

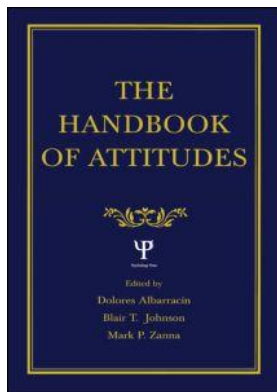
This article was downloaded by: 10.2.97.136

On: 06 Dec 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



## The Handbook of Attitudes

Dolores Albarracín, Blair T. Johnson, Mark P. Zanna

### Implicit and Explicit Attitudes: Research, Challenges, and Theory

Publication details

<https://test.routledgehandbooks.com/doi/10.4324/9781410612823.ch13>

John N. Bassili, Rick D. Brown

**Published online on: 01 Apr 2005**

**How to cite :-** John N. Bassili, Rick D. Brown. 01 Apr 2005, *Implicit and Explicit Attitudes: Research, Challenges, and Theory from: The Handbook of Attitudes* Routledge

Accessed on: 06 Dec 2023

<https://test.routledgehandbooks.com/doi/10.4324/9781410612823.ch13>

**PLEASE SCROLL DOWN FOR DOCUMENT**

Full terms and conditions of use: <https://test.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# 13

## Implicit and Explicit Attitudes: Research, Challenges, and Theory

John N. Bassili

*University of Toronto at Scarborough*

Rick D. Brown

*University of Florida*

### THEORETICAL BACKGROUND AND CHALLENGES

Assumptions about the mental representation of attitudes are critical to our understanding of attitudes. Research findings from the past two decades have added to the challenges that must be met by any representational system of attitudes. Two challenges stand out in particular. The first challenge arises from the increasingly apparent dual nature of attitudinal processes: deliberate or explicit on the one hand, and automatic or implicit on the other (Devine, 1989; Fazio, 1990; Wilson, Lindsey, & Schooler, 2000), whereas the second challenge arises from the pervasiveness of context effects in attitude experience and expression (see Blair, 2002; Schwarz & Bohner, 2001). In addition to being able to accommodate deliberative as well as relatively automatic attitude expression, the attitudinal constructs that are posited by a representational system of attitudes have to be able to account for attitudinal malleability. We begin with a discussion of the challenges created by the apparent duality of attitude expression and experience and by context effects. Given the relative novelty of research on implicit attitudes, we review this area in particular detail.

#### Two Significant Challenges to Contemporary Theories of Attitudes

##### *Implicit Versus Explicit Attitudes: The Challenge to the Unity of the Attitude Construct*

Our growing understanding of implicit psychological processes as powerful contributors to thought and behavior has formed what is undoubtedly one of the most important contributions of psychological research at the end of the 20th century. By their very nature, implicit processes exert a subtle influence on the way we think and behave, and this very subtlety has kept them for long in the shadow of experimental research. Though a multitude of historical observations

of unconscious psychological processes have been reported, it is only recently that they have been brought together under the umbrella of implicit psychological phenomena (see Schacter, 1987, for a review). Most of these observations were made in the domain of human memory, but it was not long before the notion of implicit processes became central to the discourse about attitudes.

The first clear application of implicit procedures to the study of attitudes was reported by Fazio and his colleagues (Fazio, Sanbonmatsu, Powell, & Kardes, 1986) who used a priming procedure to investigate the automatic activation of attitudes. This methodology was seminal because of its ability to reveal the automatic activation of evaluations toward an object, presented as a prime, by detecting the extent to which it facilitates or interferes with the subsequent judgment of a target word's valence. Two features of the priming paradigm are particularly noteworthy. First, because facilitation or interference of the target judgment is computed against a baseline condition in which no prime is presented, many extraneous features of the task are controlled (see Bassili, 2000). Second, by keeping the temporal parameters of the task very brief with a stimulus onset asynchrony of 350 ms, any effects of the prime on the target judgments cannot be influenced by controlled processes and can be assumed to be automatic.

At present, the most prominent implicit measure of attitudes is the implicit association test (IAT) introduced by Greenwald, McGhee, and Schwartz (1998) (see Krosnick, Judd, & Wittenbrink, chap. 2, this volume). The IAT procedure is based on categorization tasks involving two sets of stimuli. For example, one set may consist of White faces and of African American faces, and the other set of words with a positive connotation and others with a negative connotation. The participant's task across a series of blocked trials is to discriminate between the stimuli. On the first block the discrimination may be between White faces and African American faces and on the second between pleasant words and unpleasant words. The IAT procedure hinges on trials in which the same response has to be made for either of two stimuli—African American faces or pleasant words, for example. Response latency on such trials serves as an index of the participant's implicit preference of the two target categories. For example, some participants are much slower at the discrimination task when the same key has to be pressed for African American faces and pleasant words, than for White faces and pleasant words, revealing an implicit preference for Whites over African American.

Nosek and Banaji (2001) introduced the Go/No-go Association Task (GNAT), which is a variant of the IAT that involves a single target category. On one block of trials, participants respond to stimuli representing the target category and pleasant words and do nothing in response to other stimuli. On another block of trials participants respond to stimuli representing the target category and unpleasant words. As in the case of the IAT, the comparison of response latencies or error rates to such trials serves as an index of a participant's evaluation of the target category.

There are a variety of other types of implicit measures of attitudes that include word fragment completion tasks (Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997), name letter preference tasks (Koole, Dijksterhuis, & van Knippenberg, 2001), a modified evaluative priming procedure that uses a word pronunciation task in place of the word-valence task (Bargh, Chaiken, Raymond, & Hymes, 1996), and an evaluative variant of the spatial Simon task (De Houwer & Eelen, 1998; see Krosnick et al., chap. 2, this volume).

Importantly, implicit measures of attitudes have been shown to uniquely predict certain types of behavior—particularly spontaneous behavior. For example, implicit measures of prejudice have been shown to predict several spontaneous negative behaviors of Whites when interacting with African Americans (Dovidio et al., 1997; Dovidio, Kawakami, & Gaertner, 2002; McConnell & Liebold, 2001; Wilson et al., 2000). In a similar vein, Bessenoff and Sherman (2000) reported correspondence between a priming measure that used overweight and thin exemplars, and the distance at which participants later placed their own chair from an

overweight woman. Furthermore, the link between implicit attitudes and spontaneous behaviors has been demonstrated in domains other than prejudice and stereotypes. For example, Marsh, Johnson, and Scott-Sheldon (2001) showed a link between implicit attitudes toward condom use and behavior. This link, shown for casual partners but not main partners, could not be explained by explicit attitudes.

In short, what has been learned about implicit attitudes over the past two decades has presented attitude theorists with the challenge of reconciling two dramatically different modes of attitude expression. This challenge has not been trivial, as the relation between implicit and explicit attitudes has varied from strong to weak depending on many factors. A comprehensive account of both types of attitudes and of the relationship between them is, accordingly, of utmost importance at this juncture in attitude theorizing.

A note of caution is in order as we proceed with our discussion of what we will call implicit and explicit attitudes in this chapter. The use of the labels “explicit attitude” and “implicit attitude” is not meant to convey that two distinct types of attitudes are necessarily represented in memory. The framework we adopt here is distinctly more integrative than these labels suggest. The labels are used for efficiency to refer to two modes of attitudinal experience and expression, an explicit one that involves reflection on, and awareness of, the judgmental process, and an implicit one that does not.

### *Context Effects: The Challenge to Attitudes as Enduring Representations*

The tension between the properties of attitude stability and malleability has long played an important role in theorizing about attitudes (see Kruglanski & Stroebe, chap. 8, this volume). Traditional conceptions of attitudes assume that they are enduring psychological constructs that exercise a guiding function on thought and behavior. This perspective is evident in Allport’s (1935) classic definition of attitudes and prominent in theories of how attitudes should relate to behavior (Ajzen & Fishbein, chap. 5, this volume). Despite reasonable and intuitive expectations of stability in the attitude domain, this quality has proven elusive, especially in the mass public. In his classic research, Philip Converse (1964, 1974) tracked political opinions expressed by the respondents who were interviewed on a number of occasions at two-year intervals. The lack of stability of expressed opinions was dramatic, prompting Converse to develop a “black and white model” in which a small proportion of respondents give answers in a stable manner and the rest give answers at random. The black and white model fit the data unnervingly well! This pattern, along with other indications of vacuity in mass public attitudes prompted Converse to suggest that what is often measured in attitude surveys are “nonattitudes” that reflect little by means of stable sympathies.

Converse’s non-attitude thesis has been criticized on a number of grounds (for reviews see Kinder & Sears, 1985; Krosnick et al., chap. 2, this volume). Yet these skirmishes were but minor precursors to the battle about the stability of attitudes that would ensue. Schuman and Presser (1981) presented the results of an elaborate program of research on response effects, showing variations in answers caused not by substantive variations in the question but by apparently immaterial variations in things like the wording or order of questions. When asked whether the United States should forbid public speeches against democracy, for example, a larger percentage of respondents appear to be in favor of free speech than when asked if the United States should allow such speeches (Krosnick & Schuman, 1988; Rugg, 1941; for a review see Krosnick et al., chap. 2, this volume).

The apparent vacuity and lability of attitudes suggested by these findings was recently further emphasized by Norbert Schwarz and his colleagues (Schwarz & Strack, 1991; Schwarz & Bohner, 2001; for reviews, see Fabrigar et al., chap. 3, this volume; Kruglanski & Stroebe,

chap. 8, this volume; Wyer & Albarracín, chap. 7, this volume). Whereas past demonstrations of context effects were usually interpreted as revealing error that one must guard against in attitude measurement, Schwarz and colleagues' claim has been that there is little evidence that attitude expression ever rests on anything other than on-the-spot construction (Schwarz & Strack, 1991; Schwarz & Bohner, 2001).

Schwarz's case is premised on the observation that context effects impact information processing at every phase of the question answering. Question comprehension is impacted by inferences based on the nature of previous questions and other contextual considerations (Sudman, Bradburn, & Schwarz, 1996). Recall and judgment processes are also impacted by contextual influences such as information that is included versus excluded from a consideration set, a process that can variously result in assimilation or contrast effects (Stapel & Schwarz, 1998; Schwarz & Bless, 1992). Context effects are also evident at output, with respondents being sensitive to the necessity to format their answers in a way that is compatible with the response scale (Strack, 1994), as well as to social desirability expectations and self-presentational motives (DeMaio, 1984).

The growing realization that attitude expression and experience can manifest themselves implicitly as well as explicitly has raised the possibility that implicit attitudes represent a more accurate reflection of people's inner feelings than explicit attitudes, thus providing a window that is free of the effects of self-censorship. The prospect that these "true attitudes" may prove to be more enduring and less susceptible to context effects than explicit attitudes has, in fact, been the source of early optimism in meeting the challenge of attitude malleability raised by survey data. As we will see later in this chapter, the optimism went unsatisfied, implicit attitudes also proving susceptible to context effects, including effects of training and of various motives (see the section on the malleability of implicit attitudes below).

In short, the challenge presented by context effects for the conceptualization of attitudes is not insubstantial. Survey researchers have found this challenge particularly tenacious, encountering it repeatedly through the years and worrying about its impact on the validity of the attitudes they measure (Hyman & Sheatsley, 1950; Rugg, 1941; Schuman & Presser, 1981). The challenge has now come to its social psychological home to roost, and it must be met by any comprehensive theory of attitudes.

### Understanding Implicit/Explicit Attitudes and Context Effects

The building blocks of theories limit the scope of the phenomena the theories can explain. In this section we discuss the attitudinal primitives posited by popular theories of attitudes. By primitives, we mean theoretical elements that are helpful in explaining observed phenomena, and that do not themselves beg for explanation at approximately the same level as the phenomena they aim to explain. As we will see, the attitudinal primitives of most extant theories of attitudes have difficulty meeting the challenges posed by the variable relationship between implicit and explicit attitudes, and by context effects.

#### *Object-Evaluation Associations as Primitives*

Fazio (1986, 1995, 2001) offered what has arguably been the most influential view of the representation of attitudes for the past two decades. According to this view, attitudes consist of summary evaluations that are associated with the attitude object. These object-evaluation associations are presumed to be part of much broader associative networks, but the emphasis is on the object-evaluation association because its strength is assumed to be highly consequential for the way attitudes guide thought and behavior. The central tenet of the theory is that attitudes fall on a continuum defined at one end by representations of attitude objects that are not

associated with a summary evaluation, and at the other end by representations of attitude objects that are strongly associated with a summary evaluation. When the object-evaluation is strong, the attitude is accessible, so that exposure to the attitude object will activate the evaluation. The process is thought to be automatic and important to the attitude-behavior relationship because activated evaluations serve to guide thought and behavior in the presence of the attitude object (Fazio, Powell, & Herr, 1983; for a review, see Ajzen & Fishbein, chap. 5, this volume). This phenomenon has been demonstrated in experiments in which the repeated expression of an attitude makes it more accessible and more likely to guide behavior. Related research has shown that accessible attitudes are less pliable and more stable than inaccessible attitudes (Bassili, 1996; Fazio & Williams, 1986), that they guide attention and categorization (Smith, Fazio, & Cejka, 1996; Roskos-Ewoldsen & Fazio, 1992), and that they prime evaluations automatically upon exposure to the attitude object (Fazio et al., 1986; Fazio, Jackson, Dunton, & Williams, 1995).

From the perspective of associative network theories, Fazio's approach is minimalistic, keeping a singular focus on the link between the attitude object node and a corresponding evaluative node. Yet, this theory underlies a most significant development in research in the social cognition of attitudes, namely the demonstration that evaluations can be triggered automatically. Automatic evaluation activation is at the heart of implicit attitudes, and most of the present interest in implicit attitudes can be traced to Fazio's theory and research. Fazio's contribution is all that more important because it provides a framework for understanding implicit attitudes that is lacking from other accounts based on constructs such as beliefs or exemplars, that we discuss below.

Useful as object-evaluation associations have proven to be in the attitudinal domain, they are not good theoretical candidates for attitudinal primitives. Ironically, the problem is not that object-evaluation associations are too specific, because the summarizing of an entire history of experience with an attitude object in a single evaluation represents a major step in abstraction and generalization. Instead, object-evaluation associations fail as attitudinal primitives because they are too static. If attitudes consisted entirely of a summary evaluation, we should expect measures of attitudes to yield the same results across time and situations. Yet, as we have already suggested, an overwhelming amount of research over the past several decades demonstrates that attitudes are highly susceptible to context effects (see Schwarz & Bohner, 2001).

Because the object-evaluation association approach posits that attitudes fall on a continuum of strength, they should not, according to this framework, all be assumed to be stable and resistant to context effects. In fact, intuition is consistent with the object-evaluation association approach in suggesting that strong attitudes should be less susceptible to context effects than weak attitudes (Payne, 1951). One may think, for example, that feelings that are accessible, that are held with certainty, that are considered important to the self, and that are felt with intensity, ought to resist the influence of suggestions contained in an attitudinal query. Under most circumstances, however, this is not the case (Bassili & Krosnick, 2000; Krosnick & Schuman, 1988; for a review see Krosnick et al., chap. 2, this volume).

All in all, therefore, it appears that attitude malleability in the face of context shifts is the norm rather than the exception, and that this malleability is characteristic of strong and weak attitudes alike, as well as of deliberative or explicit attitudes, and of automatically activated or implicit attitudes. Any theory of attitudes, therefore, has to have the flexibility to accommodate malleability just as readily as it accommodates stability.

### *Beliefs as Primitives*

Combinatorial models of attitudes stipulate that attitudes are formed from the combination of beliefs (for reviews, see Kruglanski & Stroebe, chap. 8, this volume; Wyer & Albarracín,

chap. 7, this volume). The probabilistic model introduced by McGuire (1960) and extended by Wyer (1970), for example, posits that people hold beliefs that are related logically to a conclusion. If one thinks of the belief in the conclusion as the attitude, then beliefs in various premises relevant to that conclusion can be thought of as the elements from which the attitude is formed.

The most popular approach to the relationship between beliefs and attitudes consists of expectancy-value models that posit that attitudes are the sum of expected values of attributes associated with the object. At the heart of Fishbein and Ajzen's (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975; for reviews, see Ajzen & Fishbein, chap. 5, this volume; Wyer & Albarracín, chap. 7, this volume) influential theorizing about the relationship between attitudes and behavior is the notion that attitudes are determined by salient beliefs. The approach is thought to be "reasonable" because people form beliefs about an object by taking note of features, characteristics, outcomes and consequences associated with the object. These attributes are valued more or less positively by the person, and it is the values associated with the beliefs that ultimately yield evaluations of attitude objects.

Although people can hold a large number of beliefs about an attitude object, cognitive capacity limits the number that are salient at any given time (Doll & Ajzen, 1992; Miller, 1956). In this approach, therefore, attitudes are based on a sampling of salient beliefs about the attitude object and an integration of the values associated with these beliefs. In this fashion, we evaluate favorably objects that we believe at the moment to have largely desirable attributes or consequences, and we evaluate unfavorably objects that we believe to have largely undesirable attributes or consequences.

Beliefs are obvious candidates for attitudinal primitives, and in this capacity they offer an important advantage, but they also suffer from an important limitation. The advantage is that attitudes are composed of a sampling of evaluative elements that are salient at any moment in time. Context effects, therefore, are easily accounted for by variations in salience caused by temporary influences on the attributes of the attitude object that are brought to mind in a particular context. The limitation of belief sampling as a means of accounting for context effects is that beliefs themselves are susceptible to these effects. To illustrate, consider the belief assessed by the statement "having a baby within the next two years would make my marriage stronger" (Davidson & Jaccard, 1979). Given what we know about context effects on attitudes, it is very likely that this belief is abundantly susceptible to context effects. Is there any doubt, for example, that there would be a tendency for the answer to this question to be different if the question was preceded by a question about the stresses of child care than if it was preceded by a question about the joys of family outings? And if belief sampling is used to account for the influence of context on attitudes, how are we to account for the effect of context on beliefs?

The notion of sampling makes eminent sense for accounting for context effects. The problem is that the elements that are integrated into an attitude are defined at too high a level and share too many conceptual properties of attitudes to serve as primitives.

### *Exemplars as Primitives*

Exemplar theories stipulate that knowledge representation comprises specific instances of previously encountered exemplars. In its strict form, this approach does not postulate any abstraction from the exemplars, a judgment about a stimulus being made from the match between that stimulus and exemplars that it brings to mind. As applied to attitudes, the exemplar approach stipulates that as attitude objects are encountered, known attributes of similar exemplars that are brought to mind from memory influence judgments about the attitude object (Smith & Zárate, 1992).

Exemplar approaches have enjoyed substantial success in explaining a variety of judgment effects related to attitudes. For example, Lord and Lepper (1999) have proposed an exemplar approach to attitude representation that posits that people sample from exemplars when they answer attitude questions. When the same exemplars are sampled at two separate times or situations, the attitude judgment tends to be stable. This point is illustrated in a study by Sia and colleagues (Sia, Lord, Blessum, Ratcliff, & Lepper, 1997) in which participants were asked to think of social categories and to name the first exemplar that came to mind to represent the category. Participants engaged in the same task about a month later. The results showed that the attitudes of students who named the same exemplar on the two occasions were much more stable than the attitudes of participants who named different exemplars.

The effect of exemplars can generalize to attitude objects with which they are associated. The inclusion/exclusion model proposed by Schwarz and his colleagues provides a particularly interesting example of this phenomenon (Schwarz & Bless, 1992; Stapel & Schwarz, 1998). The model stipulates that information that is included in the mental representation of the attitude object at the time of judgment results in assimilation effects, whereas information that is excluded from the mental representation of the attitude object results in contrast effects. In one study, for example, participants were asked what party Colin Powell recently joined (Stapel & Schwarz, 1998). The question led participants to think of Colin Powell (a highly popular military leader at the time) as being part of the Republican Party, resulting in more positive evaluations of the party than when no question about Colin Powell was asked.

The exemplar approach to attitude research has produced dramatic demonstrations of how the exemplars that are brought to mind by the attitude object influence judgments about the object. The very fact that attitude judgments involve a sampling process among exemplars imparts substantial flexibility to this approach in accommodating context effects. As a theory of attitudes, however, the exemplar approach suffers from an important shortcoming. To explain an attitude in terms of the exemplar that it brings to mind begs the question: How is the attitude toward the exemplar to be explained? Whether or not the processes responsible for evaluation can influence our judgments through exemplar activation, these processes must also operate at earlier stages to impart value and affective qualities to stored exemplars.

### *Schemas as Primitives*

Greenwald and colleagues (Greenwald et al., 2002) recently provided a unified account of implicit social cognition focused on the relationships between the self and other social elements. According to this framework, information about the self is represented in schematic social knowledge structures where it is linked with constructs such as self-esteem, stereotypes, and attitudes. For example, the self concept includes a "ME" node that is linked to roles (e.g., graduate student) and traits (e.g., hard working). Self esteem is the collection of links between self concept nodes and corresponding positive and negative valence, represented as bipolar opposites. Stereotypes are collections of links with group concepts such as young person, graduate student, athlete, male and female, the last two also represented as bipolar opposites.

One interesting feature of this framework is that relations between constructs appear to conform to the principles of Heiderian balance (Heider, 1958) when measured using the IAT, but not when measured explicitly through self-report. Greenwald et al. (2002) attribute this difference to the fact that implicit measures such as the IAT are more sensitive to associations that the respondent is not aware of, and are also less susceptible to demand characteristics (Orne, 1962), evaluation apprehension (Rosenberg, 1969), and role playing (Weber & Cook, 1972), than are self-reports.

Issues of measurement aside, Greenwald et al.'s (2002) unified theory is particularly valuable because it provides a theoretical integration of some of social psychology's most important



cognitive (self-concept, stereotypes) and affective (attitude, self-esteem) constructs. In general, however, schemas do not constitute good theoretical primitives because their focus is more on structural relationships among represented constructs than on the building blocks from which the constructs are assembled. In the case of Greenwald et al.'s (2002) unified theory, for example, self-concept, stereotypes, attitudes, self-esteem, and so forth, are high level constructs that bear analysis at a lower level of primitives. In the case of attitudes, such analysis is precisely the focus of the present discussion. Thus, while schematic theories can be very valuable in providing an integrated view of knowledge representation, such theories are poor candidates for primitives.

### *Multiple Attitudes as Primitives*

One perspective on the relation between implicit and explicit attitudes is that old attitudes do not fade away, but linger in an implicit form (Wilson et al., 2000). According to the dual attitudes model, attitude change can result in a new evaluation that is explicit and consciously accessible, while the old evaluation lingers below awareness as an implicit attitude. When dual attitudes exist, the implicit attitude is more accessible and activated automatically, whereas the retrieval of an explicit attitude requires motivation and deliberate processing. Following this logic, the relative influence of implicit and explicit attitudes depends, therefore, upon the processing conditions under which a response is made. Explicit attitudes will influence deliberative responses that may occur if an individual has the opportunity to consider the consequences of his or her actions. Implicit attitudes will influence spontaneous or uncontrollable responses and responses that an individual does not view as an expression of inner feelings, and therefore does not attempt to control.

The dual attitudes model advanced by Wilson et al. (2000) provides a new perspective on attitude malleability by positing that context effects on attitude expression result from the elicitation of distinct attitudes in different circumstances. Pushed farther, the dual attitude model allows for the possibility that multiple implicit and explicit attitudes can coexist, and that a wide variety of attitudes can manifest themselves across contexts. Though theoretically possible, we believe that this perspective provides an unwieldy solution to the challenge posed by the susceptibility of attitudes to context effects. This lack of economy stems from the quasi-infinite potential diversity of contexts in which attitude objects can be encountered and far exceeds the capacity for memory to store autonomous representations of attitudes corresponding to each. There is simply too much diversity in attitude eliciting contexts for attitudinal primitives to be restricted to a set of autonomously represented attitudes, especially since these contexts are often instantiated linguistically and can thus be immensely varied.

Another problem we find with the dual attitude perspective is that it sets up an inherent opposition between implicit and explicit attitudes. Given that attitudes are relegated to the implicit realm by the formation of new explicit attitudes, one has to assume that implicit and explicit attitudes are usually different. In fact, Wilson et al. (2000) often invoke a process of "overriding" to account for the emergence of explicit attitudes in place of implicit ones. Yet, research on the relationship between implicit and explicit attitudes reveals that although the correlation between attitudes measured implicitly and explicitly is sometimes low, there are circumstances under which it is substantial (see section on response modes later in this chapter). It is unlikely, therefore, that the attitudinal primitives that underlie implicit attitudes are distinctly different from the attitudinal primitives that underlie explicit ones.

### *Connectionist Networks as Attitudinal Primitives*

In contrast to models that rely on symbolic constructs such as object-evaluation associations, beliefs, or exemplars to represent the elements responsible for aspects of psychological

functioning, connectionist models convey meaning at the level of patterns of activation distributed across units within a network (McClelland & Rumelhart, 1986). To this end, all stored knowledge is encoded in sets of connection weights that regulate the flow of activation among the individual units of a network. Retrieval of this knowledge involves reinstatement of a previously experienced pattern of activation, which can be evoked by a particular set of cues presented to the network as inputs. Within this framework, attitudinal primitives can be conceptualized as sets of units and connection weights that contribute to the emergence of distributed patterns of activation that may be associated with judgments, affect, and previous experiences with an attitude object.

Connectionist models provide explanations for many attitude relevant phenomena such as temporary and chronic accessibility, evaluative priming, and implicit processing of social information (see Smith, 1996 for a review; Masson, 1991; Smith, 1997; Wiles & Humphreys, 1993). Furthermore, by assuming differential involvement of areas of the brain that support feelings and cognition, connectionist systems can be designed to involve several types of elements that serve distinct functional roles. In this manner, affective elements that give rise to different experiential properties but that follow the same operational rules as purely evaluative ones can evoke patterns of activation associated with affective states and can interact with each other to produce the rich mix of human attitudinal and affective experience.

The flexibility and context sensitivity of connectionist networks is particularly relevant to the challenge posed by context effects in the attitude domain. Connectionist networks can be designed to satisfy parallel constraints that enable a network to settle into an overall activation pattern that best accommodates the current input in light of encoded representations of past experiences (Smolensky, 1989). As a consequence, it is unlikely that a given stored item will be evoked in precisely the same form as it was initially encoded. As a norm, encoded representations will be evoked or re-created in an imprecise form, as activation patterns are subject to influence from a person's accumulation of encoded knowledge (Smith, 1996). Furthermore, connectionist models are inherently context sensitive. Any other current sources of activation such as self-presentation concerns, expectancies, or affect, will also influence or constrain the resulting representation (Blair, Ma, & Lenton, 2001; Smith, 1996).

Connectionism can account for automatic processing in a variety of ways. For example, exposure to an attitude object that results in the automatic activation of evaluations and affect associated with the object can be modeled using pattern-completion properties of connectionist networks. Simply put, a partial pattern such as an input pattern representing the attitude object could evoke a much more elaborate pattern that incorporates several components of one's attitude toward the object (Moll, Miikkulainen, & Abbey, 1994). Automatic processing can also be modeled using a multimodule approach in which specific modules receive input patterns from sensory and perceptual modules and operate independent of, or prior to, conscious processing (Smith, 1997). In such a system, an input representing a perceived individual could result in the automatic categorization of this individual as a member of a particular stigmatized group and the automatic activation of attitudes toward this group independent of, or prior to, conscious processing of the individual.

Finally, several proponents of connectionism point out that these models are attractive because they provide a neural-like architecture for cognitive modeling (Bechtal & Abrahamsen, 1991). Generally speaking, connectionist networks are designed with a concern for neural plausibility. Although these models still involve simplifications and idealizations rather than actual matches of properties of biological neurons, their architecture is much more similar to the architecture of the brain than is the case for other extant symbolic models (Smith, 1996, 1997).<sup>1</sup>

The general properties of distributed connectionist networks make them good candidates for explaining implicit and explicit attitudes, which led us to organize our review around such a framework. For the purpose of our discussion, the inspiration afforded by connectionist

networks is more important than specific instantiations aimed at modeling attitudes and affect. Accordingly, we will be unrestricted in the presentation of our framework by the type of detail that would normally be part of a workable connectionist model. Our aim is to present a perspective on what have become challenging theoretical issues in the domain of attitudes, and to do so in a comprehensive manner. For this purpose, abstractions from connectionism are likely to be more useful to our understanding of implicit and explicit attitudes than a fully specified but more narrow connectionist model. For example, the molecularity of the elements that make up connectionist networks affords substantial flexibility in the production of implicit and explicit attitudes. Further, context effects can be understood as a recruitment of elements that are constrained by both their internal dynamics (interconnections) and by the potentiation of contextualized eliciting conditions.

### POTENTIATED RECRUITMENT: A FRAMEWORK FOR IMPLICIT AND EXPLICIT ATTITUDES

#### Overview of the Framework

We propose a framework in which attitudes are emergent properties of the activity of microconceptual networks that are potentiated by contextually situated objects, goals, and task demands. We chose the term “potentiation” over the more common term “activation” because “potentiation” places more emphasis on the source in the excitation process, an emphasis that is important in our framework. The evaluative phenomenology of attitudes in this framework derives from the potentiated recruitment of microconcepts that are imbued with evaluative information. Activity in microconceptual networks is highly fluid, its dynamic interplay being captured nicely by observations made by McIntosh (1999) on the basis of findings derived from functional neural imaging:

Most parts of the brain possess the rudimentary properties necessary for cognition. . . . When several brain regions interact at a larger scale, these rudimentary features will combine to produce a particular cognitive function (Bressler, 1995). Whether or not a region is part of a neurocognitive system depends on the specifics of the processing demands (what is the person doing?) and the interactions with other regions (what is the rest of the brain doing?). Just as an instrument in an orchestra may switch from a lead to a support role in different pieces of music, some regions may play a more prominent role in certain cognitive functions and then play a supporting role in others. (p. 540)

Likewise, we suggest that evaluation is an emergent property of activity in microconceptual networks that are recruited to a greater or lesser extent in different situations, and that imbue experience with distinct tonal quality.

Figure 13.1 depicts a framework consisting of a number of components, some representational, others stemming from conditions that impinge on the individual and others yet reflecting psychological processes. At the heart of the framework is the representational system that we term the *attitudinal cognitorium*. We borrow this label from Milton Rosenberg (1968), who coined it in the context of a discussion of cognitive consistency theories. The attitudinal cognitorium consists of microconcepts that are associated with each other in varying degrees and that have, at any given moment, a certain level of activation. Microconcepts are molecular elements of knowledge that yield meaning when assembled into networks with other microconcepts.

Several other elements of the framework identify sources of potentiation that influence the level of activation of microconcepts in the attitudinal cognitorium. There are four primary sources of such potentiation (Fig. 13.1: eliciting conditions). One source is recent information processing experiences that prime particular microconcepts in memory. Another source is current information about the attitude object. Eliciting conditions, which comprise such things

as encounters with the attitude object and queries about it, are particularly powerful sources of potentiation. Because such eliciting conditions are steeped in a context, properties of the context also exert powerful potentiating influences on the attitudinal cognitorium. The third source of potentiation consists of spreading activation between linked concepts. The cueing of one stored concept by another is well known in memory research (Anderson, 1983; Srull, 1983) and we suggest that it is responsible for the influence of general knowledge and of culture on attitudes and affect. Finally, cognitive activity in working memory (see Fig. 13.1) is an important source of potentiation, as well.

The combined effect of potentiation from preexisting activation, eliciting conditions, and prior experiences gives rise to patterns of microconcept activation from which attitudes and affect emerge. The interpretation (Fig. 13.1: Appraisal and construal) of objects and events in the social environment determines our evaluative and emotional reactions to them (Ellsworth, 1991; Scherer, Schorr, & Johnstone, 2001). We posit that microconcepts differ in their capacity to engender evaluations and feelings. The developmental history of microconcepts (whether they stem from learned evaluations or conditioned affect, for example) will produce projections onto parts of the brain that specialize in cognition (e.g., the neocortex) or emotion (e.g., the amygdala). Activation of microconcepts will engender corresponding experiences. We propose that meaning and feelings emerge from these patterns of activation in a configural manner. That is, a best fit is achieved among the informational properties of the microconcepts so as to produce coherent evaluations and feelings. Patterns of activation are highly fluid, microconcepts that are recruited into one pattern in one set of potentiating circumstances having the potential of being recruited, in combination with other microconcepts, in a different pattern in another set of circumstances. This potential for recombination is, in our opinion, essential to explain the malleability of attitudes and affect.

The response side of the potentiated recruitment framework is critical for understanding the relationship between explicit and implicit attitudes. By definition, explicit attitudes involve means of expression in which the respondent is aware of the assessment of the attitude. This type of attitude assessment is familiar and historically ubiquitous. From survey questions about matters of government policy, to friends' queries about how one liked a movie, to private commiserations about one's feelings or moods, the expression of explicit attitudes requires the involvement of controlled processes in working memory. Implicit means of attitude assessment focus the respondent's attention on features of the task that are irrelevant to those that are of interest to the researcher. The indirect nature of such assessment, therefore, minimizes the involvement of controlled processes relevant to the focal task (Krosnick et al., chap. 2, this volume). Speeded means of assessment also minimize the involvement of controlled processes in the attitudinal task. This relative uncontrollability is achieved either by using extremely quick intervals in priming procedures, or by getting the participant to make judgments in a hurry. The potentiated recruitment framework, therefore, posits that the main difference between explicit and implicit attitudes is the involvement of controlled processes at output (curved lines in Fig. 13.1).

### An Example

Evaluations are correlates of activity in networks of microconcepts. These microconcepts contain information, and this information carries evaluative and affective implications. Take the example of a woman who played competitive tennis as an adolescent. What evaluations and feelings will the sport engender in her? We propose that tennis will be represented in this woman's mind as a collection of microconcepts having to do with competition, discipline, pressure, fairness, the joys of winning, the disappointments of losing, traveling to tournaments, hanging out with other competitors, interacting with her parents, character building, etc. The list goes on, and is that much more elaborate because every one of these microconcepts is linked

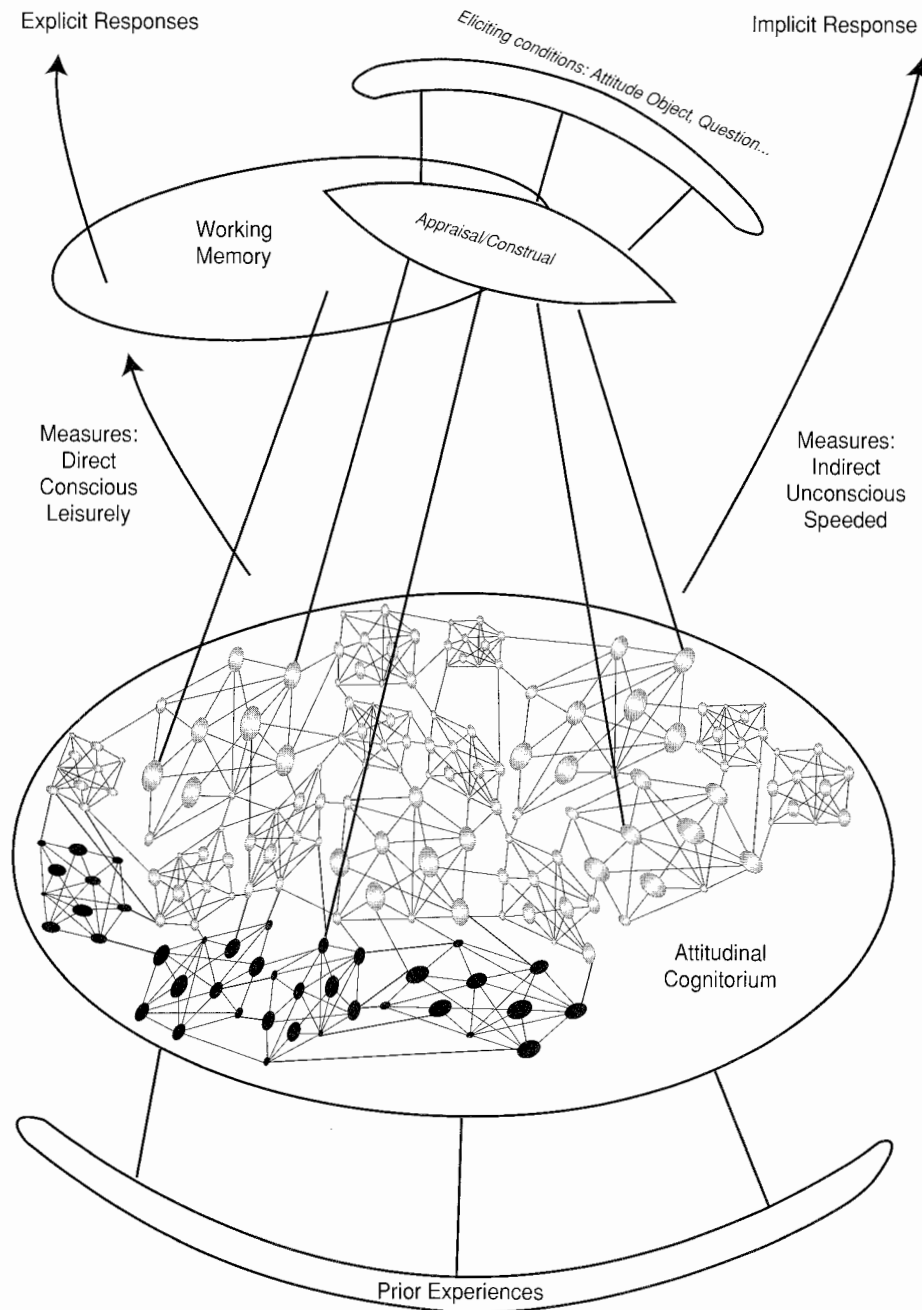


FIG 13.1. The Potentiated Recruitment Framework

Note: Potentiating factors are depicted as straight lines to, or within, the attitudinal cognitorium, whereas curved lines indicate output paths. Microconcepts in the cognitorium are depicted in two shades to indicate that some are pertinent to evaluation, whereas others are pertinent to feelings.

to information about specific instances where the concept was manifested. Competition may be linked to specific matches and tournaments, discipline to various training schedules over the years, fairness to disputes over line calls, the joys of winning to the outcome of matches, or interactions with parents, discussions and/or quarrels.

Where does this woman's attitude toward tennis come from? We propose that microconcepts are infused with evaluative and affective information and that attitudes are correlates of activation of microconcepts. Competing, winning, character building as well as many of the specific manifestations of these concepts may, for this woman, be positive and have the potential to engender positive feelings. Pressure, traveling to tournaments and quarreling with parents, along with their specific manifestations, are negative and have the potential to engender negative feelings. Evaluation at any given moment is the upshot of activity in the microconcepts that have been potentiated by eliciting, as well as prior processing conditions, and by cueing across interconnections among linked concepts (see Fig. 13.1).<sup>3</sup>

### Implicit and Explicit Attitudes

In the following sections, we review the literature on implicit and explicit attitudes. For the reasons outlined above, a potentiated recruitment framework appears ideal to explain these phenomena. Our review is, accordingly, organized around this framework.

#### *Features of the Context Potentiate Implicit and Explicit Attitudes*

Perceptual information about the attitude object constitutes one of the most powerful potentiating forces acting on the attitudinal cognitorium. As Higgins (1996) discussed, knowledge activation depends on the relation between stored knowledge units and stimulus information. "Applicability" plays a central role in the synergy between attended features of the stimulus object and those of stored information: The greater the overlap between them, the greater the potentiating force that the stimulus will exert on stored knowledge (Higgins, 1989, 1996). Likewise, the power of the attitude object to potentiate attitudes plays a crucial role in the potentiated recruitment framework. Because there are preexisting interconnections among microconcepts in the cognitorium, complex patterns of activation usually result from exposure to a subset of features of the attitude object, such patterns allowing for the influence of general knowledge.

Recent research has investigated the influence of specific features of attitude objects on the expression of implicit evaluations. For example, Livingston and Brewer (2002) demonstrated that specific physical features of faces used as primes significantly influenced automatic evaluations as measured by a race evaluative priming procedure. Mitchell, Nosek, and Banaji (2003) have also investigated the role that specific exemplars play in the expression of implicit attitudes. In a study of particular relevance, participants completed two race IATs. One IAT included racial stimuli consisting of admired African American individuals and disliked White individuals. Conversely, the other IAT included racial stimuli consisting of disliked African American individuals and admired White individuals. If IAT effects are dependent exclusively on evaluations associated with category labels (see De Houwer, 2001), the exemplars comprising each stimulus set should not affect the results. However, consistent with the findings of Livingston and Brewer (2002), implicit evaluations were dependent on the exemplars. As expected, a significant preference for Whites was demonstrated when the stimuli consisted of admired White individuals and disliked African American individuals. However, the preference for Whites was reduced when the stimuli consisted of disliked White individuals and admired African American individuals.

Another important aspect of potentiation by the attitude object is that the object is always encountered in a context, and features of the context are just as involved in the potentiation as

features of the attitude object. Context effects caused by the position of a question in the midst of others in a questionnaire are well known in survey research. A person's attitude toward free choice in abortion, for example, will appear less permissive if the issue is raised after questions in which abortion is contemplated to reduce a health risk for the fetus or the mother (Schuman & Presser, 1981). For the potentiated recruitment framework, such context effects are the norm rather than the exception, because context is part of the eliciting conditions for attitudes and affect.

Wittenbrink, Judd, and Park (2001, Study 2) provided compelling evidence that the context in which an attitude object is encountered can affect the potentiation of attitude-relevant microconcepts. In an otherwise typical race evaluative priming procedure, the images used as primes not only varied in the race of the depicted individual (White/African American), but also in social context. Independent of race, half the primes consisted of images of individuals in a positive context stereotypic of African Americans (a church) and half consisted of images of individuals in a negative context stereotypic of African Americans (a dilapidated street corner). The results indicated that facilitation scores were generally related to the context in which the target individuals were presented. Most notably, African American target individuals presented in the positive context resulted in stronger facilitation of positive words compared to negative words. This finding supports a basic tenet of the potentiated recruitment network, namely that variations in the context in which the attitude object appears influences the configuration of information that is activated in the cognitorium.

#### *Implicit Affect and Mere Exposure*

Zajonc (1980, 2000) has argued that evaluative tendencies often emerge prior to, and independently of, cognition. This assumption of independence between cognition and affect has its roots in the distinction between "reason" and "passion" (Aristotle, trans. 1991; Le Bon, 1995) and usually posits that people differ a lot more in the former propensity than the latter. As advanced by Zajonc, the notion of affective primacy is supported by phenomena in which evaluation of a stimulus is driven by processes, often subconscious, that do not appear to involve explicit cognitive processes.

The mere exposure effect (Zajonc, 1968) is one source of evidence that Zajonc (2000) has interpreted as being consistent with the notion of affective primacy. In this well-known phenomenon, repeated exposure to a stimulus increases liking for it, a finding that has proven ubiquitous and robust (Bornstein, 1989; Harrison, 1977; Smith & Bond, 1993; for a review, see Clore & Schnall, chap. 11, this volume). In some mere exposure experiments, the stimuli are presented below the threshold of awareness (Kunst-Wilson & Zajonc, 1980) and participants are unable to subsequently recognize the stimuli to which they were exposed. Yet, preferences based on mere exposure still emerge. Such findings have prompted Zajonc to argue not only that affect may stem from a separate system than cognition, but also that this system is faster acting and often takes primacy over cognition (see Damasio, 1994).

Another experimental paradigm that has yielded results consistent with Zajonc's interpretation involves subliminal primes (Murphy & Zajonc, 1993). In the relevant conditions of these experiments targets consisting of Chinese ideographs were preceded by an "affective prime" consisting of a smiling or frowning face. The participants' task on each trial was to indicate how much they liked the ideograph. When the prime was presented subliminally (for 4 ms) judgments shifted in the affective direction of the prime, participants expressing a preference for the ideograph when it was preceded by a smiling face than when it was preceded by a frowning face. When the prime was presented supraliminally (for 1 s), the effect was reversed. This pattern of results led Zajonc and his colleagues to conclude that emotional reactions can occur with minimal stimulation and can influence subsequent cognitions. An important aspect

of this logic is that the origin of affect that is elicited outside of conscious awareness is unspecified and can spill over onto the target, whereas the origin of affect that is elicited within the participant's awareness is specified and cannot easily be transferred to other stimuli.

The importance of subliminal stimulus presentation in mere exposure and priming studies for Zajonc's theory is that it presumably engages the affective system without simultaneously engaging the cognitive one. The potentiated recruitment framework does not share in this logic. A basic assumption of the potentiated recruitment framework is that most of cognition and affect is subconscious and that the potentiation of microconcepts relevant to both cognition and affect is also mostly subconscious. The finding that affective priming did not spill over onto the target in the supraliminal priming conditions of the Murphy and Zajonc (1993) studies is easily accommodated by the potentiated recruitment framework: Presumably, the engagement of working memory in these conditions (see Fig. 13.1) opened the door to untold numbers of processes with the potential to interact with the microconcepts potentiated by the subliminal prime (Martin, 1986; Strack, Schwartz, Bless, & Kuebler, 1993; Wegener & Petty, 1997). The overlaying of reflective working memory processes on top of subliminal microconcept potentiation can yield results that are very difficult to attribute to particular causes. For this reason, we do not feel that the dissociation in the effects of subliminal and supraliminal primes reported by Murphy and Zajonc (1993) provides strong evidence for either the primacy or the separateness of an affect system.

Another important element of Zajonc's distinction between cognition and affect is that cognition is always about something whereas unconscious affect is diffuse and can "become attached to any stimulus, even totally unrelated to its origin" (Zajonc, 2000, p. 54). Here again, we find the assumption regarding the delimitation of properties of affect to be overly restrictive. Cognition, like affect, is susceptible to diffuse associative effects. To illustrate, consider a study on spontaneous trait inference (Brown & Bassili, 2002). Under the guise of a study on the effects of distraction on information processing, participants were presented with trait implying information side by side with either faces or inanimate objects. The savings in relearning procedure used in this study revealed that traits were spontaneously associated with purported actors as well as with inanimate objects. The trait of superstition, for example, came to be associated with a banana! The fact that incongruous associations can develop between human characteristics such as superstition and inanimate objects such as bananas underscores the power of associative processes in the cognitorium, and raises doubts about the distinct status of affect or cognition in this regard.

The challenge of explaining the mere exposure and subliminal affective priming effects is easily met by the potentiated recruitment framework. The most common contemporary explanation of mere exposure is that repeated processing of a stimulus increases the fluency with which it is processed, an outcome that is hedonically positive and that is associated with the stimulus (Bornstein, 1989; Winkielman, Schwartz, Fazendeiro, & Reber, 2003). Fluency is a property that is very similar to accessibility, and is explained in terms of increased efficiency in the activation of assemblies of microconcepts in the potentiated recruitment framework. Subliminal affective priming can similarly be explained in terms of the recruitment by the subliminal prime of affectively laden microconcepts. In these two criterial phenomena closely related to implicit attitudes, therefore, processes that are native to potentiated recruitment replace the need for stipulation of distinct processes or systems.

### *Implicit Attitudes and Classical Conditioning*

Classical conditioning phenomena are also relevant to understanding implicit and explicit attitudes. The relevance of this quintessential learning mechanism for attitude formation has long been recognized (Razran, 1938; Staats & Staats, 1958). There has been an important



controversy, however, about whether evaluations and feelings are formed on the basis of temporal associations between the attitude object and affect-producing, unconditioned stimuli. Keisler, Collins, and Miller (1969), for example, have argued that participants' awareness of CS-UCS contingencies and of the purpose of attitude conditioning experiments may underlie all apparent attitude conditioning (see also Page, 1974; Fishbein & Middlestadt, 1995).

Concerns about the artifactual nature of the classical conditioning of attitudes, however, has recently been allayed by experiments that effectively eliminate this bias. One approach has relied on cover stories to reduce the likelihood that participants will notice the contingency between the US and CS. Tracking the extent to which participants are aware of the contingency is important in this approach. Baeyens and his colleagues (Baeyens, Eelen, & Van den Bergh, 1990) measured participant awareness of the CS-UCS contingency in a conditioning procedure that involved the association of neutral pictures with strongly liked or disliked ones. Their results revealed some level of contingency awareness, but this awareness was not related to attitude conditioning.

Other studies have addressed the possible impact of contingency awareness in attitude conditioning by relying on subliminal presentation of the US (De Houwer, Baeyens, & Eelen, 1994; De Houwer, Hendrickx, & Baeyens, 1997; Krosnick, Betz, Jussim, Lynn, & Stephens, 1992; Niedenthal, 1990). Krosnick et al. (1992) presented participants with a series of photographs of a person engaged in mundane activities such as getting into a car. For some participants, each photograph was preceded by the subliminal presentation of either positive-affect-arousing photos (a large Mickey Mouse doll) or by a negative-affect-arousing photo (e.g., a skull). The subliminal affect-arousing photos influenced attitudes toward the target person as well as beliefs about the target person's personality traits. Subliminal effects of this sort have proven somewhat inconsistent and open to alternative interpretations (Eagly & Chaiken, 1993; Olson & Fazio, 2001). A meta-analysis of five relevant experiments did, however, reveal a small reliable effect of subliminal conditioning on attitudes (De Houwer et al., 1997).

Recently, Olson and Fazio (2001) have developed a paradigm for implicit learning that uses supraliminal exposures of the US. The paradigm is based on the observation that rules can be learned in the absence of the ability to articulate them or conscious knowledge of information relevant to them (for a review see Seger, 1994). Participants are presented a series of several hundred images that can appear anywhere on the computer screen, either alone or in pairs. Their task is to hit a response key as quickly as possible when a specified target image appears. Many of the images are of Pokemon cartoon characters, two of which are of particular importance for the procedure because they serve as CS. On a number of trials, one of the Pokemon characters is paired with a positive US consisting of a word (e.g., awesome) or an image (e.g., puppies) and the other character is paired with a negative word (e.g., awful) or image (e.g., cockroach). Despite the use of supraliminal exposures, participants in these experiments do not become aware of covariation between CS-US pairs. Yet, conditioning does occur as indicated by more positive ratings in a surprise evaluation task of the Pokemon associated with the positive items than for the Pokemon associated with negative ones. This attitude conditioning effect is also manifested on an implicit measure (the IAT in Experiment 2).

Classical conditioning is undoubtedly a very important mechanism for the formation of associations in the attitudinal cognitorium. Research on fear conditioning (see LeDoux, 2000) provides a particularly compelling illustration of the power and tractability of attitude-relevant affective conditioning. In a typical animal study of fear conditioning, a rat receives a tone (the CS) followed by an electric shock (the US). Following a few tone-shock pairings (sometimes a single pairing is sufficient), defense responses such as freezing, changes in autonomic and hormonal activity, as well as changes in pain sensitivity and reflex expressions occur following the CS. This form of conditioning is widespread across species, having been observed as low in the phyla as in flies and worms, and as high as in monkeys and humans. Although fear

may be distinct among emotions in its urgent mobilization of means of protection for the organism, the lessons learned from this research, along with social psychological research on the conditioning of attitudes, have contributed important insights into processes of attitude formation.

### *Priming and Implicit Attitudes*

Higgins (1996) defined accessibility as the activation potential of available information. Activation potential has a stable component that stems from chronic accessibility of microconcepts (Higgins, King, & Mavin, 1982), and a variable component that stems from temporary influences on the activation of microconcepts (e.g., Higgins, Rholes, & Jones, 1977). It is the latter property of temporary influences on activation potential that explains the ubiquitous effect of priming on the likelihood that microconcepts will be used at a subsequent time.

The impact of priming on accessibility effects and knowledge potentiation has been the subject of substantial attention in the study of social cognition (for reviews see Higgins, 1996; Macrae & Bodenhausen, 2000). Early studies of person perception processes (Higgins et al., 1977; Srull & Wyer, 1979) usually focused on the effect of prior exposure to category information on the subsequent interpretation of ambiguous or vague behavioral information. The finding that prior use of category information renders the subsequent use of the category more likely had such an impact on social cognition research that it was once dubbed “the law of cognitive structure activation” (Sedikides & Skowronski, 1991, p. 169).

The effect of priming is usually assimilative because judgments about the target are swayed in the direction of the prime. Assimilation effects are usually attributed to heightened accessibility of primed knowledge, which raises the likelihood of its being recruited in the interpretation of the stimulus input. This ubiquitous phenomenon is fundamental to the potentiated recruitment framework. Contrast effects, in which the judgment of the stimulus is pushed away from the prime, however, have also been observed. These have been attributed to the activation of information that serves as a standard of comparison for the target (Higgins, 1989) and to motivated inhibitory processes engendered by participants’ awareness of their susceptibility to being biased by the prime (Martin, 1986).

The tension between activation and inhibition in knowledge representations continues to be the subject of intense interest and conjecture (Macrae & Bodenhausen, 2000). From the perspective of the potentiated recruitment framework, what is important is that accessibility of information in the attitudinal cognitorium is in a state of flux. Passing experiences have a temporary impact on the potentiation of microconcepts, and therefore, on their recruitability in evaluative processes. The possibility that inhibitory processes can occur in the cognitorium is theoretically significant because it can help account for the malleability of attitude expression.

### *Awareness in Implicit and Explicit Attitudes*

A controversial issue relevant to the conceptualization of implicit attitudes concerns one’s level of awareness of these evaluations. In their seminal review of implicit social cognition, Greenwald and Banaji (1995, p. 8) defined implicit attitudes as “introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects.” One implication of this definition is that people lack awareness of implicit evaluations. The definition, therefore, has ramifications for the conceptualization of attitudes and their measurement. Researchers who endorse dual-attitude models (Wilson et al., 2000) and those who do not (Fazio & Olson, 2003) agree that people may often have awareness, even if fleetingly, of automatically activated evaluations. Furthermore, Fazio and Olson (2003) have argued that current implicit measurement procedures do little to ensure that participants are unaware of these attitudes. For these reasons, Fazio and Olson

concluded that it is more appropriate to view the measure rather than the constructs as “implicit” or “explicit.” From this perspective, participants may be unaware that their attitudes are being measured while being aware that they possess these attitudes.

The focus on awareness in our framework is on whether activity in working memory adds new elements to the potentiation of microconcepts, and whether goal directed thought processes promote or block the recruitment of microconcepts underlying attitude expression (see Fig. 13.1). To the extent that conscious symbolic thought becomes involved in processing attitudinal information, there is a potential for attitude experience and expression to be influenced and modified by conscious thought. The issue, in our view, is not one of switching from one system or one process to another, but of the addition of sources of potentiation and control to the recruitment and expression of microconcepts.

### *Implicit and Explicit Attitudes as Response Modes*

There are many important factors that can influence explicit expression of attitudes. The most notable factors are related to motivational concerns such as self-presentation, consistency, and accuracy, and are primarily associated with deliberative processes. It is because implicit attitude expression is thought to preclude the involvement of deliberate processes that some theorists have been tempted to think of attitudes expressed through implicit means as “true” attitudes. According to the potentiated recruitment framework, however, this dichotomy is much too demarcated (see also Fazio & Olson, 2003). Our framework suggests that explicit and implicit attitudes share a vast base from which microconcepts conjoin into attitudes and feelings (see Fig. 13.1). Furthermore, we suggest that any factor that influences the potentiation of microconcepts, and therefore their recruitability in evaluative processes, can influence explicit and implicit expression of attitudes. By contrast, the notion of “true attitude” implies a singularity and stability that is not consistent with the pervasive context effects that influence implicit attitudes as they do explicit ones.

### *Malleability of Implicit Attitudes*

As we saw earlier, the question of whether implicit attitudes show the same kind of malleability in the face of shifting contexts as explicit attitudes is central to the conceptualization of the implicit/explicit attitude construct. The potentiated recruitment framework predicts variability in implicit attitude expression because several factors can have an impact on the potentiation of microconcepts. In addition to potentiation factors associated with chronic accessibility, implicit attitude expression can reflect the temporary influence of factors associated with recent thoughts or feelings and the construal of an attitude object in a particular goal state or social context. A review of the evidence of malleability in implicit attitude expression highlights the success researchers have had manipulating these types of potentiation factors.

*Training and Explicit Instructions to Suppress Stereotypes and Prejudice.* Kawakami and colleagues (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000) demonstrated that training in stereotype negation can significantly reduce the automatic activation of stereotypes upon category activation. The stereotype negation training exercise involved 6 blocks of 80 trials. In each trial, the image of a target individual was presented followed by the presentation of a trait word. Participants were instructed to press “NO” if the trait presented was associated with the social group to which the target individual belonged and “YES” if the trait was not associated with the social group. Stereotype negation training reduced automatic activation of stereotypic traits associated with skinheads and African Americans. The large number of repetitive trials in this research suggest that racial judgments may be proceduralized

skill-like reactions that run their course upon exposure to relevant stimulus conditions (Anderson, 1983; Smith, 1989).

In a similar vein, Karpinski and Hilton (2001) demonstrated that learning new associations with a target social group can interfere with implicit attitude expression. In particular, participants completed a bogus memory exercise attempting to learn 200 word pairings. In the experimental condition, the targets consisted of an equal number of pairings of the word "youth" with a negative word, and of the word "elderly" with a positive word. Examination of pre- and post-memory-task IAT scores revealed that participants in the experimental condition demonstrated a significant reduction in their preference for the "youth" category. In contrast, the IAT scores of participants who attempted to memorize targets consisting of an equal number of pairings of the word "youth" with a positive word and the word "elderly" with a negative word were unaffected.

In a related line of research, Rudman, Ashmore, and Gary (2001) demonstrated that the implicit expression of racial attitudes can be significantly reduced through extensive diversity training. In two studies, race (African American/White) IATs were conducted prior to and upon completion of a 14-week prejudice and conflict seminar taught by an African American instructor. In both studies, participants who had completed the diversity training exhibited significant improvement in their attitudes toward African Americans. This improvement was not observed for participants who were enrolled in an unrelated course (control group).

The preceding results come from studies in which participants are not explicitly instructed to suppress their stereotypes or prejudice. The malleability of attitudes and stereotypes in response to explicit suppression instructions has also been explored, albeit with less consistent findings. In a study by Lowery, Hardin, and Sinclair (2001, Study 3), participants who were instructed "to be as non-prejudiced as possible" demonstrated significantly less preference for Whites when compared with African Americans as indexed by error rates on a race IAT. By contrast, Blair, Ma, and Lenton (2001, Study 4) found no effect of explicit instructions to suppress the activation of stereotypes toward gender. Prior to completing a GNAT, participants in a suppression condition were told that the subsequent word detection task was a measure of gender stereotyping and they should attempt to suppress stereotypical associations between females and weakness. Yet, the results provided no evidence that these explicit instructions moderated stereotype expression.

Other studies have shown that explicit suppression instructions ultimately lead to hyper-accessibility of the target constructs. For example, Macrae, Bodenhausen, Milne, and Jetten (1994, Study 3) presented participants with the photograph of a skinhead and instructed them to describe a typical day in the life of this individual. Afterwards, participants completed an ostensibly unrelated lexical decision task that included several words associated with the skinhead stereotype. Although participants who had been instructed to suppress stereotypical thoughts during the original writing task created passages containing less stereotype content than controls, they responded more quickly to stereotype-related target words in the subsequent lexical decision task.

In a similar study, Galinsky and Moskowitz (2000, Study 1) presented participants an image of an elderly man sitting near a newspaper stand. Participants were initially instructed to write a narrative essay about a typical day in the life of this individual. Before constructing their essay, participants in a stereotype suppression condition were further instructed to try to avoid the influence of stereotypic preconceptions when thinking about the individual. Participants in a perspective-taking condition were further instructed to adopt the perspective of the individual by trying to see the world through his eyes. Stereotype accessibility was subsequently measured using a lexical decision task. Both stereotype suppressors and perspective-takers wrote less stereotypical essays of the elderly than did participants in a control condition that received no further instructions. However, participants in the stereotype suppression condition

demonstrated significantly shorter response times to stereotypic words compared to stereotype irrelevant words. This effect was not observed for participants in the perspective-taking or control conditions. In addition to supporting the notion of suppression rebound effects, therefore, this study also suggests that the expression of implicit attitudes toward stigmatized groups may be moderated by conscious perspective-taking strategies.

From the point of view of contemporary theories of thought suppression it is not surprising that strategies involving explicit instruction have met with limited success in moderating the automatic activation of evaluations and stereotypes. Theories of thought suppression suggest that conscious monitoring of unwanted thoughts can paradoxically result in repetitive priming and thus greater accessibility of the unwanted thoughts. It is only through vigilant conscious inhibitory mechanisms that expression of these thoughts is prevented (Macrae et al., 1994; Wegner & Erber, 1992; Wegner, 1994).

*Promotion of Counter Stereotypes.* Researchers have attempted to moderate the automatic activation of negative attitudes toward stigmatized groups by increasing the accessibility of counter stereotypes. Although most people have a mental representation of an admired "exception" or counter exemplar for a given stigmatized social group, this representation is usually not highly accessible. Manipulations that increase the focus of attention on positive exemplars of stigmatized groups have had significant effects on implicit measures of attitudes. For example, Blair, Ma, and Lenton (2001) demonstrated that mental imagery can be a successful strategy for moderating the automatic activation of attitudes. In a series of studies, participants who completed a counter stereotype mental imagery task demonstrated reduced expression of gender-related attitudes compared with participants who completed neutral, stereotypic, or no mental imagery tasks. Specifically, participants in the critical counter stereotype condition were instructed to imagine several aspects of a "strong woman" and then to describe their mental image in a short paragraph. This exercise had significant effects on the automatic activation of information associated with gender as measured by the IAT, the GNAT, and a false memory paradigm.

In a similar vein, Dasgupta and Greenwald (2001, Study 1) demonstrated that implicit expression of attitudes toward stigmatized groups such as African Americans and the elderly can be moderated by repeated exposure to admired group members. Prior to completing a race IAT, participants in the experimental condition were exposed to several admired African American individuals and disliked White individuals under the guise of a general knowledge test. These participants demonstrated significantly less preference for Whites than participants who had been exposed to disliked African Americans and admired Whites or to nonracial exemplars. This effect was observed immediately after the manipulation and 24 hours later. A similar pattern of reduced ageism appeared in a subsequent study using admired elderly individuals and disliked young individuals.

Wittenbrink et al. (2001, Study 1) demonstrated that the relative accessibility of counter stereotype representations can also be manipulated in more subtle ways. In one study, participants watched a short video clip and wrote an essay about the events it depicted. In the counterstereotype condition, the video showed several African American individuals enjoying a harmonious family barbecue. Participants in this condition demonstrated significantly less preference for Whites compared to participants who had viewed a video depicting several African American individuals engaged in negative stereotypic behavior.

*Ego-Defensive Motives.* Other lines of research have investigated the influence of ego-defensive motives on the construal of attitude objects. Spencer, Fein, Wolfe, Fong, and Dunn (1998) demonstrated that the motivation to restore one's self image can affect the automatic activation of attitude-relevant associations. Participants were given positive or negative feedback ostensibly based on their performance on a bogus IQ test. Afterwards, the accessibility

of stereotypic Asian (Study 1) or African American (Study 3) traits was measured using a word fragment completion task in the presence or absence of a stereotype-relevant cue. Participants whose self-esteem had been threatened by negative feedback demonstrated significant stereotype activation upon exposure to a stereotype-relevant cue. This finding is noteworthy because all participants were kept cognitively busy during the word fragment task and previous research suggested that the automatic activation of stereotypes does not occur when a person's cognitive resources are tasked (see Gilbert & Hixon, 1991).

Sinclair and Kunda (1999, 2000) suggest that the motivation to form a particular impression of an individual can influence the accessibility of the mental representations associated with the social categories to which a target individual belongs. For example, in one study (Sinclair & Kunda, 1999, Study 3) participants received positive or negative feedback, ostensibly from an African American or White medical doctor, about their performance on an interpersonal skills questionnaire. A subsequent lexical decision task was used to measure the accessibility of words associated with doctors and African Americans. The accessibility of words associated with the African American stereotype was greatest for participants who received negative feedback from the African American doctor and lowest for participants who received positive feedback from the African American doctor. In contrast, the accessibility of words associated with doctors was greatest for participants who received positive feedback from the African American doctor and lowest for participants who received negative feedback from the White doctor. In other words, participants who were motivated to decrease the credibility of an African American doctor who had criticized them, inhibited words associated with doctor and activated traits stereotypic of African Americans. In contrast, participants who were motivated to increase the credibility of African American doctor who had praised them activated words associated with doctor and inhibited traits stereotypic of African Americans.

In a related line of research, Richeson and Ambady (2001) demonstrated that the social role that a male anticipates playing with a female coworker affected his implicit expression of gender-related attitudes. Specifically, male participants who anticipated working in a subordinate position to a female on a future task demonstrated a preference for males over females as indexed by a gender IAT. Male participants who anticipated playing the role of partner or a superior, however, demonstrated a preference for females over males. Interestingly, female participants' implicit expression of gender-related attitudes was not affected by their anticipated role relative to a male coworker in a future task.

*Social-Regulatory Motives.* Research has also focused on the link between global social motives and implicit measures of attitudes. For example, Lowery, Hardin, and Sinclair (2001, Study 1) demonstrated that basic social-regulatory motives can moderate the expression of implicit attitudes toward stigmatized groups. Starting with the notion of "social tuning," Lowery et al. hypothesized that individuals would be highly motivated to reduce the expression of bias toward African Americans in the presence of an African American individual. Consistent with this prediction, White participants who interacted with an African American experimenter prior to completing a race (African American/White) IAT, revealed less preference for Whites compared to those participants who had interacted with a White experimenter. The moderating effect of social tuning was also found using a subliminal priming procedure (Lowery et al., 2001, Study 4).

#### *Social Desirability and the Relation Between Implicit and Explicit Attitudes*

According to the potentiated recruitment framework, microconcepts involved in the explicit and implicit expression of attitudes are recruited from a common base we have called the "attitudinal cognitorium" (see Fig. 13.1). In contrast to models of dual attitudes (Devine, 1989;

Karpinski & Hilton, 2001; Wilson et al., 2000), this framework does not assume unique mental representations for explicit and implicit attitudes. Instead, implicit and explicit attitudes differ in their output paths. In particular, working memory acts as a way station in the expression of explicit attitudes, whereas the expression of implicit attitudes goes through a relatively direct path. Simply put, deliberate processes active in working memory may contribute additional sources of potentiation and control to the recruitment and expression of microconcepts. This theoretical perspective is similar to that presented by Fazio as part of his motivation and opportunity as determinants of behavior model (MODE) (Fazio, 1990; Fazio & Towles-Schwen, 1999; Fazio & Olson, 2003) and by others who have distinguished between automatic and controlled aspects of attitudes (Devine, 1989).

Given the differential involvement of working memory in explicit and implicit attitude processes, the correspondence between measures of explicit and implicit attitudes should decrease as motivational influences are activated by the attitude object and social context. Researchers have long suggested that people are often more prejudiced than they would care to admit (Crosby, Bromley, & Saxe, 1980; Gaertner & Dovidio, 1986; Sigall & Page, 1971). Research examining the link between explicit and implicit measures of prejudice and stereotypes generally supports this view, as the correspondence between explicit and implicit measures of prejudice has usually been found to be low (for reviews see Blair, 2001; Brauer, Wasel, & Niedenthal, 2000; Dovidio, Kawakami, & Beach, 2001; Fazio & Olson, 2003).

The potentiated recruitment framework suggests that involvement of motivational factors operating in working memory may moderate the relationship between implicit and explicit measures of attitudes. If so, the correspondence between these measures should be weaker when the attitude object is of a nature that invokes motivational influences than when it is not. Fazio, Williams, and Sanbonmatsu (1990) offer evidence that is supportive of this view. Their research compared participants' explicit and implicit attitudes toward a variety of attitude objects ranging from the mundane (e.g., snakes, dentists) to more sensitive issues (e.g., pornography, African Americans). When the attitude object was not controversial, the correlation between implicit and explicit attitudes was high ( $r = .63$ ). When the attitude object was controversial, the correlation was weak ( $r = -.11$ ). Likewise, Nosek, Banaji, and Greenwald (2002) recently reported the results of thousands of IATs completed on an IAT internet site. As part of the online procedure, respondents completed both explicit and implicit measures of attitudes. High correlations were reported for attitude objects that are not particularly susceptible to social desirability pressures (preference for math vs. arts and for candidates in the 2000 federal election,  $r = .47$  and  $.52$ , respectively). By contrast, much smaller correlations were found for more sensitive attitudes such as those relating to race, age, gender, and self-esteem, with correlations ranging from  $.08$  to  $.24$ .<sup>4</sup>

Evidence that motivation plays a role in the overt expression of prejudice provides further support to the view that processes that operate in working memory may moderate the relationship between implicit and explicit attitudes. In one of their studies, Fazio et al. (1995, Study 4) included a measure of the motivation to control racial prejudice (see Dunton & Fazio, 1997 for details of an updated scale). The results showed that participants low in motivation to control prejudice exhibited explicit racial bias that was more consistent with their implicit attitudes than respondents high in motivation to control prejudice.

In a similar vein, Plant and Devine (1998) introduced the internal and external motivation to respond without prejudice scales (the IMS and EMS respectively). The IMS measures the extent to which one endorses, and has internalized, nonprejudiced beliefs. Higher scores on the IMS are associated with lower scores on typical explicit measures of prejudice. The EMS measures the extent to which individuals are motivated to conceal prejudiced beliefs in order to avoid negative evaluations resulting from their expression. High EMS scores were found to be associated with higher scores on traditional prejudice measures. Devine, Plant, Amodio,

Harmon-Jones, and Vance (2002) revealed that participants' levels of IMS and EMS moderated dissociation among implicit and explicit measures of prejudice. Most notably, participants with low IMS and high EMS scores exhibited highly prejudiced responses on both implicit and explicit measures. Participants with low EMS scores demonstrated consistency across the implicit and explicit measures, regardless of their IMS score.

*Measurement Issues.* As is always the case when examining relationships between different measures, one needs to be mindful of methodological factors that mitigate strong correlations. It has long been known, for example, that when different methods are used to measure a common construct, the observed correlations tend to be smaller than when a common methodology is used (Campbell & Fiske, 1959; Krosnick et al., chap. 2, this volume). Explicit and implicit measures tend to differ in specificity. Explicit measures typically target the attitude object directly (e.g., one's feelings toward an outgroup), whereas implicit measures are typically indirect and relative (e.g., response latencies from trials involving exemplars from an ingroup compared to response latencies from trials involving exemplars from an outgroup). Moreover, as we just saw, explicit measures are susceptible to systematic error resulting from motivational factors. Similarly, whereas implicit measures may not be particularly susceptible to motivational biases, they often rely on response latency and can suffer from measurement error associated with it (Cunningham, Preacher, & Banaji, 2001). For these reasons, the overall correspondence between explicit and implicit measures of attitudes may be reduced because they rely on different methodologies and response formats (see Blair, 2000; Brauer et al., 2000; Dovidio et al., 2001).

In response to such concerns, recent research by Cunningham and his colleagues (Cunningham et al., 2001) has elucidated several measurement issues associated with the use of implicit measures of attitudes. Participants completed three implicit measures and one explicit measure of racial prejudice. Each measure was taken four times at 2-week intervals. Latent variables in structural equation modeling were used to correct for measurement error. These analyses revealed that implicit measures of attitudes demonstrated a high degree of measurement error. As a consequence, Cunningham and colleagues advise caution when attempting to use implicit measures of attitudes as indices of individual differences in the absence of analytic techniques that can separate measurement error from individual differences. However, after correction for measurement error, the implicit measures of attitudes demonstrated consistency across time and measures. Of particular interest, the implicit measures formed a single latent construct that was robustly correlated with a latent construct formed by the explicit measure. These results, therefore, provide strong support for single construct models of attitudes such as the potentiated recruitment framework. Nonetheless, given the two-factor solution, the two types of attitude measures also capture unique sources of variance. This finding is consistent with aspects of the framework, such as working memory, that are posited to be involved differentially in explicit and implicit measures (see Fig. 13.1).

### General Discussion

The past decade has seen an explosion of interest in what is variously referred to as implicit or automatic attitudes. Implicit attitudes are important because of their potentially immediate impact on social functioning, and because cognitive methodologies can be used to identify evaluative reactions over which participants have little control. The first important step in this direction was taken by Fazio et al. (1986) who developed a priming methodology to reveal automatic evaluation activation upon exposure to the attitude object. As we have seen, other methodologies like the IAT have also proven successful in revealing preferences in contexts that elicit social desirability pressures (Dovidio et al., 1997; Dovidio et al., 2002;



Wilson et al., 2000). By contrast, explicit measures of attitudes often fail to correlate with behavioral or physiological indexes of preference (see Ajzen & Fishbein, chap. 5, this volume). A comprehensive theory of attitudes, therefore, must be able to accommodate characteristics of implicit and explicit attitude expression.

Given the important distinction between implicitly and explicitly measured attitudes, an understanding of the relationship between these constructs is essential. Research has revealed a high level of variability in the concordance between implicit and explicit attitudes (Fazio et al., 1990; Nosek et al., 2002) raising doubts about the independence of these constructs. The potentiated recruitment framework (see Fig. 13.1) accommodates these findings by positing that, with the exception of working memory involvement, the same processes are responsible for the emergence of implicit and explicit attitudes. Several motives can operate in working memory to influence or inhibit explicit attitudinal expression. Social desirability concerns are particularly important in some domains of attitudes, which is consistent with the potentiated recruitment framework's assumption that the relationship between explicit and implicit measures of attitudes drops as these concerns rise.

In addition to the focus on implicit attitudes as important psychological constructs, it has become increasingly apparent over the past several decades that attitudes are not necessarily enduring mental representations that are retrieved from memory at the time of judgment. As we saw earlier, Schwarz and his colleagues (Schwarz & Strack, 1991; Schwarz & Bohner, 2001) have presented a construal model of attitudes that rejects the notion that attitudes stem from stored memory representations. The strong form of this model not only suggests that attitudes are on-the-spot constructions, but by its very emphasis implies that these constructions are based on elements present in the external eliciting context rather than on elements that are represented in memory. According to this strong view, an attitudinal cognitorium plays only a minor role in judgment processes, working memory providing the forum for most attitudinal activity.

To the extent that the construal model does away with representational aspect of attitudes, we believe that it is wrong. The preponderance of evidence used to support the construal model comes from experiments that are more attuned to detecting change than to detecting stability. This is because the evidence is based on procedures in which a property of the context is manipulated to assess its effect on the reported attitude. The statistical focus is on whether the impact of the manipulation is significant, little attention being paid to the magnitude of the component of the response that remains stable across conditions. To conclude that an attitude that changes with shifting contexts is not based on a stable representation is akin to concluding that a flag that changes direction with shifting winds is not attached to a flagpole.

Schwarz's case against memory storage is focused entirely on the idea that attitudes are stored in a unitary fashion in memory, and that they can be retrieved intact regardless of the conditions that elicit them or the conditions that precede the elicitation. While it is true that some theorists like Fazio (1995) posit that attitudes consist of summary evaluations that are associated with the attitude object, there are other ways to conceive of the representation of attitudes that assume substantive representations without assuming that these representations are unitary or fixed. According to the potentiated recruitment framework, attitudes are represented as molecular elements that have the potential of being recruited in various mixes depending on the eliciting context and other potentiating factors. It is the fluidity of this process, rather than the lack of attitudinal representations, that accounts for malleability.

The emphasis in the potentiated recruitment framework is primarily on processes that are common to the emergence of all attitudes. However, the distinction between an attitudinal cognitorium in which the information is represented subsymbolically, and working memory in which symbolic processes predominate, conforms to a dual process view of attitudinal functioning. Dual process approaches are not new to research on attitudes (see Chaiken &

Trope, 1999 for a collection of such models, and Johnson, Maio, & Smith-McLallen, chap. 15, this volume, for a review). This notion is congenial to the potentiated recruitment framework, and has very wide application in psychology. In a recent article, for example, Kahneman (2003) attributed the effect of judgment heuristics to an intuitive system characterized by effortless and automatic associative processes that operate in parallel. The output of this system consists of accessible impressions that dominate judgment unless corrective steps are taken by rule governed reasoning processes that are slow, effortful, and serial in nature.

The interplay between the attitudinal cognitorium and working memory in the potentiated recruitment framework shares many of the elements of Fazio's (1990) MODE model. In both cases, automatic processes are responsible for the emergence of evaluations in consciousness, and effortful deliberative processes only occur when the motivation and cognitive capacity required by working memory are available. There is, however, a fundamental difference between the MODE model and the potentiated recruitment framework. Automatic processes in the MODE model involve the activation of attitude objects in which features are preassembled into coherent symbolic representations to which summary evaluations are attached. For reasons we presented early in this chapter, the potentiated recruitment framework dispenses with this view in favor of a representational system that is subsymbolic. That is, features of attitude objects (what we have called microconcepts) are assembled as they are recruited by potentiating influences that operate in a particular context. We contend that only such a subsymbolic system, where evaluations are attached to microconceptual features, can account for the fluidity of attitudes in changing contexts.

The MODE model accounts for attitude malleability by distinguishing between accessible attitudes that are activated automatically, and less accessible attitudes that are put together deliberately. The implication of this two-process approach is that deliberative processes are more susceptible to context effects than automatic ones. The finding that implicit attitudes are very susceptible to context effects poses a challenge to this explanation of attitude malleability. By contrast, the potentiated recruitment framework accommodates widespread malleability by positing that the sources of potentiation for an attitude are highly variable across contexts. Because implicit attitudes are as much the product of potentiating factors as explicit ones, there is no surprise in the fact that they are also as susceptible to context effects.

We began by noting that advancement in attitude research currently faces two particular challenges. The first challenge is in understanding the relationship between explicit and implicit attitudes. The correlation between explicitly and implicitly measured attitudes having varied across studies, it becomes important to reconcile these constructs in a coherent theoretical framework. The second challenge comes from the pervasive susceptibility of attitude experience and expression to context effects. Although it has been tempting to think of these context effects as applying only to weakly held attitudes (see Fabrigar, MacDonald, & Wegener, chap. 3, this volume), and to be driven by conscious influences on response processes, it has become apparent that context effects run deep and that they affect implicit as well as explicit attitude experience and expression. Given this malleability, it appears that attitudes emerge from a constructive process rather than from the retrieval of precomputed stored evaluations. Research on implicit and explicit attitudes has gone a long way in illuminating this process.

#### ACKNOWLEDGMENTS

The writing of this chapter was supported by Social Sciences and Humanities Research Council of Canada grant 410-2001-0554.

## ENDNOTES

<sup>1</sup>Connectionism is not new to social psychology. In the title of his influential paper on the approach, Smith (1996) raises the question "What do connectionism and social psychology offer each other?" The outlook in the answer he provides is optimistic, with phenomena like accessibility, social interaction, and affect and motivation, standing to gain from the synergy between the two approaches. In a similar paper that explores the import of connectionism for social psychology, Read, Vanman, and Miller (1997) argued that connectionist modeling "has great importance for understanding issues of both historical and current concern for social psychologists" (p. 26). They went on to point out similarities between connectionist models and fundamental Gestalt principles as manifested in impression formation, cognitive consistency and goal-directed behavior. It is not without fanfare, therefore, that the connectionist revolution that has unfolded in cognitive psychology has been greeted in some social psychological circles.

Despite the attention given to the promise of connectionism in social psychology, a relatively small number of models that focus specifically on social psychological phenomena have been produced. A few examples of these models will help situate elements of the approach that hold substantial promise for theorizing about attitudes. First, we need to highlight an important distinction between types of connectionist models. The distinction is between "localist" or "symbolic" models, and "distributed" or "subsymbolic" models. In localist models, each processing unit, or node, represents an entire concept, whereas in distributed models nodes represent features, or subsymbolic microconcepts, with higher order concepts emerging from activity in networks of nodes. Although this distinction may appear minor and inconsequential in the connectionist bag of tools, it is crucial to the challenge posed by context effects to attitude theories, subsymbolic models offering a lot more flexibility for dealing with context effects than localist models.

Several localist connectionist approaches have been formulated to model social psychological phenomena. In one such model, Read and Marcus-Newhall (1993) focus on the evaluation of explanations by postulating connection weights between explanations such as "Cheryl is pregnant" and the fact that "she has an upset stomach," "has been gaining weight" and "is feeling tired." The constraints imposed by these connection weights in the spreading of activation among the nodes eventually lead the network to settle in a state where the broadest, most parsimonious, unique, and best developed explanation has a higher state of activation than competing explanations. In a similar parallel constraint satisfaction model, Kunda and Thagard (1996) account for the joint effect of stereotype and individuating information in impression formation by postulating that upon observing, for example, that an African American (or White) person pushed someone, activation spreads between conceptual representations of these observations and other concepts such as "aggressive," "violent push," and "jovial shove." Connection weights between these concepts cause it to settle with activation levels that conform to empirical findings regarding how stereotypes, behaviors, and traits affect each other's meanings, and how multiple stereotypes have an impact on impressions.

The preceding examples illustrate localist models where nodes contain entire concepts. Localist models that are more directly relevant to attitudes are also possible. For example, Read and Miller (1994) present a parallel constraint satisfaction model of the Festinger and Carlsmith (1959) cognitive dissonance experiment where the nodes contain concepts such as "I do not lie without a good reason," "I was paid \$20," "the task was boring," "the task was interesting," etc. Beliefs, of course, are excellent candidates for concepts in localist connectionist networks. We have already seen, however, that beliefs are themselves susceptible to context effects. The assessment that a task is boring, for example, depends on what task preceded it, alternative available tasks, the thoughts one had while engaging in the task, etc. This is why we do not feel that localist networks built on entire symbolic concepts are ideally suited to meeting the challenge posed by the highly fluid nature of attitudinal experience.

By contrast to localist networks in which each unit represents an entire concept, subsymbolic networks posit that semantically meaningful representations consist of patterns of activation across a number of processing units (McClelland & Rumelhart, 1986). In a recent model of stereotyping processes, for example, each stimulus consisted of 40 activation values that were applied to 40 units in the network (Queller & Smith, 2002). Learning rules controlled changes in weights over a large number of training inputs. The stereotypicality of the output of the network was tested by comparing the activation values on the units with missing inputs with those on stereotypic activation values. In another recent subsymbolic model, Eiser, Fazio, Stafford, and Prescott (2003) trained a network consisting of three layers of processing units to discriminate "good" from "bad" inputs. Outputs from this network were characterized by "landscapes" of activation patterns, successful learning being revealed when the topology of these landscapes mimicked that of "good" and "bad" inputs.

Subsymbolic distributed networks are, by necessity, abstract. That is, mental representations and processes in these networks do not map in a simple linear manner onto people's responses or onto the symbolic constructs that theorists often find useful in accounting for thought, feelings, and behavior. This lack of correspondence may have been responsible for the slow adoption of such models in social psychology, and to Carlston and Smith's (1996) observation that "In coming years, social psychologists' theoretical thinking will no doubt mature beyond the simple tendency to assume a one-to-one correspondence between function (e.g., semantic priming) and structure (an associative link between two concepts)" (p. 198).

Matters of theoretical maturity aside, we believe that conceiving of attitudes as emergent properties of activity in networks of subsymbolic elements holds substantial promise in meeting the two challenges that we outlined at the

outset, namely the challenge of context effects in attitude responses and experience, and the challenge of understanding the relationship between implicit and explicit attitudes.

<sup>2</sup>As we saw in our earlier discussion of attitudinal primitives, distributed knowledge representations offer a number of advantages over localized ones. One property that is of particular importance in the domain of attitudes is flexibility in accounting for context effects.

<sup>3</sup>The microconcepts used in the present example are, by necessity, described at a fairly abstract level. We suggest, however, that in the same way that a global attitude toward tennis is made up of microconcepts like the ones described here, so too are these microconcepts made up of finer elements. Quarreling, for example, may be made up of microconcepts such as disagreement, hostility, dominance, frustrations, etc. For communicative purposes, it is more fruitful to focus on microconcepts at a level of analysis that matches the evaluations under consideration than to aim for the finest level of analysis possible.

<sup>4</sup>Some studies, however, have failed to find correspondence between explicit and implicit attitudes toward mundane objects (Karpinski & Hilton, 2001).

## REFERENCES

- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Allport, G. W. (1935). Attitudes. In C. Murchinson (Ed.), *A handbook of social psychology* (pp. 798–844). Worcester, MA: Clark University Press.
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning & Verbal Behavior*, 22, 261–295.
- Aristotle. (1991). *The art of rhetoric*. H. C. Lawson-Tancred, Trans. London: Penguin.
- Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective-evaluative learning. *Cognition & Emotion. Special Issue: Evaluative Conditioning*, 4, 3–18.
- Bargh, J., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology*, 32, 104–128.
- Bassili, J. N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. *Journal of Personality & Social Psychology*, 71, 637–653.
- Bassili, J. N. (2000). Cognitive Indices of Social Information Processing. In A. Tesser & Norbert Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 68–88). Malden, MA: Blackwell.
- Bassili, J. N., & Krosnick, J. A. (2000). Do strength-related attitude properties determine susceptibility to response effects? New evidence from response latency, attitude extremity, and aggregate indices. *Political Psychology. Special Issue: Response latency measurement in telephone surveys*, 21, 107–132.
- Bechtal, W., & Abrahamson, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Cambridge, MA: Basil Blackwell.
- Bessenoff, G. R., & Sherman, J. W. (2000). Automatic and controlled components of prejudice toward fat people: Evaluation versus stereotype activation. *Social Cognition*, 18, 329–353.
- Blair, I. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed.), *Cognitive social psychology: The Princeton symposium on the legacy and future of social cognition* (pp. 359–374). Mahwah, NJ: Lawrence Erlbaum Associates.
- Blair, I. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, 6, 242–261.
- Blair, I., Ma, J., & Lenton, A. (2001). Imagining stereotypes away: the moderation of automatic stereotypes through mental imagery. *Journal of Personality and Social Psychology*, 81, 828–841.
- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106, 265–289.
- Brauer, M., Wasel, W., & Niedenthal, P. (2000). Implicit and explicit components of prejudice. *Review of General Psychology*, 4, 79–101.
- Bressler, S. L. (1995). Large-scale cortical networks and cognition. *Brain Research Reviews*, 20, 288–304.
- Brown, R. D., & Bassili, J. N. (2002). Spontaneous trait associations and the case of the superstitious banana. *Journal of Experimental Social Psychology*, 38, 87–92.
- Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Carlston, D. E., & Smith, E. R. (1996). Principles of mental representation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 184–210). New York: Guilford.
- Chaiken, S., & Trope, Y. (1999). *Dual-Process Theories in Social Psychology*. New York: Guilford.

- Converse, P. E. (1964). The nature of beliefs systems in mass publics. In D. Apter (Ed.), *Ideology and discontent* (pp. 206–261). New York: The Free Press.
- Converse, P. E. (1974). Comment: The status of nonattitudes. *American Political Science Review*, *68*, 650–660.
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of Black and White discrimination and prejudice: A literature review. *Psychological Bulletin*, *87*, 546–563.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170.
- Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York: Putnam.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–814.
- Davidson, A. R., & Jaccard, J. J. (1979). Variables that moderate the attitude-behavior relation: Results of a longitudinal survey. *Journal of Personality and Social Psychology*, *37*, 1364–1376.
- De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*, 443–451.
- De Houwer, J., Baeyens, F., & Eelen, P. (1994). Verbal evaluative conditioning with undetected US presentations. *Behaviour Research and Therapy*, *32*, 629–633.
- De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion*, *12*, 45–61.
- De Houwer, J., Hendrickx, H., & Baeyens, F. (1997). Evaluative learning with “subliminally” presented stimuli. *Consciousness and Cognition*, *6*, 87–107.
- DeMaio, T. J. (1984). Social desirability and survey measurement: A review. In C. F. Turner & E. Martin (Eds.), *Surveying Subjective Phenomena* (Vol. 2, pp. 257–281). New York: Russell Sage Foundation.
- Devine, P. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.
- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*, 835–848.
- Doll, J., & Ajzen, I. (1992). Accessibility and stability of predictors in the theory of planned behavior. *Journal of Personality and Social Psychology*, *63*, 754–765.
- Dovidio, J. F., Kawakami, K., & Beach, K. R. (2001). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. In R. Brown & S. L. Gaertner (Eds.), *Blackwell handbook of social psychology: Intergroup processes*. Malden, MA: Blackwell.
- Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interactions. *Journal of Personality and Social Psychology*, *82*, 62–68.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*, 510–540.
- Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316–326.
- Eagly, A. H., & Chaiken, S. (1993). *The Psychology of Attitudes*. New York: Harcourt Brace Jovanovich.
- Eiser, J. R., Fazio, R. H., Stafford, T., & Prescott, T. J. (2003). Connectionist simulation of attitude learning: Asymmetries in the acquisition of positive and negative evaluations. *Personality and Social Psychology Bulletin*, *10*, 1221–1235.
- Ellsworth, P. C. (1991). Some implications of cognitive appraisal theories of emotion. In K. Strongman (Ed.), *International review of studies on emotion* (pp. 143–161). New York: Wiley.
- Fazio, R. H. (1986). How do attitudes guide behavior? In R. M. Sorrentino & T. E. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 204–243). New York: Guilford.
- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrated framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). New York: Academic Press.
- Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences*. Ohio State University series on attitudes and persuasion (Vol. 4, pp. 247–282). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion. Special Issue: Automatic Affective Processing*, *15*, 115–141.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297–327.

- Fazio, R. H., Powell, M. C., & Herr, P. M. (1983). Toward a process model of the attitude-behavior relation: Accessing one's attitude upon mere observation of the attitude object. *Journal of Personality and Social Psychology*, *44*, 723-735.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229-238.
- Fazio, R. H., & Towles-Schwen, T. (1999). The MODE model of attitude-behavior processes. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 97-116). New York: Guilford.
- Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations: An investigation of the 1984 presidential election. *Journal of Personality and Social Psychology*, *51*, 505-514.
- Fazio, R. H., Williams, C. J., & Sanbonmatsu, D. M. (1990). *Toward an unobtrusive measure of attitude*. Unpublished manuscript, Indiana University, Bloomington, IN.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203-210.
- Fishbein, M., & Ajzen, I. (1975). *Beliefs, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Fishbein, M., & Middlestadt, S. (1995). Noncognitive effects on attitude formation and change: Fact or artifact? *Journal of Consumer Psychology*, *4*, 1469-1480.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61-89). San Diego, CA: Academic Press.
- Galinsky, A. D., & Moskowitz, G. D. (2000). Perspective-taking: Decreasing stereotype accessibility and in-group favoritism. *Journal of Personality and Social Psychology*, *78*, 708-724.
- Gilbert, D., & Hixon, G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*, 509-517.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4-27.
- Greenwald, A. G., Banaji, R. M., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellot, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem and self-concept. *Psychological Review*, *109*, 3-25.
- Greenwald, A. G., McGhee, D. E., & Schwartz, L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464-1480.
- Harrison, A. A. (1977). Mere exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 39-83). San Diego, CA: Academic Press.
- Heider, F. (1958). *The psychology of interpersonal relations*. Oxford, UK: Wiley.
- Higgins, E. T. (1989). Knowledge accessibility and activation: Subjectivity and suffering from unconscious sources. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought: The limits of awareness, intention and control* (pp. 75-123). New York: Guilford.
- Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133-168). New York: Guilford.
- Higgins, E. T., King, G. A., & Mavin, G. H. (1982). Individual construct accessibility and subjective impressions and recall. *Journal of Personality and Social Psychology*, *43*, 35-47.
- Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141-154.
- Hyman, H. H., & Sheatsley, P. B. (1950). The current status of American public opinion. In J. C. Payne (Ed.), *The teaching of contemporary affairs* (pp. 11-34). New York: National Education Association.
- Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist*, *58*, 697-720.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology*, *81*, 774-788.
- Kawakami, K., Dovidio, J. F., Moll, J., Hermsen, S., & Russin, A. (2000). Just say no (to stereotyping): Effects of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology*, *78*, 871-888.
- Keisler, C., Collins, B. E., & Miller, N. (1969). *Attitude change*. New York: Wiley.
- Kinder, D. R., & Sears, D. O. (1985). Public opinion and political action. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., Vol. 2, pp. 659-741). New York: Random House.
- Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, *80*, 669-685.
- Krosnick, J. A., Betz, A. L., Jussim, L. J., Lynn, A. R., & Stephens, L. (1992). Subliminal conditioning of attitudes. *Personality and Social Psychology Bulletin*, *18*, 152-162.
- Krosnick, J. A., & Schuman, H. (1988). Attitude intensity, importance, and certainty and susceptibility to response effects. *Journal of Personality and Social Psychology*, *54*, 940-952.

- Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review*, *103*, 284–308.
- Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognized. *Science*, *207*, 557–558.
- Le Bon, G. (1995). *The crowd*. Somerset, NJ: Transaction Publishers.
- LeDoux, J. E. (2000). Emotion circuits in the brain. *Annual Review of Neuroscience*, *23*, 155–184.
- Livingston, R. W., & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology*, *82*, 5–18.
- Lord, C. G., & Lepper, M. R. (1999). Attitude Representation Theory. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 265–343). San Diego, CA: Academic Press.
- Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855.
- Lowery, B. S., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*, 842–855.
- Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual Review of Psychology*, *51*, 93–120.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology*, *67*, 808–817.
- Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. J. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift für Experimentelle Psychologie*, *48*, 161–175.
- Martin, L. L. (1986). Set/reset: Use and disuse of concepts in impression formation. *Journal of Personality and Social Psychology*, *51*, 493–504.
- Masson, M. E. J. (1991). A distributed memory model of context effects in word identification. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 233–263). Hillsdale, NJ: Lawrence Erlbaum Associates.
- McClelland, J. L., & Rumelhart, D. E. (1986). A distributed model of human learning and memory. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: explorations in the microstructure of cognition* (Vol. 2, pp. 170–215). Cambridge, MA: MIT Press.
- McConnell, A. R., & Liebold, J. M. (2001). Relations between the implicit association test, explicit racial attitudes, and discriminatory behavior. *Journal of Experimental Social Psychology*, *37*, 435–42.
- McGuire, W. J. (1960). A syllogistic analysis of cognitive relationships. In C. I. Hovland & M. J. Rosenberg (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 65–111). New Haven, CT: Yale University Press.
- McIntosh, A. R. (1999). Mapping cognition to the brain through neural interactions. *Memory*, *7*, 523–548.
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.
- Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*, 455–469.
- Moll, M., Miikkulainen, R., & Abbey, J. (1994). The capacity of convergence-zone episodic memory. In *Proceedings of the twelfth national conference on artificial intelligence* (Vol. 1, pp. 68–73). Menlo Park, CA: American Association for Artificial Intelligence.
- Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology*, *64*, 723–739.
- Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology*, *26*, 505–527.
- Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625–666.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, *6*, 101–115.
- Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413–417.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*, 776–783.
- Page, M. M. (1974). Demand characteristics and the classical conditioning of attitudes experiment. *Journal of Personality and Social Psychology*, *30*, 468–476.
- Payne, S. L. (1951). *The art of asking questions*. Princeton, NJ: Princeton University Press.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–832.
- Queller, S., & Smith, E. R. (2002). Subtyping versus bookkeeping in stereotype learning and change: Connectionist simulations and empirical findings. *Journal of Personality and Social Psychology*, *82*, 300–313.

- Razran, G. H. S. (1938). Conditioning away social bias by the luncheon technique. *Psychological Bulletin*, 37, 481.
- Read, S. J., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology*, 65, 429–447.
- Read, S. J., & Miller, L. C. (1994). Dissonance and balance in belief systems: The promise of parallel constraint satisfaction processes and connectionist modeling approaches. In R. C. Schank & E. Langer (Eds.), *Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson* (pp. 209–235). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Read, S. J., Vanman, E. J., & Miller, L. C. (1997). Connectionism, parallel constraint satisfaction processes, and Gestalt principles: (Re)introducing cognitive dynamics to social psychology. *Personality and Social Psychology Review*, 1, 26–53.
- Richeson, J. A., & Ambady, N. (2001). Who's in charge? Effects of situational roles on automatic gender bias. *Sex Roles*, 44, 493–512.
- Roskos-Ewoldsen, D. R., & Fazio, R. H. (1992). On the orienting value of attitudes: Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality and Social Psychology*, 63, 198–211.
- Rosenberg, M. J. (1968). Hedonism, inauthenticity, and other goads toward expansion of a consistency theory. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 73–111). Chicago: Rand McNally.
- Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifacts in behavioral research* (pp. 279–349). New York: Academic Press.
- Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology*, 81, 856–868.
- Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly*, 5, 91–92.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 501–518.
- Scherer, K. R., Schorr, A., & Johnstone, T. (2001). Appraisal processes in emotion: Theory, methods, research. *Series in affective science*. London: Oxford University Press.
- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. New York: Academic Press.
- Schwarz, N. (1996). Survey research: Collecting data by asking questions. In G. Semin, R. Guen, & R. K. Fiedler (Eds.), *Applied social psychology* (pp. 65–90). London: Sage.
- Schwarz, N., & Bless, H. (1992). Scandals and the public's trust in politicians: Assimilation and contrast effects. *Personality and Social Psychology Bulletin*, 18, 574–579.
- Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & Norbert Schwarz, (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (pp. 436–457). Malden, MA: Blackwell.
- Schwarz, N., & Strack, F. (1991). Context effects in attitude surveys: Applying cognitive theory to social research. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 31–50). Chichester, UK: Wiley.
- Sedikides, C., & Skowronski, J. J. (1991). The law of cognitive structure activation. *Psychological Inquiry*, 2, 169–184.
- Seger, C. A. (1994). Implicit learning. *Psychological Bulletin*, 115, 163–196.
- Sia, T. L., Lord, C. G., Blessum, K. A., Ratcliff, C. D., & Lepper, M. R. (1997). Is a rose always a rose? The role of social category exemplar change in attitude stability and attitude-behavior consistency. *Journal of Personality and Social Psychology*, 72, 501–514.
- Sigall, H., & Page, R. (1971). Current stereotypes: A little fading, a little faking. *Journal of Personality and Social Psychology*, 18, 247–255.
- Sinclair, L., & Kunda, Z. (1999). Reactions to a Black professional: Motivated inhibition and activation of conflicting stereotypes. *Journal of Personality and Social Psychology*, 77, 885–904.
- Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, 26, 1329–1342.
- Smith, E. R. (1984). Model of social inference processes. *Psychological Review*, 91, 392–413.
- Smith, E. R. (1989). Procedural efficiency and on-line social judgments. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 19–37). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology*, 27, 107–123.
- Smith, E. R. (1996). What Do Connectionism and Social Psychology Offer Each Other? *Journal of Personality and Social Psychology*, 70, 893–912.
- Smith, E. R. (1997). Preconscious Automaticity in a Modular Connectionist System. In R. R. Wyer, Jr. (Ed.), *Advances in social cognition* (Vol. 10, pp. 187–202). Mahwah, NJ: Lawrence Erlbaum Associates.
- Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology*, 71, 888–898.



- Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review*, *99*, 3–21.
- Smith, P. B., & Bond, M. H. (1993). *Social psychology across cultures: Analysis and perspectives*. New York: Harvester Wheatsheaf.
- Smolensky, P. (1989). Connectionist modeling: Neural computation/mental connections. In L. Nadel, L. A. Cooper, P. Culicover, & R. M. Harnish (Eds.), *Neural connections, mental computation* (pp. 49–67). Cambridge, MA: MIT Press.
- Spencer, S. J., Fein, S., Wolfe, C. T., Fong, C., & Dunn, M. A. (1998). Automatic activation of stereotypes: The role of self-image threat. *Personality and Social Psychology Bulletin*, *24*, 1139–1152.
- Srull, T. K. (1983). Organizational and retrieval processes in person memory: An examination of processing objectives, presentation format, and the possible role of self-generated retrieval cues. *Journal of Personality and Social Psychology*, *44*, 1157–1170.
- Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, *37*, 1660–1672.
- Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *Journal of Abnormal and Social Psychology*, *11*, 182–192.
- Stapel, D. A., & Schwarz, N. (1998). The Republican who did not want to become president: Colin Powell's impact on evaluations of the Republican party and Bob Dole. *Personality and Social Psychology Bulletin*, *24*, 690–698.
- Strack, F. (1994). Response processes in social judgment. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition, Vol. 1: Basic processes; Vol. 2: Applications* (pp. 287–322). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Strack, F., Schwartz, N., Bless, H., & Kuebler, A. (1993). *European Journal of Social Psychology*, *23*, 53–62.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass/Pfeiffer.
- Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin*, *77*, 273–295.
- Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). San Diego, CA: Academic Press.
- Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review*, *101*, 34–52.
- Wegner, D. M., & Erber, R. (1992). The hyperaccessibility of suppressed thoughts. *Journal of Personality and Social Psychology*, *69*, 903–912.
- Wiles, J., & Humphreys, M. S. (1993). Using artificial neural nets to model implicit and explicit memory test performance. In P. Graf & M. E. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology* (pp. 141–165). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126.
- Winkielman, P., Schwartz, N., Fazendeiro, T. A., & Reber, R. (2003). The hedonic making of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *Affective processes in cognition and emotion* (pp. 189–217). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology*, *81*, 815–827.
- Wyer, R. S. (1970). Quantitative predictions of belief and opinion change: A further test of a subjective probability model. *Journal of Personality and Social Psychology*, *16*, 559–570.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology*, *9*, 1–27.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, *35*, 151–175.
- Zajonc, R. B. (2000). Feeling and thinking: Closing the debate over the independence of affect. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition. Studies in emotion and social interaction, second series* (pp. 31–58). New York: Cambridge University Press.