Implicit Self and Identity

THIERRY DEVOS
MAHZARIN R. BANAJI

When William James (1890) wrote about the unique problem of studying self and identity, he immediately noted the peculiar blurring of the otherwise clear demarcation between the knower and the known. The object of scrutiny, the self, was also the agent doing the scrutinizing. This illicit merger of the knower and the known has created an epistemological unease that philosophers have worried about and psychologists have either ignored or turned into an assumption of their theorizing. The human ability for self-awareness and self-reflection is so unique that tapping it as a vital source of information about mind and social behavior has come at the expense of confronting the severe problems of the knower also being the known and of using introspection as the primary path to discovery. In this chapter, we argue that at least one circumstance can disentangle the knower from the known in the study of self; when self-as-knower does not have full introspective access to self-as-known. When knowledge about oneself resides in a form that is inaccessible to consciousness but can indeed be tapped indirectly, the self-as-knower and the self-as-known can be dissociated in a manner that is epistemologically more pleasing. In this chapter, we focus on states of unconscious thought and feeling—those marked by a lack of conscious awareness, control, intention, and self-reflection.

Over the past two decades, the study of implicit social cognition has created new paradigms for studying several traditional fields (for reviews, see Banaji, Lemm, & Carpenter, 2001; Greenwald & Banaji, 1995; Wegner & Bargh, 1998). At first sight, this trend, it would seem, has little to say about the topic of self and identity. Indeed, it is a common assumption that the studies of self centrally involve experiences of reflexive consciousness (Baumeister, 1998): Individuals reflect on their experiences, self-consciously evaluate the contents of consciousness, and introspect about the causes and meaning of things. In addition, the self is often viewed as playing a consciously active role in making meaning, implementing choices, pursuing goals, and initiating action. Studies that focus on unconscious modes of thinking and feeling, when applied to self and identity processes, question these assumptions, and they do so based on the discovery of mental acts that are fully meaningful and lawful but that appear to arise without introspective access or deliberative thought.

In this chapter, we provide an overview of
research on the implicit social cognition of self and identity. No attempt is made to exhaustively review the literature at hand. Rather, we focus on reflections of self and identity in a particular social context—the context in which thoughts and feelings about oneself are shaped by membership in a larger collective and in which such thoughts and feelings go beyond the self as target to represent and shape a view of the collective (Banaji & Prentice, 1994; Walsh & Banaji, 1997). Such a focus places us in the respectable company of others who also assume or demonstrate that the individual self is meaningfully considered in reference to social entities that transcend the individual self (Cooley, 1902; Mead, 1934; Turner, Oakes, Haslam, & McGarty, 1994). Our unique position limits the coverage to aspects of the self that emerge when (1) viewed in the context of social group memberships and (2) measured via unconscious expressions of thought and feeling. We begin with research paradigms that link the study of self with social group and proceed to specific analyses of basic preference for the ingroup and other attributes associated with the self. We then include analyses of implicit self and identity processes as viewed in research on self-evaluation, performance, and behavior, and goal pursuit. In the next major section we attend to the top-down influence of societal and cultural influences on the construction of implicit self and identity. Together, the research we review reveals the plasticity of the self as it develops and exists in close response to the demands of social group and culture.

The term "implicit" is used to refer to processes that occur outside conscious awareness. Evaluations of one's self, for example, may be influenced by group membership, even though the individual is not aware of such an influence. An Asian woman may come to view herself as excelling in math when her ethnic identity is implicitly brought to the foreground but as weak in math skills when her gender is highlighted (see Shih, Pittinsky, & Ambady, 1999, for a demonstration of such group membership effects on math performance). There are multiple ways in which one may be unaware of the source of influence on thoughts, feelings, and behavior. For example, one may be unaware of the existence of the source of influence, whereas in other circumstances one may consciously and accurately perceive the source of influence while being unaware of its causative role in self-evaluation. The term "implicit" is also applied to those processes that occur without conscious control. Here, the circumstances are such that one may be perfectly aware of the contingencies that connect a particular stimulus to a response but be unable to change or reverse the direction of the thought, feeling, or action. A person may have a view of herself as egalitarian but find herself unable to control prejudicial thoughts about members of a group, perhaps including groups of which she is herself a member. Although empirical investigations focus on one or another of these aspects of unconscious social cognition, as well as on those that select intention and self-reflection, we use the term "implicit" here to encompass both processes that occur without conscious awareness and those that occur without conscious control.

Self and Social Group

Since at least the 1970s, self-concept has been profitably studied by representing it as an information structure with empirically tractable cognitive and affective features (for reviews, see Greenwald & Pratkanis, 1984; Kihlstrom & Cantor, 1984; Kihlstrom et al., 1988). From such a theoretical vantage point came the idea that the self-concept, like other representations, could be viewed as possibly operating in automatic mode and that aspects of self may be hidden from introspective awareness (e.g., Bargh, 1982; Bargh & Tota, 1988; Higgins, Van Hook, & Dorfman, 1988; Markus, 1977; Rogers, Kuiper, & Kirker, 1977; Strauman & Higgins, 1987).

Although strongly social in focus, research in the American social cognition tradition, focused on the interpersonal aspects of self and identity, whereas another tradition with European roots, emphasized the association between self and social group, resulting in an intergroup emphasis. The latter's most articulate and encompassing formulation, labeled self-categorization theory (Turner, 1985; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987), holds that, under
particular conditions, group members perceive themselves as exemplars of the group rather than as unique individuals. In this mode, they highlight the similarities between themselves and other ingroup members, and they apply characteristics typical of the ingroup to the self (self-stereotyping). In other words, the representations of self and ingroup become inextricably linked. Until recently, tests of this hypothesis mainly involved self-report measures (Billiet, Vescio, & Green, 1996; Hogg & Turner, 1987; Simon & Hamilton, 1994; Simon, Patalese, & Mummendey, 1995). However, several empirical investigations have revealed that the process by which the ingroup may be said to become part and parcel of the self can operate at an implicit level.

Adapting a paradigm developed by Aron, Aron, Tudor, and Nelson (1991), Smith and Henry (1996) examined people's psychological ties to significant ingroups. Participants were asked to rate themselves, their ingroup, and an outgroup on a list of traits. Next, they indicated, as quickly and accurately as possible, whether each trait was self-descriptive or not. Self-descriptiveness judgments were faster for traits on which participants matched their ingroup than for traits on which they mismatched. On the contrary, no such facilitation was observed for traits rated as matching or mismatching the outgroup. This finding has been taken to illustrate that the ingroup becomes part of the representation of self, and a follow-up study (Smith, Coats, & Walling, 1999) demonstrated that the reverse was also true: Characteristics of the self influenced evaluations of the ingroup. Using a similar procedure, Smith and colleagues (1999) found that participants were faster to make ingroup descriptiveness judgments for traits that matched their self-perceptions. As another example, fraternity or sorority members are faster to make liking judgments for attitude objects (e.g., parties, tattoos, science, beer) on which they match the ingroup rather than on those on which there is a mismatch with the ingroup (Coats, Smith, Claypool, & Banner, 2000). Together, these results support the idea of a mental fusion of the self and social group.

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is a technique developed to assess the strength of implicit associations between concepts (e.g., self, group) and attributes (e.g., evaluation of good–bad, specific traits). The assumption underlying the technique is that the more closely related a concept and an attribute are (e.g., ingroup and good, outgroup and bad), the more quickly information representing the concept and the attribute should be paired. For purposes of obtaining a baseline with which such pairings can be compared, the task includes a measure of response time to contrasting pairs that are made simultaneously (ingroup and bad, outgroup and good).

Recent experiments have used this technique and variations of it to investigate the strength of self + group association, referring to this pairing as a measure of automatic identity with the social group. For example, Devos and Banaji (2001) used this procedure to capture the strength of implicit national identity among citizens of the United States. Participants were asked to categorize, as quickly as possible, stimuli presented on a computer screen. Some stimuli were pictures of American or foreign symbols (e.g., flags, coins, maps, monuments), whereas other stimuli were pronouns frequently used to designate ingroups (e.g., "we," "ourselves") or outgroups (e.g., "they," "other"). Participants completed this task twice. In one case, American symbols were paired with words representing the ingroup (e.g., "we," "ourselves"), and foreign symbols were combined with words representing the outgroup (e.g., "they," "other"). In another case, American symbols were combined with outgroup words, and foreign symbols were paired with ingroup words. Results indicated that participants performed the categorization task more quickly when American symbols and ingroup words shared the same response key. In other words, it was easier to associate American symbols with items such as "we" or "ourselves" rather than with "they" or "other." American symbols may be seen here as automatically evoking belonging and implying that, at least when unable to control their responses, this sample of Americans strongly identified with their national group.

Using the same technique, other empirical investigations have demonstrated strong im-
explicit associations between self and attributes, roles, or domains stereotypical of gender categories. For instance, female participants could more easily associate idiographic information (e.g., their names, their home-towns) or pronouns such as “me” or “mine” with feminine traits (e.g., gentle, warm, tender) rather than with masculine attributes (e.g., competitive, independent, strong), whereas the opposite was true for male participants (Greenwald & Farnham, 2000; Lenn & Banaji, 1998). Similarly, strong automatic associations between self and the concept “math” for men, and the concept “arts” for women have been obtained repeatedly (Nosek, Banaji, & Greenwald, 2002b). In addition, Lane, Mitchell, and Banaji (2001) have shown that implicit identification with a new ingroup could occur early on and without extensive contact with the group. As predicted, Yale students showed stronger implicit identity with Yale as an institution (rather than with Harvard), but strength of identity was equally strong among those who had been on campus for a few days and those who were starting their fourth and final year. Theories of implicit social cognition that assume slow learning (through long-term experience) need to explain the presence of such fast-to-form and fast-to-stabilize implicit identities.

The previous examples indicate that group membership comes to be automatically associated with self and that people automatically endorse attributes stereotypic of their group as also being self-descriptive. Recently, von Hippel, Hawkins, and Schooler (2001) identified circumstances under which the opposite is true: Counter-stereotypic attributes become strongly associated to the self. For example, African Americans were more likely to endorse the trait “intelligent” rapidly (i.e., to be schematic for this attribute) than white Americans if they performed well academically, and white Americans were more likely than African Americans to be schematic for “athletic” if they performed well in that domain. These findings are consistent with the idea that characteristics or features that make one distinctive from others are particularly likely to be represented in the self-concept (McGuire & McGuire, 1988). Both sets of seemingly opposing results may be explained by theories that emphasize the strong association between self and social group. In both cases, one’s knowledge or understanding of the association between group + attribute comes to influence implicit self-perception.

A Preference for Ingroups

The links between self and ingroup are not only visible in implicit knowledge and thought but also present in measures of attitude or evaluation. Tajfel (1974) emphasized this point when he defined “social identity as that part of an individual’s self-concept which derives from his knowledge of his membership of a social group (or groups) together with the emotional significance attached to that membership” (p. 69).

A growing body of research shows that people evaluate ingroup members more favorably than outgroup members (Brewer, 1979; Mullen, Brown, & Smith, 1992), and we examine those studies that used measure of implicit attitude or evaluation. The literature on implicit attitudes clearly suggests that groups unconsciously or automatically trigger more positive affective reactions when they are associated to self. Assessments of ethnic attitudes without perceivers’ awareness or control consistently reveal that white Americans have more positive feelings toward white Americans than toward African Americans (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald et al., 1998; Wittenbrink, Judd, & Park, 1997). Likewise, members of other groups also show implicit preference for the ingroup (Japanese Americans vs. Korean Americans; Greenwald et al., 1998), and findings such as these are expected to be obtained with a wide variety of groups (affiliations with nation, state, and city, with school and sports team, with family and friends). Some studies indicate that implicit intergroup biases in the realm of ethnic relations include both ingroup favoritism (positivity toward white Americans) and outgroup derogation (negativity toward black Americans) (Dovidio, Evans, & Tyler, 1986; Nosek & Banaji, 2001), whereas others suggest that individuals no longer differentially
Ascribe negative characteristics to ethnic groups but continue to associate positive attributes to a greater extent to whites than to blacks (Gaertner & McLaughlin, 1983). Research also shows that undergraduate students hold a more favorable attitude toward the category “young” than toward the category “old” (Nosek, Banaji, & Greenwald, 2002; Perdue & Gurman, 1990; Rudman, Greenwald, Mellott, & Schwartz, 1999). Strong implicit preferences for American symbols have been revealed in several studies (Ashburn-Nardo, Voils, & Monteith, 2001; Devos & Banaji, 2001; Rudman et al., 1999). To this list of implicit ingroup preference we add findings that show implicit positive attitudes toward groups based on sexual orientation (Lemmi & Banaji, 2000) or social class (Cunningham, Nezlek, & Banaji, 2001). In fact, Cunningham and colleagues (2001) have shown implicit positive associations to the category white (rather than black), rich (rather than poor), American (rather than foreign), straight (rather than gay) and Christian (rather than Jewish) among students known to be white, American, and Christian, a majority of whom were also assumed to be high on the social class dimension and to be heterosexual. Cunningham and colleagues have taken the extra step of claiming that these implicit preferences do not develop in isolation and that an individual difference marks the pattern: Those who show higher preference for one ingroup also show higher preference for all other ingroups, which is evidence for an implicit ethnocentrism dimension.

In most of the research described, researchers have assessed the implicit attitudes of only people belonging to one particular group. That is, white participants have been shown to have greater implicit liking for whites over blacks, but equivalent data from black participants have not always been available. Of the few studies that do measure both sides, symmetry has been found under some circumstances. Greenwald and colleagues (1998) reported data from both Japanese Americans and Korean Americans, each of whom showed a more positive implicit attitude toward their own ethnic group. The level of immersion in Asian culture moderated this pattern of implicit preferences. More precisely, participants who were immersed in their particular Asian culture (i.e., had a high proportion of family members and acquaintances who were from that culture and were familiar with the language) showed greater ingroup preference. Depending on their religious affiliation, individuals exhibit an implicit preference for Christian or Jewish people (Rudman et al., 1999). The relations between Bavarians and North Germans have also shown strong implicit ingroup bias (Neumann et al., 1998).

Using linguistic patterns to indicate intergroup bias, Maass, Salvi, Arcuri, and Semin (1989) revealed another interesting form of implicit ingroup favoritism. They show that individuals usually described positive behaviors in more abstract language terms (“X is helpful”) when performed by an ingroup member than when performed by an outgroup member (“X gave them directions to go to the station”). The opposite held for negative behaviors (see also Karpinski & von Hippel, 1996; von Hippel, Sekaquaptewa, & Vargas, 1997). This linguistic bias does not appear to be under volitional control (Franco & Maass, 1996). Recent research indicates that participants from Belgium and Spain automatically attributed secondary emotions, such as affection, pride, and remorse, more to ingroups than to outgroups (Leyens et al., 2000). Given that such secondary emotions are also seen to be uniquely human characteristics, this finding is consistent with the idea that ingroups are viewed as being more human than outgroups.

The tendency to favor the ingroup attitudinally (e.g., along a good–bad dimension) sometimes underlies implicit stereotyping (e.g., the assignment of specific qualities that may also vary in evaluation). For example, both men and women hold similar implicit gender stereotypes, but they are exhibited to a stronger extent when they reflect favorably on their own group (Rudman, Greenwald, & McGhee, 2001). Male participants are more likely to differentiate men and women with respect to an attribute such as “power,” whereas female participants are more likely to do so on a trait such as “warmth.” In other words, each group emphasizes stereotypes in an ingroup-favorable direction. These findings
parallel observations based on self-report measures as well (Lindeman & Sundvik, 1995; Peabody, 1968).

Using measures of consciously accessible cognition, the ingroup bias has been shown to emerge under minimal conditions: The mere categorization of individuals into two distinct groups elicits a preference for the ingroup (Diehl, 1990; Tajfel, Billig, Bundy, & Flament, 1971). There is now evidence that a minimal social categorization is sufficient to automatically or unconsciously activate positive attitudes toward self-related groups and negative or neutral attitudes toward non-self-related groups. For example, Perdue, Dovidio, Gurzman, and Tyler (1990) found that participants responded faster to pleasant words when primed with ingroup pronouns (such as “we” or “us”) rather than with outgroup pronouns (such as “they” or “them”), even though they were unaware of the group-designating primes. Thus the use of words referring to ingroups or outgroups might unconsciously perpetuate intergroup biases. More recently, Otten and Wentura (1999) showed that neutral words automatically acquired an affective connotation, simply by introducing them as group labels and by relating one of them to participants’ self-concepts. The self-related group label functioned equivalently to a priori positive primes, whereas the other label functioned similarly to a priori negative primes. In other words, as soon as a word designated an ingroup, it acquired a positive connotation, whereas words referring to an outgroup immediately conveyed a negative valence. Using a different procedure, Otten and Moskowitz (2000) also found evidence for implicit activation of positive affect toward novel ingroups. A minimal intergroup setting was combined with a well-established procedure measuring spontaneous trait inferences (Uleman, Hon, Roman, & Moskowitz, 1996). Results showed that when behaviors implied a positive trait about the ingroup, they were more likely to be categorized in a manner consistent with the implied trait than when the behaviors were performed by an outgroup member or when the traits implied were negative. These experiments suggest that the ingroup bias occurs automatically or unconsciously under minimal conditions (see also Ashburn-Nardo et al., 2001).

Preferences for Self Extend to Attributes Associated with Self

These findings are reminiscent of research showing that the mere ownership of an object or its association to the self is a condition sufficient enough to enhance its attractiveness. Nuttin (1985) found that when individuals were asked to choose a preferred letter from each of several pairs consisting of one alphabetical letter from their names and one not, they tended reliably to prefer alphabets that constitute their names. This finding, known as the “name letter effect” (NLE), has been replicated in many countries (Nuttin, 1987) and with samples from very different cultures (Hoorens, Nuttin, Herman, & Pavakanun, 1990; Kitayama & Karasawa, 1997). In addition, the preference for name letters has been shown to be stable over a 4-week period (Koole, Dijkstra, & van Knippenberg, 2001).

In order to test whether the preference for name letters depended on a conscious decision, Nuttin (1985) invited participants to search for a meaningful pattern in the pairs of letters presented. Despite the fact that no time limit was imposed and that a monetary award was promised to anyone who could correctly identify the rearranged pattern of letters, not a single participant could come up with the solution. This finding supports the idea that the NLE does not stem from a conscious recognition of the connection between the attribute and one’s self. Several alternative interpretations of the effect have been ruled out. For instance, the NLE does not seem to be a reminder of the positive mastery affect most people experienced when they first succeeded in reading or writing their own names (Hoorens & Todorova, 1988; Hoorens et al., 1990). In addition, the NLE does not appear to be due to an enhanced subjective frequency of own name letters as compared with non-name letters (Hoorens & Nuttin, 1993). The most convincing interpretation of this effect at present is that preference for letters in one’s name reflects an unconscious preference for self, and its generality is shown
through research on preference for other similar self-related information, such as birthdays over other numbers (Kitayama & Karasawa, 1997; Koole et al., 2001). Finch and Cialdini (1989) showed that this effect extended to greater liking for otherwise undesirable characters (e.g., Rasputin), and Prentice and Miller (1992) showed that sharing a birthday with another person led participants to be more cooperative in a competitive situation. Recently, Pelham, Mirenberg, and Jones (2002) have shown that the extent of implicit positive evaluation of self influences where people choose to live and what people choose to do for a living. Across a dozen studies, they found that people are more likely to live in cities or states and to choose careers whose names share letters with their own first or last names. For example, a person named Louis is disproportionately likely to live in St. Louis and individuals named Dennis or Denise are overrepresented among dentists. If corroborated with continued evidence of such relationships, these findings suggest that personal choices may be constrained by linkage to self that is not noticed, not consciously sought, and even surprising.

Interesting as such research is in suggesting automatic attitude preference for self, the studies are quasi-experimental in that name letters and birth dates are not manipulated variables. Feys (1991) has provided experimental evidence for the idea that mere ownership of an object is a sufficient condition to enhance its attractiveness. Participants first learned to discriminate four computer-displayed graphic icons (which represented the participant in a computerized game) from four others that represented the participant's opponent (the computer). When participants subsequently judged all eight patterns for aesthetic attractiveness, the self-associated patterns received higher ratings. Thus individuals evaluated an object more favorably simply from the act of owning it (see also Beggarly, 1992). Objects attached to the self are immediately endowed with increased value. These findings, along with the classic attitude-similarity effect, were interpretable as implicit self-esteem effects (Greenwald & Banaji, 1995); they reveal introspectively unidentified (or inaccurately identified) effects of the self-attitude on evaluations of associated objects.

Balancing Self and Social Group

Work reviewed so far highlights the cognitive and affective ties between self and group memberships and stresses the fact that individuals are not necessarily fully aware of these bounds on their thinking or that they are aware but unable to control their operation. Now we turn to the relationships between the cognitive and affective components that make up the self system. Several theories predict some consistency between constructs that represent self and social group. For example, social identity theory (Tajfel, 1974; Tajfel & Turner, 1986) assumes some interrelations between self-esteem, group identification, and ingroup bias. According to the theory, social identification serves as a source of self-esteem. Generally speaking, individuals strive to maintain or increase their self-esteem. They can derive a sense of self-worth through favorable intergroup comparisons: Group membership contributes to positive self-esteem if the ingroup can be positively differentiated from outgroups. Thus self-esteem should be enhanced by membership in a valued group, and strong identification with the group should go hand in hand with positive evaluation of the ingroup. Evidence for the role of self-esteem in intergroup comparisons is mixed (Abrams & Hogg, 1988; Brown, 2000; Hogg & Abrams, 1990; Rubin & Hewstone, 1998). Moreover, support for the idea that there should be a positive correlation between group identification and ingroup favoritism is not overwhelming (Brown, 2000; Hinkle & Brown, 1990); this relationship is obtained only under specific circumstances (Brown et al., 1992). The absence of expected relationships has led to examinations of these constructs using implicit measures. For example, Knowles, Peng, and Levy (2001) found that the strength of the automatic association between self and whites (ingroup identifications) was positively correlated with the intensity of the pro-white implicit attitude (ingroup favoritism) and also accounted for
the extent to which individuals possessed a restrictive representation of their ethnic group (ingroup exclusiveness).

Recently, Greenwald and colleagues (2002; see also Greenwald et al., 2000) proposed a unified theory of social cognition that predicts patterns of interrelations between group identification, self-esteem, and ingroup attitude. Their approach draws its inspiration from theories of affective-cognitive consistency that dominated social psychology in the 1960s (Abelson et al., 1968) and allows them to integrate a range of otherwise isolated findings obtained with the IAT (Greenwald et al., 1998). This approach is based on the assumption that social knowledge (including knowledge about the self) can be represented as an associative structure. From this point of view, the structure of the self is a network of associations: The self is linked to traits, groups, concepts, or evaluations. A core principle of the theory is that attitudes toward self and concepts closely associated with self (i.e., components of self-concept or identity) should tend to be of similar valence. In other words, according to the balance-congruity principle, if someone holds a positive attitude toward the self and considers that a particular concept (e.g., a group, an attribute, or a domain) is part of his or her self-concept, this person should also hold a positive attitude toward that particular concept.

A study on women’s gender identity illustrates this principle more concretely. For women, one would typically expect an association between self and the concept “female” (gender identity or self + female) and a positive association toward the self (positive self-esteem or self + good). Based on the balance-congruity principle, these two links should also be accompanied by a third link: a positive association toward the concept “female” (liking for female or female + good). More precisely, the strength of the positive attitude toward “female” should be a joint (or interactive) function of the strength of the associations between self and positive and between self and female. Data supported this prediction: As gender identity increased, so did the positive relation between self-esteem and liking for women (see also Farnham, Greenwald, & Banaji, 1999). Further support for similar hypotheses was found in studies on race and age identities. Interestingly, the balance-congruity principle always received stronger support when tested with implicit than with explicit measuring tools.

Another key principle of the theory, the imbalance-dissonance principle, is that the network resists forming new links that would result in a component being tied to pairs of bipolar-opposed constructs. A study on self-concept and gender stereotypes about mathematics demonstrates the heuristic value of this principle (Nosek et al., 2002b). Relative gender stereotypes link the concept “math” with “male,” and the concept “arts” with “female.” Both men and women displayed automatic associations fitting these stereotypes. Given that women also associate the self to the concept “female,” the imbalance-dissonance principle leads to the prediction that women should therefore not associate themselves with “math.” Using a variety of social groups and differing clusters of attributes that measure attitude, stereotype, and self-esteem, Greenwald and colleagues (2002) show evidence for balanced identities: The stronger the connection between self and social group, the more likely are the preferences and beliefs to follow in stereotypic fashion. Moreover, such effects were primarily obtained when implicit measures of self and group identity were used and appeared in weaker form on measures of conscious affect and cognition.

Self-Evaluation

Having shown the presence of self and social group connections on attitude and beliefs, we turn to research that demonstrates shifts in self-evaluation that also occur without conscious intention. For example, the unconscious activation of significant others has implications for self-evaluation. Baldwin (1992) proposed that the internalization of relationships involves the development of relational schemas; these cognitive structures represent regularities in patterns of interpersonal interactions. Often, the sense of self can be derived from such well-learned scripts of interpersonal evaluations. In other words, activated relational schemas shape self-evaluative reactions, even when
these schemas are primed below the level of awareness. Indeed, subliminal exposure to
the name of a critical versus an accepting significant other led participants to report
more negative versus positive self-evaluations (Baldwin, 1994). Similarly, undergraduate students evaluated their own research
ideas less favorably after being subliminally exposed to the disapproving face of their
department chair rather than the approving face of another person (Baldwin, Carrell, &
Lopez, 1990). These effects occurred only when the prime was a significant other. For
instance, Catholic participants rated themselves more negatively after exposure to the
disapproving face of the Pope, but not after exposure to the disapproving face of an
unfamiliar person. In addition, if the Pope did not serve as a figure of authority, self-evaluation
remained unaffected by the priming
manipulation.
Mitchell, Nosek, and Banaji (2000) have shown that subtly varying the race and gen-
der composition of the context in which a particular social group is evaluated strongly
affected implicit attitudes. For example, they found that white women expressed a
more negative attitude toward black females when race (rather than gender) was the
distinctive categorization criterion. On
the other hand, highlighting gender elicited a relatively negative attitude toward white
males, whereas making race salient trig-
gerated a more positive attitude. Although
these studies do not speak directly to the issue of self and identity, they lend support to
the idea that intergroup biases vary as a function of the frame through which a given
situation is filtered.
Unobtrusively making a social identity salient can also influence self-evaluation. S.
Sinclair, Hardin, and Lowery (2001) asked participants to indicate how others and
they themselves evaluated their verbal and math ability. Before they made these
judgments, participants reported either their
gender or ethnic identity. This subtle manip-
ulation affected impressions attributed to
others. For example, when their gender
identity was salient, Asian American
women reported that others evaluated them
higher in verbal ability than math ability,
but when their ethnic identity was salient,
they stated that others evaluated them higher in math ability than verbal ability. This
social identity manipulation not only affect-
ed perceived evaluations of others but also
translated into self-stereotyping effects:
Changes in self-evaluations paralleled changes in evaluations of the self by others.
In other words, individuals were more likely to endorse the stereotype associated with
their group when their membership in that particular group was subtly implied. It is
quite unlikely that participants were aware
of the impact of the manipulation on their
self-evaluations, and the results may be taken as evidence of implicit self-stereotyping.
Internalized expectations about one's social
group can shape self-evaluations, though
perhaps only or especially when they are
unobtrusively activated.
Research on implicit self-esteem also indi-
cates that contextual variations can produce
an effect on unconscious or automatic pref-
rences. For example, Bylsma, Tomaka,
Luhtanen, Crocker, and Majori (1992) demon-
strated that self-descriptive judgments
were faster for positive adjectives after posi-
tive than after negative feedback. Using the
name-letter test described earlier, a recent
study shows that such an effect did not oc-
cur after participants had received failure
feedback on an alleged IQ test but that it
remerged once participants were given the
opportunity to affirm a personally impor-
tant value (Koole, Smeets, van Knippen-
berg, & Dijkstra Huis, 1999). Thus it ap-
pears that a failure on an alleged
intelligence test increases the accessibility of failure-related cognitions and reduces, at
least temporarily, participants' implicit self-
estee m. Affirming an important aspect of
one's self-concept permits counteracting the
negative consequences of the feedback. Such
work illustrates the dynamic nature of self-
related processes (Markus & Wurf, 1987). It
is a fact of modern life that people belong to
a range of social groups, both chosen and
given. As societies become more heterogeneous, the opportunity for comparing and
contrasting oneself to others will only in-
crease. Across time and situations, varying identities may come forward or recede from
consciousness. Effects that appear to be un-
 systematic and unpredictable may be quite
lawful when unconscious social influences
on self-evaluations are considered. Routine
use of measures of implicit social cognition
will need to be considered in a wide range
of research topics, rather than relegating them to a subset of research on particular topics, as appears to be the case at present.

Performance and Behavior

If thoughts and feelings are transformed by the activation of social group membership, it is expected that behavior should be influenced as well. Yet, because cognition and affect are much better understood systems than behavior, studies of the latter have been less frequently reported. Perhaps for this reason, and because behavior is the gold standard in this science, studies that show the influence of social group on self-relevant behavior are attention getting. This is certainly true of work on stereotype threat, or situations in which the presence of a negative stereotype about one's group can handicap the test performance of members of the group (Steele, 1997; Steele & Aronson, 1995). According to the proponents of this theoretical framework, when African American students perform a scholastic or intellectual task, they face the threat of confirming a negative stereotype about their group's intellectual ability. This threat, it is speculated, interferes with intellectual functioning and can lead to detrimental impact on performance. Support for this argument has now been obtained in many experiments showing the influence of subtle activation of race/ethnicity, gender, class, and age distinctions on performance on standardized tests. For example, Steele and Aronson (1995) found that stereotype threat can affect the performance of African American college students, who performed significantly worse than white Americans on a standardized test when the test was presented as diagnostic of their intellectual abilities. This effect did not occur when the test was presented as nondiagnostic of their ability. Other studies have demonstrated that women underperform on tests of mathematical ability when the stereotype associated with their group was made salient (Spencer, Steele, & Quinn, 1999). Shih and colleagues (1999) showed that activating gender identity or ethnic identity among Asian women shifted performances to be respectively inferior or superior on a math test. Students of low socioeconomic status (SES) performed worse than students of high SES only when a test was presented as diagnostic of their intellectual abilities (Croizet & Claire, 1998). The manipulations producing these effects are often rather subtle. In some cases, it is sufficient to ask participants to indicate their group membership just prior to assessing their performance (Steele & Aronson, 1995). Using age as the social category, Levy (1996) showed that subliminally activated negative stereotypes about the elderly produced decrements in the memory performance of elderly participants. In contrast, the activation of positive stereotypes of the elderly improved their memory performance. Interestingly, the same manipulations did not affect the performance of young individuals, who should not be susceptible to the threat posed by the negative stereotype about elderly individuals.

There is now considerable evidence that the activation of trait constructs or stereotypes can automatically or unconsciously influence social behavior. When trait constructs or stereotypes are primed in the course of an unrelated task, individuals subsequently are more likely to act in line with the content of the primed trait construct or stereotype (Bargh, Chen, & Burrows, 1996). For instance, if participants were exposed to words related to rudeness, they were more likely to interrupt an ongoing conversation. If, instead, they were primed with words related to politeness, chances were higher that they would respond in a polite fashion, waiting for the conversation to end without interruption. Priming the stereotype of the elderly caused participants to walk more slowly down the hallway after leaving the experiment. Not only did the subliminal presentation of African American faces lead participants to behave with greater hostility, it also increased the hostility of the person they were interacting with (Chen & Bargh, 1997). Thus the effect is not restricted to commonly accepted social groups, such as gender, class, or race/ethnicity, but works through the meaningful activation of information (as in the polite–rude study). Yet social groups tend to be among the dimensions of social life that provide clear and consensual stereotypes and may be particularly effective at producing a connection to oneself.
For instance, studies showed that priming a stereotype of professors or the trait "intelligent" enhanced performance on a general knowledge task (similar to Trivial Pursuit) and that priming the stereotype of soccer hooligans or the trait "stupid" decreased performance on the test (Dijksterhuis & van Knippenberg, 1998). More important, recent findings suggest that these effects are mediated by passive perceptual activity and are direct consequences of environmental events (priming manipulations). Indeed, manipulations or factors known to produce changes in perception also affected behaviors. For example, priming stereotypes of social categories produced assimilation effects like the ones we just described, whereas activating specific exemplars of the same categories led to contrast effects (Dijksterhuis et al., 1998). More precisely, if participants were primed with the category "professors" (rather than "supermodels"), their own intellectual performance was enhanced (assimilation effect), but if they were primed with the exemplar "Albert Einstein" (rather than "Claudia Schiffer"), a decrement in their performance resulted (contrast effect).

The strength of stereotypic associations also mediates the impact of experience on behavior (Dijksterhuis, Aarts, Bargh, & van Knippenberg, 2000). Those who report having high contact with the elderly showed a stronger association between the category "elderly" and the attribute "forgetful" than people who had less contact with the elderly. We speculate that such an effect of greater stereotyping that relies on direct experience can exist in the presence of a positive attitude toward the group. Those with a positive attitude toward the elderly but with experience that strengthens a stereotype (perceiving them to be forgetful) may have no recourse but to show the negative stereotype even in the face of a consciously positive attitude.

Other studies have demonstrated that individuals can fail to detect changes in their actions when those actions were induced implicitly. For example, people can be unaware that their behaviors shift in accordance with the behaviors of others. Chartrand and Bargh (1999) coined the term "chameleon effect" to describe the tendency to mimic unconsciously the postures, mannerisms, or facial expressions of one's inter-action partners. They showed that the mere perception of another's behavior automatically increased the likelihood of engaging in that behavior oneself. Individuals were more likely to rub their faces or shake their feet if they interacted with someone who was performing that behavior. Such an effect is assumed to serve an adaptive function by facilitating smooth social interaction through increases in liking between individuals involved in the interactions.

Vorauer and Miller (1997) also observed that people engaged in behavior matching and that they were unaware of doing so. Individuals who read another's positive self-description conveyed more positive impressions of their own satisfaction with academic achievements and university experiences compared with those who first read another's negative self-description. Participants' own sense of the impression they had conveyed was not affected by this manipulation. Thus people failed to detect the extent to which they were induced to present a particular impression of their feelings and experiences. The extensive information that individuals possess about their general personal qualities, as well as about their feelings in the moment, may make it difficult to perceive social influence on their actions. In each of these cases, the irony is that, of all the special domains of knowledge, self-knowledge is assumed to be well known and well defended against intrusion. Such discoveries both highlight the pervasiveness of social influences on the self and point to the inadequacy of introspection as the only tool for obtaining self-knowledge. Indirect measures reveal the subtle but important ways in which the social construction of self unfolds and the lack of conscious access to the process.

Self-Motives and Goal Pursuits

In recent years, research on self and identity has put a greater emphasis on the motivational mechanisms that propel social behavior. In particular, the role of self-enhancement has been the focus (Baumeister, 1998; Kunda, 1990; Sedikides, 1993; Swann, 1990), with investigations of a desire for positive feedback about the self and self-protective reactions unleashed by
threatening experiences. Several lines of research suggest that unconscious or automatic processes are triggered when the self or the ingroup is threatened. Spencer, Fein, Wolfe, Fong, and Dunn (1998) showed that threat to self-image can automatically activate stereotypes of social groups even under conditions that otherwise do not produce such activation (i.e., under cognitive load; see Gilbert & Hixon, 1991). Following negative feedback, participants were more likely to view others in stereotypic ways, and this reaction has been taken to suggest a dynamic process of restoring a positive self-image (Fein & Spencer, 1997). Recently, L. Sinclair and Kunda (1999) demonstrated that a self-enhancement motive not only prompted stereotype activation but also led to an inhibition of applicable stereotypes. Participants motivated to disparage a black doctor (because of his criticism of them) inhibited the stereotype of doctors and used a race stereotype, whereas participants motivated to esteem the same target (because of his praise of them) inhibited the race stereotype and relied on the doctor stereotype. In these examples, the threat was directed toward the self, but similar reactions might also be at stake when the ingroup is threatened. Maass, Ceccharelli, and Rudin (1996) showed that the linguistic intergroup bias (the propensity to describe in more abstract ways behaviors depicting the ingroup favorably) was intensified when an outgroup was seen to threaten the ingroup, supporting the idea that this bias is a result of a motivation to protect the ingroup. Pratto and Shih (2000) investigated the links between implicit intergroup biases and the personality variable captured by social dominance orientation (SDO), the degree to which people endorse ideologies that justify hierarchical relationships among groups in society (Pratto, Sidanius, Stallworth, & Malle, 1994). Using a priming technique, they found that individuals high in SDO exhibited stronger implicit intergroup biases than people who generally support group equality, but this difference occurred only following a threat to participants' social identity.

Individuals do not merely seek to preserve or establish a positive self-image; they also feel the need to be similar to others (need for differentiation). According to the optimal distinctiveness model (Brewer, 1991), identity stems from a tension between these two needs. Individual security and well-being are threatened when one of these needs is not met: Excessive depersonalization no longer offers a basis for self-definition, and excessive individualization renders one vulnerable to isolation and stigmatization. These motives affect people's identification even at the implicit level. Using a measure of speed of response adapted by Smith and Henry (1996), Brewer and Picket (1999) found that the need for assimilation increased students' implicit identification with their university, whereas the need for differentiation decreased their group identification.

Work based on Bargh's (1990) automatic model is centrally relevant to this discussion, beginning with the idea that pursuit of goals can occur automatically and nonconsciously. Goals activated outside of awareness, control, or intention are pursued similarly to goals chosen through deliberate or conscious means. For example, Chartrand and Bargh (1996) demonstrated that information processing goals, such as impression formation or memorization, can be automatically activated and pursued. Gardner, Bargh, Shellenbou, and Bessenhoff (2000) observed that the same brain region reacts whether an evaluation goal is consciously or unconsciously activated. Individuals primed nonconsciously with an achievement goal performed better on an achievement task and were more likely to persist at the task than individuals who were not primed with such a goal (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001). In addition, Chartrand (2000) found that success or failure at a nonconsciously activated goal led to the same consequences that resulted from success or failure at conscious goal pursuit. Specifically, succeeding at a nonconscious goal improved people's mood, and failure depressed it, paralleling shifts in mood usually occurring in the presence of consciously activated goals.

This research highlights the similarities between conscious and nonconscious self-motives or goals. Other researchers have examined the extent to which implicit and explicit motives differ. McClelland, Koestner,
and Weinberger (1989) investigated the differences between self-reported motives (explicit) and motives identified in associative thoughts (implicit). Among other things, they make the case that implicit motives predict spontaneous behavioral trends over time, whereas self-attributed motives predict immediate responses to specific situations. They also report evidence suggesting that implicit motives represent a more primitive motivational system derived from affective experiences, whereas self-attributed motives are based largely on more cognitively elaborated constructs. This finding is consistent with research by Woike (1995) on the relationship between implicit or explicit motives and most memorable experiences. Implicit motives were associated with affective experiences about the implicit motive, whereas explicit motives were related to routine experiences corresponding to self-descriptions.

Levesque and Pelletier (2000) argued that automatic processes could regulate intrinsic and extrinsic motivations and that these processes were functionally distinct from their explicit counterparts. Intrinsic motivation involves doing activities for the pleasure and satisfaction inherent in doing them, whereas extrinsic motivation refers to behaviors that are performed for instrumental reasons, that is, in order to attain a goal or an outcome. They found that cognitive structures relevant to intrinsic and extrinsic motivations could temporarily be made accessible through priming of associated constructs and that they influenced self-reported motivation. In addition, explicit and implicit motivations predicted different outcomes: Self-reported intrinsic motivation predicted immediate intentions about long-term behavior, whereas chronically accessible intrinsic motivation predicted the actual long-term behavior.

Research on self and identity has documented the pervasiveness of self-presentational concerns (Baumeister, 1982; Leary, 1995), and a common claim is that techniques assessing implicit attitudes or beliefs are usually free of self-presentational concerns. However, such an argument assumes that when people try to make a good impression, they are fully aware of doing so. Research on implicit self-motives and goals raises the possibility that such motives or goals may operate unconsciously and that self-presentation itself is a complex process that could include components that are strategic but inaccessible to conscious awareness and control.

Societal and Cultural Foundations

We now turn our attention to the influence of societal and cultural realities on implicit identities. We have indicated already that stereotypes about social groups have a profound impact on the implicit self. Automatic associations involving the self often reflect an internalization of cultural stereotypes. Other lines of research bring to the fore the societal roots of implicit processes. We begin with the premise that more often than not, relations between groups are hierarchically organized (Sidanius & Pratto, 1999). In other words, social groups rarely occupy interchangeable positions, and groups that enjoy greater social favors usually remain in that position for extended periods, whatever may be the criteria that characterize the hierarchy (e.g., numerical status, social status, or power; Sachdev & Bourhis, 1991). What is the impact of these factors on social identities? To what extent do members of dominant and subordinate groups exhibit a preference for their own group? On this issue, contrasting predictions can be formulated. On one hand, one would expect that members of subordinate groups engage in more ingroup bias than members of dominant groups. This would be consistent with the idea that the former have a stronger need to achieve a positive social identity, which should be satisfied by increasing favorable intergroup distinctions. On the other hand, we might hypothesize that members of subordinate groups are less likely than members of dominant groups to display a preference for their group because social conditions consistently impose a less favorable evaluation of the subordinate group. At least in the case of natural groups, the evidence at hand seemed to support the first alternative. A meta-analysis conducted by Mullen and colleagues (1992) indicated that members of low-status groups had a tendency to exhibit a stronger ingroup bias than members of high-status groups. In the case of ethnic comparisons in the United
States, African Americans often display more ethnocentric intergroup perceptions than white Americans (Judd, Park, Ryan, Brauer, & Kraus, 1995).

However, a different pattern of findings has emerged with some regularity when implicit social identity has been examined. Data from a demonstration Web site, www.yale.edu/implicit, provide some insights on this issue (Nosek et al., 2002) that support those from more traditional laboratories (Greenwald et al., 2002). Web data show that on a measure of conscious feeling, white respondents report a preference for the group "white" over the group "black," and black respondents report an opposite and even stronger preference for their own group. This strong explicit liking reported by black respondents stands in sharp contrast to performances on the implicit measure. Unlike white respondents, who continue to show a strong preference for white over black on the implicit measure of liking, black respondents show no such preference. This pattern of results mimics laboratory data obtained with college students (Banaji, Greenwald, & Rosier 1997): Black students exhibited strong explicit liking and identification with their own ethnic group (compared with white students), whereas a reversed pattern was observed on implicit measures (with white students showing stronger implicit ingroup preference). In addition, Jost, Pelham, and Carvallo (2001) found that students from high- and low-status universities both implicitly associated academic characteristics with the higher status group and extracurricular activities with the lower status group. Moreover, students from the high-status university exhibited significant ingroup favoritism on an implicit measure, whereas students from the low-status university did not.

These findings illustrate that the expected ingroup preference effect is moderated by sociocultural evaluations of social groups. On explicit measures, disadvantaged group members exert effort to report positive attitudes, but the lower social standing of their group is internalized sufficiently so as to be detected in their failure to show an implicit preference for their own group. On the other hand, advantaged group members' preferences show the combined benefit of both ingroup liking and the sociocultural advantage assigned to their group. Such results are consistent with the notion of system justification (Jost & Banaji, 1994), or the idea that beyond ego justification and group justification lies the more insidious tendency to justify the system or status quo, even when it reflects poorly on one's self or group. Members of dominant groups share thoughts, feelings, and behaviors that reinforce and legitimize existing social systems, which is in their interest, but, surprisingly, so do members of less dominant groups. Examples reviewed in this section indicate that ideological bolstering can occur outside conscious awareness, and this prevents perceivers and even targets of prejudice from questioning the legitimacy of social arrangements.

It would be erroneous to claim that members of subordinate groups cannot display strong implicit preference for their group. It depends on the comparison group. Devos and Banaji (2001) found that Asian American and African American students showed only a slight implicit preference for their ethnic group when it was compared with white Americans. However, when their group was compared with another minority group, ingroup favoritism was more pronounced: Asians displayed strong preference for Asian Americans relative to African Americans, whereas African Americans showed the opposite effect. This finding is consistent with the idea that members of minorities seek to maintain a positive social identity through downward social comparisons (Tajfel & Turner, 1986; Wills, 1981). In addition, we should stress that members of ethnic minorities often display strong implicit preferences for the self relative to others (Banaji et al., 1997; Nosek, Banaji, & Greenwald, 2002a), suggesting that members of disadvantaged groups (who show low group esteem) do not show lower self-esteem; they are able to avail themselves of opportunities to protect self-esteem (Crocker & Major, 1989). Thus, to the degree that members of subordinate groups are influenced by cultural evaluations of their group, they may develop a negative social identity, but it may not necessarily translate into a negative view of self (Pelham & Hetts, 1999).

Very little research has analyzed the relationship between self and identities that may be in conflict. We have chosen to study
these by viewing connections between national and ethnic identity (Devos & Banaji, 2001). The United States is a perfect testing ground, being, as it is, a pluralist society composed of identifiable ethnic groups that vary in length of association, immersion into mainstream culture, and conditions of immigration. We investigated the extent to which ethnic groups are implicitly conceived as being part of America in a culture, that explicitly holds that all groups be treated equally. We assumed that the hierarchy present in American society would structure implicit beliefs about the links between ethnicity and American identity. More precisely, we hypothesized that white Americans are unconsciously viewed as being more essentially American and as exemplifying the nation, whereas ethnic minorities are placed psychologically at the margins (Sidanius, Feshbach, Levin, & Pratto, 1997; Sidanius & Petrocik, 2001).

Using techniques developed to assess implicit attitudes or beliefs (Greenwald et al., 1998; Nosek & Banaji, 2001), we examined the extent to which various ethnic groups (white, Asian, and African Americans) were associated with the category “American” (relative to “foreign”). For example, we asked participants to pair, as quickly as possible, American or foreign symbols (e.g., flags, maps, coins, monuments) with faces that varied in ethnicity (white, Asian, and African Americans) but were clearly understood to be American. Although participants were aware that all individuals were American, irrespective of ethnicity, consistently the data indicated that African and Asian Americans were less strongly associated with the category “American” than were white Americans. Participants categorized items faster when American symbols were paired with white American faces than with Asian or African American faces. These findings did not merely reflect the fact that members of ethnic minorities were viewed as more foreign; using the appropriate technique (Nosek & Banaji, 2001), we showed that they are excluded from the category “American.” Such implicit associations are sometimes consistent with people’s explicit beliefs. For example, Asian Americans are viewed as less American than white Americans at both explicit and implicit levels of responding. Indeed, people stated explicitity that they did not consider Asian and white Americans to be equally American even when both held citizenship; they also considered Asian Americans to have weaker ties to American culture than whites.

In other cases, we found strong discrepancies between explicit and implicit beliefs (Devos & Banaji, 2001). For instance, at the explicit level, African and white Americans were not differentiated. In particular, both groups were perceived to be strongly and equally tied to American culture. However, in sharp contrast to the parity expressed at the explicit level, African Americans were viewed unconsciously as less American than were white Americans. These implicit associations were revealed even when explicit beliefs or knowledge showed the contrary. For example, on explicit measures, in a domain such as track and field sports, in which black Americans dominate, we found that black athletes were more strongly associated with the category “American” than were white athletes. Participants explicitly endorsed the statement that black athletes represent to a greater extent than white athletes “what America is all about” or that they contribute to a greater extent “to the glory of the American nation.” However, at the implicit level, black athletes were less strongly associated with the category “American” than were white athletes. In an attempt to find a condition that would surely remove this bias, we tested the association between ethnic and national category by blatantly selecting exemplars who are known Asian Americans and known white foreigners. We still could not shake off the strong implicit association between white and American. Participants were able to associate “American” with known white foreigners (e.g., Gérard Depardieu, Katarina Witt, Hugh Grant) more swiftly than with known Asian Americans (e.g., Connie Chung, Kristi Yamaguchi, Michael Chang). In other words, even though people were fully aware that someone like Gérard Depardieu is not American and that Connie Chung is American, it remained easier to make the white + American connection. We conclude from these studies that the national identity of being American is associated with the ethnic identity of being white—and sufficiently intimately so as to be irremovable even when it is consciously rejected.
In the previous studies, participants were all white Americans. In separate data collections (Devos & Banaji, 2001) we examined the role of group membership and found that Asian American participants displayed very similar implicit associations. Among other things, they viewed their own group as being less American than the group “white,” showing an internalization that is detrimental to their personal and group interests. Indeed, such implicit association potentially hurts their national identity. African Americans, on the other hand, did not display the same pattern of associations. Although viewed by white participants to be less American, they themselves perceived their own group to be as American as the group “white” and more American than the group “Asian American.” Implicit associations are rooted in experiences, they bear the mark of cultural socialization, and they reflect differences between ethnic groups at these levels. African Americans, perhaps because of the presence of other minorities who may be seen as less American, do not internalize the belief that resides in the minds of the advantaged majority, whereas Asian Americans do.

In sum, despite declarations of equality before the law, under a variety of circumstances, white Americans are unconsciously conceived of as being more American than Asian Americans or African Americans. Such implicit beliefs both reflect the hierarchy between ethnic groups within American society and contribute to the preservation of existing social arrangements (Jost & Banaji, 1994).

Research on culture and self-concept shows that members of different cultures often define and evaluate the self in different ways. A major distinction in cross-cultural psychology is the opposition between collectivist and individualist societies (Triandis, Bontempo, Villareal, Asai, & Lucca, 1988; Triandis, McCusker, & Hui, 1990). In collectivist cultures, people define themselves as members of groups, subordinate their personal goals to group goals, and show strong emotional attachment to the group. In individualist cultures, people place a strong emphasis on self-reliance, individual achievement, and personal goals. In their work on the self-concept, Markus and Kitayama (1991, 1994) argued that the self is defined in terms of interdependence in Asian cultures. In other words, the self is inherently collective in these cultures. In contrast, the typically Western conception of self is one in which individuals see themselves as distinct and independent from others. Hetts, Sakuma, and Pelham (1999; see also Pelham & Hetts, 1999) used this distinction to compare the implicit and explicit self-concepts of people who varied in their exposure to individualistic cultures but who were currently living in the same culture. More precisely, they examined to what extent explicit and implicit self-evaluations of recent Asian immigrants differed from those of European Americans and Asian Americans reared in the United States. At the explicit level, they found little difference between these groups. In particular, Easterners emigrating to a Western culture seemed to endorse the kind of self-concept promoted in individualistic societies. However, a different picture emerged at the implicit level. Using response latency and word-completion techniques, Hetts and colleagues found strong differences between groups in terms of personal versus group regard. For people reared in an individualistic culture, ideas that were automatically associated with the individual and collective identities were relatively positive. For people socialized in a collectivist culture, the group or collective identity automatically elicited positive thoughts, but ideas tied to individual identity were neutral, ambivalent, or even negative. Such discoveries are consistent with the idea that the need for positive regard is expressed through social or collective identities in some cultures and in individualistic ways in others.

In sum, the cultural context can overshadow differences in cultural experiences when measured through explicit self-evaluations, but implicit self-evaluations continue to reveal the mark of cultural socialization. These results could be taken as evidence that implicit self-evaluations are less influenced by normative demands than their explicit counterparts. More fundamentally, they also suggest that implicit self-evaluations are slower to change. Because they are overlearned associations rooted in experiences, it may take substantial or highly salient contradictory experiences to shift them. Consistent with this argument, Hetts
and colleagues (1999) report that implicit self-evaluations of recent immigrants to the United States become increasingly individualistic over time.

Implicit and Explicit Self-Concept

So far, we have placed the emphasis on research demonstrating that self-related processes do occur unconsciously or automatically. On several occasions, we pointed out that findings at the implicit level were convergent with observations based on self-report measures. In other cases, we stressed the fact that investigations of unconscious or automatic processes revealed a different picture from what we knew based on assessments of explicit self-concepts or identities. In this section, we examine more carefully how implicit and explicit self-related processes might be intertwined.

This issue has been addressed mostly in the domain of self-esteem or self-evaluation. For example, Bosson, Swann, and Pennebaker (2000) examined systematically the correlations between various measures of implicit and explicit self-esteem. Although some implicit measures correlated significantly with explicit measures, the magnitude of the observed correlations was relatively small (all $r < .27$). Using confirmatory factor analysis, Greenwald and Farnham (2000) demonstrated that implicit self-esteem and explicit self-esteem were distinct constructs (positively, but weakly, correlated). Research by Hetts and colleagues (1999) on the influence of culture on the self-concept also indicated that measures of implicit self-regard or group-regard were generally uncorrelated with explicit measures of self-concept. Spalding and Hardin (1999) found a correlation of $-.05$ between the Rosenberg Self-Esteem Scale (Rosenberg, 1965) and a priming technique designed to assess implicit self-esteem. In addition, they showed that implicit self-esteem predicted some behavioral consequences, whereas explicit self-esteem did not. In their study, participants took part in either a self-relevant or a self-irrelevant interview and were then rated by the interviewer on their anxiety. When the interview was self-relevant, participants low in implicit self-esteem appeared more anxious than participants high in implicit self-esteem. Explicit self-esteem did not predict participant's apparent anxiety. Interestingly, individuals' own ratings of anxiety were linked to explicit self-esteem but not to implicit self-esteem.

Several studies support the idea that, under some circumstances, self-descriptions may switch from a controlled mode to an automatic mode. One would expect automatic self-evaluations to be much more positive than controlled self-evaluations because people have a lifetime of practice describing themselves mostly in positive terms. Paulhus and Levitt (1987) demonstrated that a shift in self-description occurred when people were emotionally aroused. Participants were asked to indicate the self-descriptiveness of a set of traits. These traits were presented on a computer screen, along with emotional or nonemotional distractors. The presence of emotional distractors induced participants to claim more of the positive and fewer of the negative traits than they did in the presence of nonemotional distractors. A similar positivity effect was found when participants' attentional capacity was reduced (Paulhus, Graf, & Van Selst, 1989).

Kooi and colleagues (2001) demonstrated that encouraging participants to rely on their feelings clearly led to implicit self-esteem effects (positivity biases for name letters and birthdate numbers). In contrast, these forms of implicit self-esteem were no longer apparent when participants were encouraged to find reasons why they felt the way they did. Thus deliberative forms of processing can override automatic self-evaluations. In addition, Kooi and colleagues found that the ability to engage in conscious self-reflection affected the degree of congruence between implicit self-esteem and self-reported evaluations of the self. For example, slow self-evaluations were less congruent with implicit self-evaluation than fast self-evaluations. Similarly, when participants' cognitive resources were deprived (high cognitive load), implicit self-evaluations predicted self-reported evaluations, but that was not the case when cognitive resources were available (low cognitive load). These findings support the idea that when the capacity or the motivation to engage in conscious self-reflection is lacking, implicit self-evaluations will be the prevailing influ-
ence on phenomenological experience. In sum, the evidence at hand suggests that implicit and explicit self-concepts are distinct constructs, although, at least under some circumstances, connection may also be detected.

Conclusions

The question of how we know ourselves and what we know about ourselves is of fundamental interest to understanding how personal knowledge is represented, the degree to which such knowledge is constructed in social context, and its implications for health and well-being. Yet the epistemological quagmire inherent in the empirical assessment of knowledge about oneself has always posed a problem, as noted at the start of this chapter. We recommended that analyses of unconscious self-processes may assist in this regard, and we focused on the social aspect of self and identity, restricting our attention to a particular aspect of the self—one that emerges in the context of social group membership. From the initial research using implicit or indirect measures of self and identity, we already have evidence about the visible role of social group membership in creating a sense of self and self-worth.

The work reviewed in this chapter raises several possible issues that need to be incorporated into an understanding of self. Processes that capture group identity can operate without introspective access or deliberative thought. Such group identity and even knowledge about social groups (that is automatically learned even if consciously denied) can have indirect influence on assessments of the self. An unspoken assumption has been that implicit attitudes, beliefs, and motives about oneself are hard to change—given that they are overlearned associations about a well-known object. Several findings reported in this chapter would suggest, to the contrary, that implicit associations are not rigid and that shifts in self-definitions or self-evaluations can occur without conscious awareness or intention. Situational or contextual manipulations reveal the plasticity of self-related implicit social cognition. Finally, several lines of research reported in this chapter show the subtle but crucial ways in which sociocultural realities shape self-related mental processes. In many instances, the impact of sociostructural influences on psychological processes become more obvious when research is focused on the nitty-gritty of mental processes that are not consciously accessible but may nevertheless be found using indirect measures. In that regard, work on implicit processes offers the promise to renew thinking about the obvious interplay between the psychological and the social, the individual and the collective.

Notes

1. For a sample of such tasks, readers may visit www.yale.edu/implicit or www.tolerance.org.
2. A sample task is available at www.tolerance.org or may be obtained by writing to the authors.

Acknowledgments

Preparation of this chapter was supported by grants from the National Institute of Mental Health (No. MII-57672) and from the National Science Foundation (No. SBR-9709924) to Mahzarin R Banaji and by a Swiss National Science Foundation Fellowship (No. 8210-056562) to Thierry Devos.

References


8. Implicit Self and Identity

are watching me from the back of my mind. *Journal of Experimental Social Psychology*, 26, 435–454.
Dijksterhuis, A., & van Knippenberg, A. (1998). The rela-
tion between perception and behavior, or how to win a game of Trivial Pursuit. Journal of Personality and Social Psychology, 74, 865–877.


