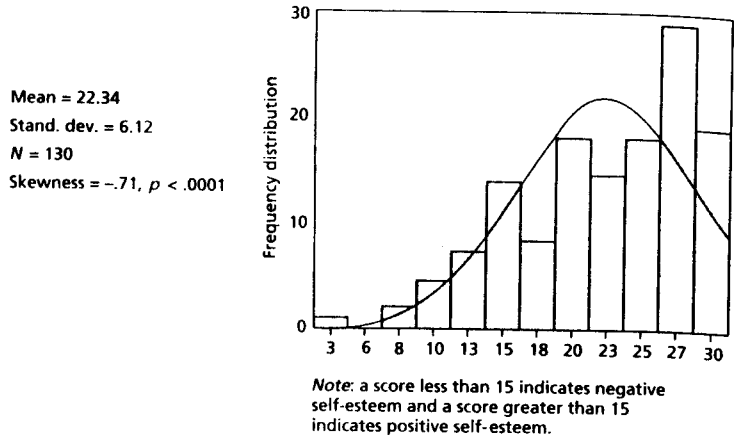


Figure 10.6 Rosenberg SES, raw scores are not normally distributed.

In addition to completing the IAT, subjects responded to seven explicit measures (see table 10.1). The IAT measure of self-positivity had only weak positive correlations with these other measures (see first row of table 10.1). These low correlations indicate that the IAT does not measure the same construct that is represented by shared variance among the set of explicit measures. Although it is not the only explanation of such low correlations, this conclusion is consistent with the supposition that the explicit measures are more sensitive to self-presentation strategies than to self-esteem.

The IAT's sensitivity to the self-positivity bias, or *implicit self-esteem*, described by Greenwald and Banaji (1995) is encouraging. Implicit self-esteem is conceived as a transfer of affect from one's self-attitude to the concepts or objects associated with the self. If, as we suspect, self-report measures of self-esteem are highly sensitive to self-presentation, then assessment of the construct validity of the IAT's self-esteem measure cannot rely on correlations with the self-report measures. Consequently, our program of research is pursuing alternative construct validation strategies.

Using the IAT to measure the relationship between implicit ingroup bias, implicit self-esteem, and social identity

If the IAT is really measuring implicit self-esteem, then it ought to correlate positively with ingroup bias. In addition, the level of this correlation should depend on the level of identification with that group. To test this prediction, we had female subjects complete three IATs measuring implicit self-esteem, implicit

Table 10.1 Intercorrelations between the IAT, measures of self-esteem, self-deception, and impression management

	IAT	*Rosenberg SES*	*Thermometer scale*	*Self-affect scale*	*Semantic differential*	*Trait agreement scale*	*Self-attributes*	*Self-deception*	*Impression management*
IAT	1	.05	.04	.26	.26	.23	.24	.18	.17
Rosenberg SES		1	.44	.84*	−.07	.74*	.41*	.48*	.07
Thermometer scale, self–other[a]			1	.41	.18	.22	.13	.22	−.23
Self-affect scale[a]				1	−.06	79*	.38	.52*	.08
Semantic differential, self–other[a]					1	−.06	−.06	.13	−.10
Trait agreement scale[a]						1	.52*	.52*	.09
Self-attributes questionnaire							1	.50*	.09
Self-deception (BIDR)								1	.37*
Impression management (BIDR)									1

[a]$N = 54$, otherwise $N = 125$; **bold** = $p < .05$, *italics* = $p < .005$, *** = $p < .0005$

The first six measures are self-esteem measures. IAT = log latency for task with me/pleasant minus log latency for task with me/unpleasant. The Rosenberg SES (Rosenberg, 1979), self-attributes questionnaire (Pelham and Swann, 1989), trait agreement scale (Brown, 1993), self-deception scale and impression management scale (BIDR; Paulhus, 1991) were developed in previous research. All other measures (the thermometer scale, semantic differential, and self-affect scale) were developed for the current research.

ingroup bias (favoring females over males), and implicit identification with being female (Farnham and Greenwald, 1997). We measured self-esteem combining pleasant/unpleasant words with me/not-me words, as described above. Positivity towards females was measured combining female/male words (such as female, girl, woman, male, boy, and man) with pleasant/unpleasant words. Identification with female was measured by combining me/not-me words with male/female words.

In general, the 62 female subjects strongly favored females over males (IAT effect = 303 ms, $d = 2.34$), had positive implicit self-esteem (IAT effect = 303 ms, $d = 2.06$), and identified with being female (IAT effect = 204 ms, $d = 1.34$). More importantly, we did find that implicit self-esteem was correlated with implicit ingroup bias ($r = .34$, $p < .05$), and that the relationship between self-esteem and bias towards females depended on level of identification with being female, $F(1,55) = 3.48$, $p = .03$, one-tailed (see figure 10.7). Thus ingroup favoritism at least to some extent is an implicit self-esteem effect, one that depends on the person's level of association between the group and the self. A promising study

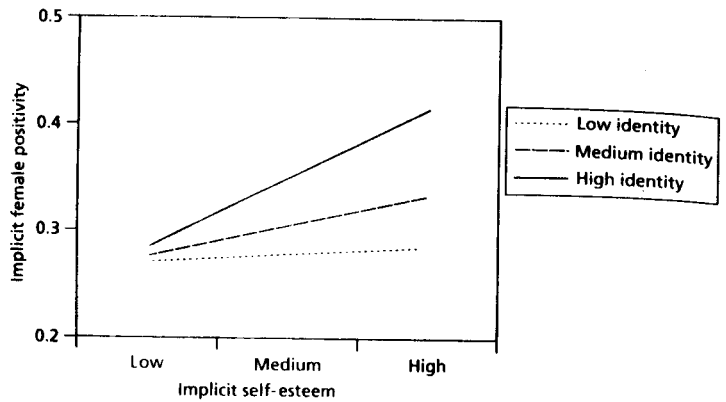


Figure 10.7 Relationship between implicit self-esteem and gender positivity depends on gender identity.

for the future is to examine whether the IAT may be able to predict people's behaviors in the minimal group paradigm.

Conclusions

The construct validity of self-reported self-esteem is questionable, because both (1) responses to these measures can be deliberately managed (i.e., they are sensitive to self-presentation and impression management strategies) and (2) the target construct of affective self-regard may be unavailable to introspection. The Implicit Association Test (IAT) provides a potential avenue around these threats to validity. As expected, the IAT reveals (1) a strong positivity bias in favor of the ingroup, or *implicit ingroup favoritism*, and (2) a strong positivity bias in favor of the self, or *implicit self-esteem*. Low correlations between the self-esteem IAT and explicit measures of self-esteem demand the conclusion that the IAT and the explicit measures assess different constructs that are no more than moderately related.

We also found that identification with group moderated the relationship between personal self-esteem and attitudes toward the ingroup. That this effect was found using measures of implicit self-esteem, implicit identification with the group, and implicit evaluation of the group, suggests that Tajfel and Turner's (1979) minimal group effect does occur on the automatic, implicit level. A person's self-affect either transfers to the ingroup as soon as the person has made the association between the self and the group, or the person dissociates the self and

group if the group is perceived to be negative. Ironically, our research examining the relationship between self-esteem and social identity suggests that the field of social identity needs to take another look at balance theory explanations of ingroup favoritism (Heider, 1958), rather than focusing entirely on self-esteem theories of ingroup favoritism. Heider's balance theory, much like Festinger's cognitive dissonance theory (1957), argues that people seek balanced relationships among their attitudes and identities. Thus, the closer one identifies the self with one's social identity, the greater is the imperative to keep feelings for the self and group consistent. The IAT shows that at least on an automatic level, the relationship between the self and gender identity adheres to a balanced pattern.

References

Abrams, D. and Hogg, M.A. (1988). Comments on the motivational status of self-esteem in social identity and intergroup discrimination. *European Journal of Social Psychology*, 18, 317–34.

Banaji, M.R. and Prentice, D.A. (1994). The self in social contexts. *Annual Review of Psychology*, 45, 297–332.

Bargh, J.A. and Tota, M.E. (1988). Context-dependent automatic processing in depression: Accessibility of negative constructs with regard to self but not others. *Journal of Personality and Social Psychology*, 54, 925–39.

Baumeister, R.F. (1982). Self-esteem, self-presentation, and future interaction: A dilemma of reputation. *Journal of Personality*, 50, 29–45.

Baumeister, R.F. (ed.) (1993). *Self-esteem: The puzzle of low self-regard*. New York: Plenum Press.

Baumeister, R.F., Heatherton, T.F., and Tice, D.M. (1993). When ego threats lead to self-regulation failure: Negative consequences of high self-esteem. *Journal of Personality and Social Psychology*, 64, 141–56.

Baumeister, R.F. and Jones, E.E. (1978). When self-presentation is constrained by the target's knowledge: Consistency and compensation. *Journal of Personality and Social Psychology*, 36, 608–18.

Baumeister, R.F., Tice, D.M., and Hutton, D.G. (1989). Self-presentational motivations and personality differences in self-esteem. *Journal of Personality*, 57, 547–79.

Blascovich, J. and Tomaka, J. (1991). Measures of self-esteem. In J.P. Robinson, P.R. Shaver, and L.S. Wrightman (eds), *Measures of Personality and Social Psychological Attitudes*, Vol. 1. San Diego: Academic Press, pp. 115–60.

Bretherton, I. (1985). Attachment theory: Retrospect and prospect. *Monographs of the Society for Research in Child Development*, 50, 3–35.

Brown, J.D. (1991). Accuracy and bias in self-knowledge. In C.R. Snyder and D.F. Forsyth (eds), *Handbook of Social and Clinical Psychology: The health perspective*. New York: Pergamon Press, pp. 158–78.

Brown, J.D. (1993). Self-esteem and self-evaluation: Feeling is believing. In J. Suls (ed.), *Psychological Perspectives on the Self*, Vol. 4. Hillsdale, NJ: Erlbaum, pp. 22–58.

Brown, J.D., Collins, R.L., and Schmidt, G.W. (1988). Self-esteem and direct versus indirect forms of self-enhancement. *Journal of Personality and Social Psychology*, 55, 445–53.

Campbell, J.D. (1986). Similarity and uniqueness: The effects of attribution type, relevance, and individual differences in self-esteem and depression. *Journal of Personality and Social Psychology*, 50, 281–94.
Cassidy, J. (1988). Child–mother attachment and self in six-year-olds. *Child Development*, 59, 121–34.
Cohen, A. (1959). Some implications of self-esteem for social influence. In C. Hovland and I. Janis (eds), *Personality and Persuasibility*, Vol. 2. New Haven: Yale University Press, pp. 102–20.
Collins, A.M. and Loftus, E.F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82, 407–28.
Coopersmith, A. (1959). A method for determining types of self-esteem. *Journal of Abnormal and Social Psychology*, 59, 84–7.
Crocker, J., Luhtanen, R., Blaine, B., and Broadnax, S. (1994). Collective self-esteem and psychological well-being among white, black, and Asian college students. *Personality and Social Psychology Bulletin*, 20, 503–13.
Crocker, J. and Major, B. (1989). Social stigma and self-esteem. The self-protective properties of stigma. *Psychological Review*, 96, 608–30.
Crocker, J. and Schwartz, I. (1985). Prejudice and ingroup favoritism in a minimal intergroup situation: Effects of self-esteem. *Personality and Social Psychology Bulletin*, 11, 379–86.
Crocker, J., Thompson, L., McGraw, K., and Ingerman, C. (1987). Downward comparison, prejudice, and evaluation of others: Effects of self-esteem and threat. *Journal of Personality and Social Psychology*, 52, 907–16.
Crowne, D.P. and Marlowe, D. (1964). *The Approval Motive*. New York: Wiley.
Demo, D.H. (1985). The measurement of self-esteem: Refining our methods. *Journal of Personality and Social Psychology*, 48, 1490–1502.
Dovidio, J. and Fazio, R. (1992). New techniques for the direct and indirect assessment of attitudes. In J. Tanur (ed.), *Questions about Questions*. New York: Russell Sage, pp. 103–27.
Farnham, S.D. and Greenwald, A.G. (1997). Implicit balance between personal and social identity: I am female + I am good = female is good. Poster presented at the meeting of the Society of Experimental Social Psychology, Toronto, October 1997.
Fazio, R.J., Jackson, J.R., Dunton, B.C., and 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, 1013–27.
Fazio, R.H., Sanbonmatsu, D.M., Powell, M.C., and Kardes, F.R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–38.
Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
Fiske, S. and Pavelchak, M. (1986). Category-based versus piecemeal-based affective responses: Developments in schema-triggered affect. In R. Sorrentino and E. T. Higgins (eds), *Handbook of Motivation and Cognition: Foundations of social behavior*. New York: Guilford, pp. 167–203.
Fleming, J.S. and Courtney, B.E. (1984). The dimensionality of self-esteem: II. Hierarchical facet model for revised measurement scales. *Journal of Personality and Social Psychology*, 46, 404–21.
Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Doubleday.
Greenberg, J., Solomon, S., Pyszczynski, T., Rosenblatt, A., Burling, J., Lyon, D., Simon, L., and Pinel, E. (1992). Why do people need self-esteem? Converging evidence that self-esteem serves an anxiety-buffering function. *Journal of Personality and Social Psychology*, 63, 913–22.
Greenwald, A.G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35, 603–18.
Greenwald, A.G. and Banaji, M.R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102, 4–27.
Greenwald, A.G. and Breckler, S.J. (1985). To whom is the self presented? In B.R. Schlenker (ed.), *The Self and Social Life*. New York: McGraw-Hill, pp. 126–45.
Greenwald, A.G., Klinger, M.R., and Liu, T.J. (1989). Unconscious processing of dichoptically masked words. *Memory and Cognition*, 17, 35–47.
Greenwald, A.G., McGhee, D.E., and Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74, 1464–80.
Greenwald, A.G. and Pratkanis, A.R. (1984). The self. In R.S. Wyer, Jr and T.K. Srull (eds), *Handbook of Social Cognition*. Hillsdale, NJ: Erlbaum, pp. 129–78.
Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.
Hewitt, J. and Goldman, M. (1974). Self-esteem, need for approval, and reactions to personal evaluations. *Journal of Experimental Social Psychology*, 10, 201–10.
James, W. (1890). *Principles of Psychology*, Vol. 1. New York: Henry Holt.
Kaplan, H.B. (1975). *Self-attitudes and Deviant Behavior*. Pacific Palisades, CA: Goodyear Publishing Company.
Kihlstrom, J.F. and Cantor, N. (1984). Mental representations of the self. In L. Berkowitz (ed.), *Advances in Experimental Social Psychology*, Vol. 17. Orlando, FL: Academic Press, pp. 1–47.
Lindeman, M. and Verkasalo, M. (1995). Personality, situation, and positive–negative asymmetry in socially desirable responding. *European Journal of Personality*, 9, 125–34.
Lobel, T.E. and Levanon, I. (1988). Self-esteem, need for approval, and cheating behavior in children. *Journal of Educational Psychology*, 80, 122–3.
Lobel, T.E. and Teiber, A. (1994). Effects of self-esteem and need for approval on affective and cognitive reactions: Defensive and true self-esteem. *Personality and Individual Differences*, 16, 315–21.
Luhtanen, R. and Crocker, J. (1992). A collective self-esteem scale: Self-evaluation of one's social identity. *Personality and Social Psychology Bulletin*, 18, 302–18.
Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 35, 63–78.
Markus, H. and Wurf, E. (1986). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, 38, 299–337.
Marsh, H.W. (1986). Global self-esteem: Its relation to specific facets of self-concept and importance. *Journal of Personality and Social Psychology*, 51, 1224–36.
Mikulincer, M. (1995). Attachment style and the mental representation of the self. *Journal of Personality and Social Psychology*, 69, 1203–15.
Mueller, J.H. and Grove, T.R. (1991). Trait actualization and self-reference effects. *Bulletin of the Psychonomic Society*, 29, 13–16.
Murphy, S.T. and Zajonc, R.B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, 64, 723–39.
Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, 106, 226–54.

Paulhus, D.L. (1986). Self-deception and impression management in test responses. In A. Angleitner and J.S. Wiggins (eds), *Personality Assessment via Questionnaire* New York: Springer-Verlag, pp. 143–65.

Pelham, B. and Swann, W. (1989). From self-conceptions to self-worth: On the sources and structure of global self-esteem. *Journal of Personality and Social Psychology*, 57, 672–80.

Raskin, R., Novacek, J., and Hogan, R. (1991). Narcissism, self-esteem, and defensive self-enhancement. *Journal of Personality*, 59, 19–38.

Rosenberg, M. (1965). *Society and the Adolescent Self-image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). *Conceiving the Self*. New York: Basic Books.

Ross, M.J., Jurek, A.W., and Oliver, J.M. (1996). Stimulus emotionality and processing information about the self. *Journal of Social Behavior and Personality*, 11, 43–50.

Rudman, L.A., Greenwald, A.G., and McGhee, D.E. (1998). Sex differences in gender stereotypes revealed by the implicit association test. Unpublished manuscript. Rutgers University and University of Washington.

Sackeim, H.A. and Gur, R.C. (1978). Self-deception, self-confrontation and consciousness. In G.E. Schwartz and D. Shapiro (eds), *Consciousness and Self-regulation: Advances in research*, Vol. 2. New York: Plenum Press, pp. 139–97.

Schlenker, B.R. (1980). *Impression Management*. Monterey, CA: Brooks-Cole.

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

Schneider, D.J. and Turkat, D. (1975). Self-presentation following success or failure: Defensive self-esteem models. *Journal of Personality*, 43, 127–35.

Shiffrin, R.M. and Schneider, W. (1977). Control and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127–90.

Shrauger, J.S. (1975). Responses to evaluations as function of initial self-perception. *Psychological Bulletin*, 82, 581–96.

Swann, W. B. (1990). To be adored or to be known? The interplay of self-enhancement and self-verification. In R. M. Sorrentino and E. T. Higgins (eds), *Handbook of Motivation and Cognition*, Vol. 2. New York: Guilford Press, pp. 408–48.

Tajfel, H. and Turner, J.C. (1979). An integrative theory of intergroup conflict. In W.G. Austin and S. Worchel (eds), *The Social Psychology of Intergroup Relations*. Monterey, CA: Brooks-Cole, pp. 38–48.

Tajfel, J. and Turner, J.C. (1986). The social identity theory of intergroup behaviour. In S. Worchel and W.G. Austin (eds), *Psychology of Intergroup Relations*, 2nd edn. Chicago: Nelson-Hall, pp. 7–24.

Taylor, S.E. and Brown, J. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.

Tice, D. (1991). Esteem protection or enhancement? Self-handicapping motives and attributions differ by trait self-esteem. *Journal of Personality and Social Psychology*, 60, 711–25.

Wells, L.E. and Marwell, G. (1976). *Self-esteem: Its conceptualization and measurement*. Beverly Hills, CA: Sage.

Zajonc, R.B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35, 151–75.

11

Joining Groups to Reduce Uncertainty: Subjective Uncertainty Reduction and Group Identification

MICHAEL A. HOGG AND BARBARA-A. MULLIN

Henri Tajfel's early articulation of cognitive processes and structures with social processes and structures (e.g., Tajfel, 1969) quickly developed into social identity theory as an attempt to understand intergroup behavior (e.g., Tajfel and Turner, 1979; also see Hogg and Abrams, 1988). Although initially a distinctly European approach (cf. Tajfel, 1984), social identity theory now spans national boundaries and has had a significant impact on a revitalized social psychological interest in group processes (Abrams and Hogg, 1998; Hogg and Moreland, 1995; Moreland, Hogg, and Hains, 1994).

Social identity theory specifies how social categorization and social comparison processes work in conjunction with social belief structures to produce specific forms of group behavior. The theory has come to emphasize self-evaluation and the need for self-esteem as the principal motivational mechanism (e.g., Turner, 1982; cf. Hogg and Abrams, 1990), and as a consequence, the motivational analysis has become associated with social comparison and self-enhancement rather than social categorization. Evidence for the role of self-esteem in social identity processes is mixed – a state of affairs which may reflect methodological, operational, conceptual, or level of analysis issues, or a combination of all three (Abrams and Hogg, 1988; Hogg and Abrams, 1990, 1993; Long and Spears, 1997; Mullin, 1998; Rubin and Hewstone, 1998).

With the rapid development of contemporary social cognition in the 1980s, some social identity theorists shifted their conceptual emphasis away from social comparison and onto social categorization, to produce self-categorization theory (Turner, 1985; Turner, Hogg, Oakes, Reicher, and Wetherell, 1987). Farr (1996, see p. 10), in his historical analysis of the roots of modern social psychology, considers self-categorization theory to be the expression of modern social cognition in social identity theory. Indeed, many scholars, including advocates of social identity theory, would agree and some might lament this development (cf. the diversity of perspectives represented by chapters in Robinson, 1996). An alternative view (e.g., Hogg, 1996; Hogg, Terry, and White, 1995), however, is