The automaticity of evaluation

Melissa J. Ferguson

Cornell University

To appear in J. A. Bargh (Ed.), *Automatic Processes in Social Thinking and Behavior*. Psychology Press.

Authors' note: Preparation of this chapter was supported in part by grants R03-MH067877 and R03-MH73629 from the National Institute of Mental Health (NIMH). Correspondence should be directed to the author at the Department of Psychology, Cornell University, 211 Uris Hall, Ithaca, NY 14853. Email address: mjf44@cornell.edu

The automaticity of evaluation

People's ability to assess the evaluative nature of stimuli in their environment has been a central topic of study in psychology throughout the last 100 years (e.g., Allport, 1935; Brown, 1998; Eagly & Chaiken, 1993; Higgins & Brendl, 1996; McGuire, 1969, 1985; Osgood, Suci, & Tannenbaum, 1957; Rosenberg, 1965; Tesser & Martin, 1996; Zajonc, 2000). Although the scholarly questions concerning evaluations have been distinct and wide-ranging, a substantial amount of the work has been conducted using the methodology of direct self-reports (see Himmelfarb, 1993; Krosnick, Judd, & Wittenbrink, 2005). This method entails asking people to overtly describe their attitudes and evaluations, typically by identifying the number along a scale (e.g., a number between 1 and 11) that indicates the degree to which a given stimulus is pleasing or displeasing. In this way, the evaluations, and evaluative processes, that have been examined over the last century can be primarily described as intentionally generated and consciously accessible.

The last 20 years of findings in attitude research stand in stark contrast with this tradition. This research has benefited from advances in methodology that allow the indirect and subtle measurement of evaluations. This work has demonstrated that in addition to consciously, carefully, and deliberately assessing whether something is good or bad, people also evaluate stimuli nonconsciously and unintentionally upon the immediate perception of the stimuli. People assess stimuli as positive or negative within half a second after perceiving them, without intending or being aware of such assessments, or even necessarily of the stimuli themselves (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Fazio, 2001; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Greenwald, Klinger, & Liu, 1989; Greenwald, McGhee, & Schwartz, 1998; Zajonc, 1980). For instance, people can process whether a face is positive or negative based solely on a

subliminal, 4-millisecond presentation of that face on a screen, even while remaining unaware that a face was even presented (e.g., Murphy & Zajonc, 1993; Niedenthal, 1990; Öhman, 1986). The fact that evaluative processes can occur without the perceiver's intention, awareness, or control has led to their characterization as *automatic* (see Bargh, 1994; for evidence of effortlessness, see Hermans, Crombez, & Eelen, 2000).

Interest in the existence and limitations of *automatic evaluation* has especially increased over the last ten years, and has produced a sizable body of empirical findings. For example, the *Journal of Personality and Social Psychology* and *Cognition and Emotion* have each devoted journal issues to implicit measures of evaluation in recent years (2001 and 2002 respectively), suggesting the topic's centrality in contemporary research (for reviews see Banaji, 2001; Bassili & Brown, 2005; Blair, 2002; Fazio, 2001; Fazio & Olson, 2003; Musch & Klauer, 2003). The focus on how evaluative processes in particular can operate automatically is part of a broader, developing conceptualization of human cognition as driven by both controlled and automatic processing (e.g., Bargh & Ferguson, 2000; Chaiken & Trope, 1999; Greenwald, & Banaji, 1995; Hassin, Uleman, & Bargh, 2005; Sloman, 1996), as is evidenced by the other chapters in this volume.

The current chapter reviews the research on automaticity in evaluation over the last several decades with a focus on the last ten years. After some initial comments on the terminology in this literature, findings are discussed with regard to the measures that capture automatic attitudes and evaluations, the degree to which such evaluations are contextually independent, the correspondence between evaluations that are generated automatically versus deliberately, the extent to which automatic evaluations can be generated toward novel or unfamiliar stimuli, and the range of downstream consequences of such evaluations on subsequent

thinking, feeling, and acting. Lastly, the chapter considers possible underlying cognitive architectures for evaluative processing, some aspects of which are constrained and informed by recent findings.

Some comments on terminology

Before beginning a review of the measures that are used to assess automatic evaluations, some comments on the terminology used in the area seem necessary. The first issue concerns the use of the terms *attitude* and *evaluation*. Although the two terms are often used interchangeably, they may suggest different assumptions about the way in which the construct is represented in memory and generated on perception of a given stimulus. *Attitude* is a long-standing construct in social psychology, and has been defined in dozens of ways over the last century (e.g., Allport, 1935; Doob, 1947; Osgood, Suci, & Tannenbaum, 1957; Sarnoff, 1960; M. B. Smith, Bruner, & White, 1956; Thurstone, 1931). One relatively recent and widely accepted definition comes from Eagly and Chaiken (1993), who define it as a "psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor." This definition depicts the theme of *liking* or *disliking* that is common across many of the definitions.

Over the last twenty years, however, the attitude construct has been increasingly understood as an association in memory between the corresponding object and its evaluation (Fazio, 1986; Fazio, Chen, McDonel, & Sherman, 1982). Because this definition includes an assertion about the representational basis of the construct, it has been highly influential within social cognitive research, which focuses on the processes and representational format that might underlie social psychological constructs (e.g., Kunda, 1999; Moskowitz, 2005). For example, research suggests that the strength of the association between a given object and its corresponding evaluation determines the degree to which that attitude will influence a range of

downstream behavior and judgment toward the object (e.g., Fazio, 1989, 1990; Fazio & Williams, 1986; Petty & Krosnick, 1995). From the perspective of much of this research, an *attitude* refers to a stored, summary index of positivity or negativity that is associated with the object representation, and that remains dormant and inert (unchanged) until it is acted upon by independent retrieval processes (Eagly & Chaiken, 1993; Fazio, 1986; Fazio et al., 1982; Fiske & Pavelchek, 1986; for alternative views, see Bassili & Brown, 2005; Duckworth, Bargh, Garcia, & Chaiken, 2002; Ferguson & Bargh, 2002, 2004; Mitchel, Nosek, & Banaji, 2003; Schwarz & Bohner, 2001; Tesser, 1978).

The term *evaluation* also refers to the positive or negative assessment of a stimulus, but has been less formally defined in the literature compared with the term attitude (though see Tesser & Martin, 1996). The use of the term *evaluation* in this literature probably derives in part from the use of the term "automatic evaluation," which describes the act of evaluating a stimulus without intention or awareness (e.g., Bargh, Chaiken, Raymond, & Hymes, 1996; Bargh et al., 1992; Chaiken & Bargh, 1993). Whereas an *attitude* might be assumed to reflect (only) preexisting evaluative information associated with the object representation, an *evaluation* seems to invoke fewer assumptions about the underlying architectural possibilities. The term *evaluation* denotes the end-product of an evaluative process, which is necessarily on-line and does not explicitly constrain the source(s) of the evaluative information (e.g., Bassili & Brown, 2005; Brendl & Higgins, 1996; Duckworth et al., 2002; Ferguson & Bargh, 2002, 2004; Gawronski, Strack, & Bodenhausen, in press; Mitchel et al., 2003; Schwarz & Bohner, 2001; Tesser, 1978). That is, the evaluation of any given object might be constructed based on multiple sources of evaluative information, beyond the information associated with the object representation itself (e.g., see Bassili & Brown, 2005). Thus, although the terms *evaluation* and

attitude are interchangeable based on their similar meaning (i.e., an assessment of positivity or negativity), the former term seems to invoke fewer theoretical implications concerning the potential underlying representational format. Although both terms will appear in the present chapter, it should be noted that the term *attitude* here does not imply any particular underlying cognitive architecture. The issue of representation and generation is discussed at the end of the chapter.

It is customary in the literature on automatic evaluation and attitudes to refer to the targets of evaluation as "attitude objects" (e.g., Allport, 1935; Bargh et al., 1992; Fazio, 2001; Fazio et al., 1986; Sarnoff, 1960; M. B. Smith, et al., 1956; Thurstone, 1931). Although the primary definition of the word *object* is "something material that may be perceived with the senses," the use of the term in the attitude literature is based on its secondary meaning of "something mental or physical toward which thought, feeling, or action is directed" (Merriam-Webster, 2005). In other words, it can denote anything that is discriminable, or a subject of thought (Eagly & Chaiken, 1993). This can include concrete inanimate (e.g., fruit, bottle) as well as animate stimuli (e.g., people, animals), event and issues (e.g., abortion, death penalty), and abstract notions and values (e.g., liberty, freedom). As Thurstone (1931) noted, an attitude refers to the "affect for or against a *psychological* object" (p. 261; italics added).

Another source of potential confusion about the terminology in this area concerns the measures that are used to assess attitudes in an implicit fashion (these measures are discussed in more detail in the following section). Some controversy exists about whether the dissociation between attitudes that are measured implicitly versus explicitly (e.g., see Blair, 2002; Dovidio, Kawakami, Johnson, Johnson & Howard, 1997; Fazio & Olson, 2003) reflects two distinct underlying attitude constructs (i.e., an implicit versus explicit attitude; Devine, 1989; Wilson,

Lindsey, & Schooler, 2002) or merely differences in the way in which a single underlying construct is measured (Fazio, Jackson, Dunton, & Williams, 1995; Fazio & Olson, 2003; Brendl, Markman, & Messner, 2005). Because there is as of yet no definitive answer as to whether the two kinds of measures tap the same (e.g., Fazio & Olson, 2003), related (e.g., Hofman, Gawronski, Gschwendner, Le, & Schmitt, in press; Nosek, in press), or independent constructs (e.g., Wilson et al., 2000), some theorists have cautioned against referring to implicitly measured attitudes as "automatic attitudes," or "implicit attitudes", which gives the impression that an attitude measured implicitly is distinct from one measured explicitly (see De Houwer, in press; Fazio & Olson, 2003). In this chapter, the terms "automatic attitudes" and "implicit attitudes" will be used with the proviso that they do not imply any assumptions about the qualitative distinction between such attitudes and those attitudes measured explicitly. The issue of whether differences between implicitly and explicitly measured attitudes reflect substantive versus merely methodological differences is discussed in a subsequent section.

Finally, the measures that are used to assess automatic evaluations are typically referred to as "implicit" (Banaji, 2001; Greenwald & Banaji, 1995) because they assess the attitude indirectly or covertly, that is, without asking the respondent to report her or his attitude. This is an important difference from the traditional, explicit and direct way in which attitudes have been measured throughout the past century (e.g., Himmelfarb, 1993; Krosnick et al., 2005). However, as some have noted (see De Houwer, in press; Fazio & Olson, 2003), the term *implicit* is used in cognitive psychology to describe an influence of past experience that cannot be recalled, regardless of effort and intention (Roediger, 1990; Squire & Kandel, 1999; Tulving & Craik, 2000). This definition of the term may not accurately describe the way in which attitudes are measured in social psychology as there is typically no evidence that participants who are

completing the "implicit" measures are unable to access their evaluations of the pertinent stimuli. Thus, it should be noted that there are some differences in the way in which the term *implicit* is used across social and cognitive psychology.

Evidence for automatic evaluation

Psychologists have traditionally measured attitudes by directly asking people how they feel toward certain people, groups, objects, issues, and concepts (see Albarracín, Johnson, & Zanna, 2005). For example, initial attempts to assess people's attitudes toward stigmatized groups often involved asking respondents to report their agreement with stark statements such as "Black people are generally not as smart as whites" (McConahay, 1986). Given the rise over the second half of the twentieth century in social pressures for equal treatment and egalitarian behavior toward all people irrespective of their group membership(s) (e.g., McConahay, 1986; Myers, 1993), researchers began to suspect that people might be reluctant to openly admit inter-group negativity (see also Dovidio, Mann, & Gaertner, 1989; Jones & Sigall, 1971; Katz & Hass, 1988; Sears, 1988). To circumvent this reluctance, researchers devised more subtle attempts to gauge a person's attitudes, and these efforts led to the development of implicit attitude measures. In the present section, the implicit measures of attitudes that are predominant in contemporary research are described (for other implicit attitude measures see De Houwer, 2003; De Houwer & Eelen, 1998; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Koole, Dijksterhuis, & van Knippenberg, 2001; Niedenthal, 1990; Nosek & Banaji, 2001; Payne, Cheng, Govorun, & Stewart, in press; von Hippel, Sekaquaptewa, & Vargas, 1997).

Evaluative priming paradigm

In 1986, Fazio and colleagues published the first empirical article testing whether people's evaluations or attitudes toward stimuli are automatically activated from memory on the

mere perception of the stimuli (see also Fiske, 1982). Fazio et al. (1986) modified a semantic priming paradigm that was originally developed to examine automatic and controlled processing in semantic memory (Logan, 1980; Meyer & Schvaneveldt, 1971; Neely, 1976, 1977; Posner & Snyder, 1975; Shiffrin & Schneider, 1977). Neely (1977) was interested in testing whether the perception of a stimulus leads to the automatic activation in memory of knowledge semantically related to that stimulus. To explore this question, Neely constructed a series of prime-target pairs that were or were not semantically related, and assessed whether the perception of a prime stimulus, such as *BIRD*, led to faster responses to semantically related targets, such as *ROBIN*, compared with when the preceding prime was a nonsense stimulus (XXX). Faster responses to the target *ROBIN* when it was preceded by the prime *BIRD* (versus XXX) would suggest that semantically related knowledge about birds (including knowledge about robins) was automatically activated upon reading *BIRD,* which would allow faster encoding of related targets. Neely (1977) found that when the stimulus onset asynchrony (SOA) between the prime and target was shorter than 500 ms, knowledge semantically related to the prime stimulus was indeed activated regardless of the perceiver's intentions and strategic processing (see also Meyer & Schvanevedlt, 1971; Neely, 1991; Posner & Snyder, 1975).

Fazio and colleagues tested whether the perception of a stimulus also led to the automatic activation of *evaluative* information about that stimulus, along with other semantically related information. To do so, Fazio et al. (1986) constructed prime-target pairs that were unrelated semantically, except for sharing (or not sharing) a positive or negative valence (e.g., sunshine-wonderful, death-excellent). Based on the research by Neely (1977), Fazio et al. (1986) used a brief SOA (300 ms) in the paradigm to ensure that any effect of the primes on the speed of responses to targets would reflect non-strategic and unintentional (i.e., automatic) processing. In

three experiments, participants were faster to respond to the target adjectives when the targets and primes shared (versus did not share) the same valence. Fazio et al. concluded that this *evaluative priming effect* suggested that the evaluative information about primes was automatically activated, and then allowed faster encoding and responding to similarly evaluative targets. This evaluative priming paradigm thus provided the first evidence that evaluative information about stimuli (i.e., attitudes, evaluations) is automatically activated on perception of those stimuli.

The evaluative priming effect has since been replicated numerous times using a variety of experimental stimuli and response tasks. For example, the effect has emerged when the prime stimuli consist of words (e.g., Bargh et al., 1992; Fazio et al., 1986), pictures (Giner-Sorolla, Garcia, & Bargh, 1999), and odors (Hermans Baeyens, & Eelen, 1998). Although much of the research in this area has used an evaluation response task ("Is this target word positive or negative?"), the effect has also emerged when participants are asked to pronounce the targets (Bargh et al., 1996; Duckworth, Bargh, Garcia, & Chaiken, 2002; Glaser & Banaji, 1999; Hermans, DeHouwer, & Eelen, 1994), or generate a speeded motor response task to the primes alone (Chen & Bargh, 1999; Duckworth et al., 2002; Wentura, 2000).

There has been considerable debate concerning the generality of the effect across stimuli (i.e., whether attitude strength moderates automatic attitude activation; Castelli, Zogmaister, Smith, & Arcuri, 2004; Chaiken & Bargh, 1993; Fazio, 1993) and response tasks (e.g., Klauer & Musch, 2003; Wentura, 1999, 2000), as well as possible underlying mechanisms and boundary conditions of the effect (e.g., Fazio, 2001; Ferguson & Bargh, 2003; Klauer & Stern, 1992; Klauer & Musch, 2003; Klinger, Burton, & Pitts, 2000; Wentura, 1999). Nevertheless, much of this debate has centered on the consequences of automatic evaluations once they have been

generated (e.g., when and how will they influence how subsequent stimuli are processed?); the evidence that evaluations are automatically generated is considerable and largely non-controversial.

The evaluative priming paradigm has also been used as an implicit attitude measure (e.g., Fazio et al., 1995; Ferguson & Bargh, 2004; Wittenbrink, Judd, & Park, 1997; Wittenbrink, Judd, & Park, 2001). Researchers have used versions of this paradigm to assess participants' automatic evaluation of a particular set of prime stimuli by investigating how the perception of the primes influences subsequent responses to positive versus negative adjectives, compared to some comparison set of prime stimuli. If participants respond faster to positive versus negative adjectives after a given prime stimulus (compared to a control prime stimulus), then one can assume that the participants generated a positive evaluation of that prime. If on the other hand participants respond faster to negative versus positive adjectives, one can conclude that they exhibited a negative evaluation of the prime stimulus (e.g., Fazio et al., 1995; Ferguson & Bargh, 2004; Wittenbrink et al., 2001). For instance, it is possible to measure whether participants automatically generate a positive or negative evaluation of a Black face by seeing whether the perception of the face prime facilitates their subsequent responses to unrelated positive versus negative adjectives.

Implicit Association Test (IAT)

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwarz, 1998) assesses the degree to which people implicitly associate a class of objects with pleasant versus unpleasant stimuli (for reviews see Fazio & Olson, 2003; Greenwald & Nosek, 2001; Nosek & Banaji, chapter X of this volume). In this task, participants are asked to perform two sorting tasks simultaneously. If researchers want to gauge participants' automatic attitudes toward women

versus men, for example, one sorting task would be to press a certain response key if the stimulus that appears on the computer screen is a member of the category female names (e.g., *Linda*) or a member of the positively valenced category flowers (e.g., *daffodil*). Participants are asked to press a different response key if the stimulus is a member of the category male names (e.g., *Frank*) or a member of the negatively valenced category bugs (e.g., *roach*). The second sorting task is a reversal of the first one -- participants are asked to group female names with bugs, and male names with flowers. Researchers can then assess via response latencies and error rates the relative ease of the two sorting tasks in order to infer whether participants are faster to group female names with flowers versus bugs, compared to male names. If it is generally easier to group female names with flowers, then it is assumed that participants more strongly associate women (versus men) with positive versus negative information.

Research on the IAT has now yielded an extraordinary amount of data, particularly through its implementation on the worldwide web (see Nosek, Banaji, & Greenwald, 2002). For example, researchers have investigated self-esteem (Greenwald & Farnham, 2000), prejudice (e.g., Ashburn-Nardo, Voils, & Monteith, 2001, Blair, Ma, & Lenton, 2001; McConnell & Leibold, 2001), social identity (Greenwald, Banaji, Rudman, Farnham, & Nosek, 2002), gender bias in mathematics (Nosek, Banaji, & Greenwald, 2002), and personality traits (e.g., Jordan, Spencer, & Zanna, 2003; Marsh, Johnson, & Scott-Sheldon, 2001). Although researchers are still clarifying the precise nature of the associations captured by the IAT (e.g., what kinds of behavior they predict, their external and predictive validity; see Hofman et al., in press; Karpinski & Hilton, 2001; Mierke & Klauer, 2003; Nosek, Greenwald, & Banaji, 2005; Olson & Fazio, 2004), the IAT has proven highly effective in implicitly tapping the relative ease with which people associate certain stimuli with evaluative information.

There are some differences between the IAT and the evaluative priming paradigm in terms of why they are considered implicit measures. The automaticity of evaluations captured by the evaluative priming paradigm is assumed because the SOA normally used in the paradigm (less than 500 ms) is too brief a period to allow strategic, intentional processing (Neely, 1976, 1977). In addition, participants are never asked to explicitly evaluate the prime stimuli, so their evaluations are necessarily spontaneous, especially in paradigms in which participants are not asked to explicitly evaluate the target stimuli (see Bargh et al., 1996). Finally, the evaluative priming effect has emerged even when the primes are presented subliminally (e.g., Ferguson, Bargh, & Nayak, 2005; Greenwald et al., 1989; Murphy & Zajonc, 1993; Niedenthal, 1990), demonstrating that the activation of evaluative information does not require participants' awareness of the stimuli or their intention to evaluate.

In comparison, the associations captured by the IAT are considered implicit primarily because participants are not aware that their attitudes toward the stimuli are being measured, and also because they are instructed to respond as quickly as possible, thereby minimizing strategic processing. Nevertheless, whereas associations captured by the IAT can be considered implicit (i.e., indirect; see Fazio & Olson, 2003), the extent to which they rely on strategic processing is not yet clear. For example, there is some evidence that the evaluations captured by the IAT are more subject to controlled processing than those measured by an evaluative priming paradigm (see Payne, 2006; Sherman, Rose, Koch, Presson, & Chassin, 2003). Indeed, the complexity of the classification task of the IAT (sorting the target stimuli according to two orthogonal dimensions) is greater than that of the evaluation decision task or lexical decision task typically used in the evaluative priming paradigm (sorting the stimuli according to one dimension), and

thus perhaps necessarily involves more strategic (versus automatic) processing (see Balota & Lorch, 1986).

One interesting area of research concerns the application of process dissociation procedures to various tasks (PDP; Jacoby, 1992, 1997; Jacoby, Yonelinas, & Jennings, 1997; see also Conrey, Sherman, Gawronski, Hugenberg, & Groom, in press; Sherman, in press). The process dissociation framework assumes that any measurement task will capture aspects of both automatic and controlled processing – in other words, no response is *process pure* in that it results from only one kind of processing. For example, when there is no conflict between a correct response and automatic tendencies on a trial in a task, controlled and automatic processes will work together in synchrony to produce the response. However, when automatic tendencies and the correct response are in opposition, the two types of processes will be working against one another to produce the response. PDP involves the computation of estimates for each type of processing, and then the identification of the relative contribution of each to a pattern of responses (for more detail on process dissociation procedures, see Payne, chapter X in this volume). Researchers have begun to apply process dissociation techniques to implicit attitude measures (e.g., Payne, 2001; Payne, Jacoby, & Lambert, 2005; Sherman & Payne, in press), and this will potentially shed light on the relative influence of automatic versus controlled processing across the various measures.

Brain imaging methodologies

Brain imaging methodologies represent a new direction in the study of implicit evaluation. Researchers have begun to use these techniques both to gather evidence for implicit evaluation, and also to investigate the particular brain areas implicated in automatic versus strategic evaluation (e.g., Amodio, Harmon-Jones, & Devine, 2003; Cunningham, Johnson,

Raye, Gatenby, Gore, & Banaji, 2003; Ito & Cacioppo, 2000; Phelps, O'Connor, Cunningham, Funayama, Gatenby, Gore, & Banaji, 2000). Ito and Cacioppo (2000), for example, asked participants to categorize a series of social (e.g., a photograph of a couple embracing; mourners at a graveside) and nonsocial (e.g., chocolate bar; littered beach) stimuli. Participants were asked to categorize the stimuli according to a non-evaluative dimension (i.e., "Were people present or absent?"), and while they were doing so their event-related brain potentials (ERP) were recorded. Even though participants were not explicitly relying on evaluative information to complete the categorization task, there was increased electroencephalographic activity when the stimuli suggested an evaluative inconsistency (negative stimulus in a positive context, or the reverse), suggesting participants' unintentional sensitivity to the evaluative dimension.

Cunningham et al. (2003) have recently reported data that suggest which brain regions are involved in implicit evaluative processing. Participants underwent functional magnetic resonance imaging (fMRI) while they were asked to explicitly assess famous names along an evaluative ("good or bad?") or non-evaluative ("past or present fame?") dimension. The findings suggested greater amygdala activity for negative versus positive famous names, regardless of whether the task was evaluative or not. These data suggest that participants assessed the evaluative implications of the stimuli unintentionally while they were focused on a non-evaluative dimension. Other research as well has found that the amygdala is particularly active in response to negative stimuli (e.g., LeDoux, 2000; Phelps, O'Connor, Gatenby, Gore, Grillon, & Davis, 2001; Zald & Pardo, 1997), even when the negative stimuli are nonconsciously processed (e.g., Cunningham, Johnson, Raye, Gatenby, Gore, & Banaji, 2004; Morris, Ohman, & Dolan, 1998).[1]

Research on brain imaging during evaluative processing also suggests that different brain regions are implicated in evaluative processing depending on the demands of the evaluative task. Whereas the amygdala and right inferior prefrontal cortex (PFC) tend to be more active in response to negatively versus positively valenced stimuli, regardless of whether the perceiver is intentionally evaluating the stimuli (e.g., see Cunningham et al., 2003), activity in the medial and ventrolateral PFC is greater when participants are explicitly making evaluative judgments about the stimuli, especially when the stimuli are evaluatively complex. This suggests that more evaluatively complex information might demand greater deliberate evaluative processing (Cunningham et al., 2003; Cunningham, Raye, & Johnson, 2004). Data from research using brain imaging techniques have converged with behavioral evidence to suggest that automatic evaluation is a pervasive, constant, and spontaneous activity, and at the same time suggest that intentional and unintentional evaluative processing may each depend on different neural substrates.

The influence of context on automatic evaluation

Contextual independence

One of the central questions guiding research on automatic evaluation over the past five years has been the degree to which such evaluations are stable across time and contexts. Initially, the attitudes and evaluations captured by implicit measures were embraced as potentially reliable indices precisely because they were assumed to be independent of the context in which they were measured (e.g., Banaji, 2001; Bargh, 1999; Bargh et al., 1992; Bargh et al., 1996; Devine, 1989; Fazio et al, 1995; Greenwald et al., 1998; Wilson & Hodges, 1992; Wilson et al., 2000), unlike the contextual influences inherent in explicit attitude measurement (see Banaji, 2001; Fazio et al., 1995; Schwarz & Bohner, 2001). That is, because implicit measures

can assess participants' evaluations without their awareness, participants were assumed to be unable to strategically edit or modify their responses. Accordingly, those studying socially sensitive topics such as prejudice toward groups began to rely on implicit (versus explicit) attitude measures in order to obtain a cleaner index of people's socially undesirable thoughts and feelings (e.g., Bargh, 1999; Devine, 1989; Fazio et al., 1995; Swanson, Rudman, & Greenwald, 2001).

The assumptions of independence were bolstered by findings suggesting that while participants' responses on explicit measures showed little or no negativity toward traditionally stigmatized groups, their automatic reactions were considerably more negative (e.g., Devine, 1989; Fazio et al., 1995; Greenwald et al., 1998). For example, Fazio et al. (1995) demonstrated that even though some participants' automatically activated attitudes toward Blacks were negative, their explicitly reported attitudes toward Blacks as assessed by the Modern Racism Scale (MRS; McConahay, 1986) were highly positive. The reactivity of the MRS was also demonstrated when participants reported less prejudice when the experimenter was Black versus White (Fazio et al., 1995, study 3). Theorists speculated that whereas people's desire to avoid appearing prejudiced prompts them to edit and modify (i.e., hide) their true feelings on explicit measures, no such modification is possible on the implicit measures and thus their true feelings are captured (though see Arkes & Tetlock, 2004; Banaji, Nosek, & Greenwald, 2004).

Initial presumptions of contextual independence and stability were also supported by the early findings from the evaluative priming paradigm. Researchers noted the consistency between items that were explicitly evaluated and those that were spontaneously evaluated (e.g., Bargh et al., 1992). For example, Bargh et al. (1992) gathered normative data on the prime stimuli used in their studies, and found that participants' automatic evaluations largely

corresponded with the normative, explicit ratings of the respective objects. In other words, objects that had been explicitly classified as positive (e.g., puppy, sunshine) did in fact seem to facilitate response latencies to positive versus negative adjectives, relative to objects that had been classified as negative (e.g., death, cockroach). The signature evaluative priming effect therefore inherently depended on some degree of correspondence between implicit and explicit evaluations.

Contextual-dependence

As research on the two types of attitude measures accumulated, evidence began to emerge that implicitly measured attitudes were often weakly correlated or completely unrelated to attitudes measured explicitly (e.g., Banaji & Greenwald, 1995; Fazio et al., 1995; Karpinski & Hilton, 2001; though see Cunningham, Preacher, & Banaji, 2001; McConnell & Leibold, 2001; Wittenbrink, Judd, & Park, 1997). Although some degree of disconnect between the two types of measures was expected given the differences in the nature of the measures, a complete lack of correspondence worried researchers, and some questioned the construct validity of implicitly measured attitudes and evaluations (see Banaji, 2001 for a discussion). If implicit measures were tapping people's "true" attitudes and preferences, then they should at least partially correspond with related measures under some circumstances, in line with basic conventions regarding convergent and criterion validity. This concern provoked considerable research efforts at examining the stability and contextual independence of implicitly measured attitudes and the relation between implicit and explicit measures in general (e.g., for a review see Blair, 2001; Brauer, Wasel, & Niedenthal, 2000; Dovidio, Kawakami, & Beach, 2001; Fazio & Olson, 2003; Nosek, 2005).

The findings from this research activity suggest a host of contextual influences on implicitly measured evaluations and attitudes, contrary to the initial assumptions of contextual independence. Specifically, findings suggest that the direction and strength of an automatic evaluation of a given object vary depending on the type of recently activated, or repeatedly learned, object-relevant information (e.g., Barden, Maddux, Petty, & Brewer, 2004; Dasgupta & Greenwald, 2001; Karpinski & Hilton, 2001; Livingston & Brewer, 2002; Lowery, Hardin, & Sinclair, 2001; Mitchell et al., 2003; Wittenbrink et al., 2001). For example, researchers found that participants displayed significantly less negative automatic evaluations toward group members who are commonly targets of prejudice after being exposed to pro-elderly stimuli (Karpinski & Hilton, 2001), exemplars of well-liked African-Americans and disliked white people (Dasgupta & Greenwald, 2001, Experiment 1), or presentations of a movie clip of an African-American family enjoying themselves at a picnic (Wittenbrink et al., 2001, Experiment 1).

Recent work has suggested that automatic evaluations are also influenced by the type of goal or objective that the perceiver is currently pursuing (e.g., Ferguson & Bargh, 2004; Lowery et al., 2001; Moors & De Houwer, 2001; Moors, De Houwer, & Eelen, 2004; Sherman et al., 2003, study 2) as well as by the perceiver's chronic motivations (Maddux, Barden, Brewer, & Petty, 2005). For example, Lowery et al. (2001) found that participants who completed the IAT under the direction of a Black (versus white) experimenter exhibited significantly reduced negative attitudes toward Blacks, demonstrating that automatic attitude measures may be susceptible to social influence pressures. Mitchell et al. (2003) recently demonstrated that automatic evaluations are also dependent on participants' categorization goal as they encounter the respective objects. For example, Mitchell et al. (2003) selected Black targets who were liked

athletes, and White targets who were disliked politicians. When participants were instructed to classify the Black and White targets in an IAT in terms of career, African-American exemplars were more easily associated with positively (versus negatively) valenced stimuli compared with when they were classified in terms of race. These findings together show that different contextual parameters, such as task objectives and instructions, can influence the type of information that is activated with regard to the object(s) of interest.

Further evidence of the goal dependent nature of automatic evaluations comes from Ferguson and Bargh (2004). In this work, participants who were currently pursuing a goal (or not) completed an evaluative priming paradigm that measured their automatic evaluations of objects that varied in their relevance to the goal. The results suggest that objects that were relevant to the goal were automatically evaluated as most positive when the perceiver was still pursuing the goal versus had already completed it. For example, participants who were thirsty implicitly evaluated the highly thirst-relevant objects (e.g., water, juice) as most positive, compared with the other objects (e.g., chair, table), and the participants who had just quenched their thirst. These findings demonstrate that automatic evaluations are sometimes prospective with regard to the utility of the objects, and are not always solely a function of recent experience with the objects. As such, the findings are in accord with the classic notion that attitudes are functionally tied to the perceiver's current motivational priorities (e.g., Cacioppo, Priester, & Berntson, 1993; Chen & Bargh, 1999; Katz, 1960; Lewin, 1935; Pratkanis et al., 1989; Rosenberg, 1956; Wentura, Rothermund, & Bak, 2000).

People's chronic goals can also influence automatic evaluations. Maddux et al. (2005) recently showed that the impact of contextual cues on participants' automatic evaluations of Blacks depended on participants' chronic motivation to avoid being prejudiced. Those

participants low in this motivation exhibited negative automatic evaluations toward Blacks in contexts that were threatening (e.g., a prison cell) compared to non-threatening (e.g., a church), which is consistent with previous research (Wittenbrink et al., 2001). In contrast, however, those participants high in the motivation to avoid being prejudiced actually showed less negative evaluations of Blacks in the threatening context, compared with other participants overall and also with high-motivation participants in the non-threatening context. Interestingly, these participants' less negative evaluations resulted from an inhibition of negative information in the threatening condition. This work suggests that people's chronic motivations can determine the way in which they respond to contextual cues regarding the nature of the evaluated stimuli (see also Moskowitz, Gollwitzer, Wasel, & Schaal, 1999).

Compatibility of findings on independence versus dependence

Given the array of findings over the last five years suggesting the contextual dependence of automatic evaluation, how is it possible that early work suggested stability and contextual-independence? Firstly, none of the initial experiments on automatic attitude activation manipulated the nature of object-relevant, recently activated information (e.g., Bargh et al., 1992; Bargh et al., 1996; Fazio et al., 1986). In other words, it may be that given a default context (i.e., one in which there is no manipulation of preceding, object-relevant information), the perception of the word *dentist*, for example, may automatically and explicitly evoke a negative evaluation much of the time for most participants. Yet, this does not preclude the possibility that the accessibility of positive dentist-related information (e.g., a prestigious profession) would lead to a positive automatic evaluation of *dentist* (Ferguson & Bargh, 2003). Furthermore, the objects that were used in the early studies were strongly normatively positive (e.g., gift) or negative (e.g., poison), compared to the more evaluatively ambiguous attitude objects that have been

studied in more recent articles. The more an object is associated in memory with evaluatively divergent information, the more likely the object will be automatically evaluated in different ways across time and situations.

Early research on implicitly measured attitudes also rarely examined particular objects as the dependent variable. Most often the dependent variables of interest consisted of groups of positive and negative attitude objects, and groups of positive and negative target words (e.g., see Bargh et al., 1992, 1996; Fazio et al., 1986) and five or six objects typically comprised each valence group. Because of this grouping, it may be that although the intensity or even direction of the attitude for each object varied to some extent, when grouped together with others the group remained positive or negative. This method of grouping can hide the fluctuations of the evaluations of individual objects.

Although research over the last five years clearly demonstrates that both implicit and explicit attitudes are contextually dependent, recent research suggests that the two types of measurements are not equivalently sensitive to the same set of factors. For instance, one interesting area of recent research suggests that implicit attitudes may not reflect certain situational influences that are easily integrated into explicit attitudes. For example, Gawronski and colleagues have argued that whereas explicit attitudes are sensitive to dissonance pressures, implicit attitudes are not (see Gawronski et al., in press; Gawronski & Strack, 2004). In a recent set of experiments, Gawronski and Strack (2004) demonstrated that people who had written a counterattitudinal essay under low perceived situational pressure exhibited more pro-issue explicit attitudes than those who had written it under high perceived pressure, in line with the classic findings by Festinger and colleagues (e.g., Festinger, 1957; Festinger & Carlsmith, 1957). However, there was no difference in participants' implicit attitudes across the situational

pressure conditions, suggesting that such attitudes might not represent the (potentially nonconscious; Liberman, Ochsner, Gilbert, & Schacter, 2001) cognitive restructuring that occurs in dissonance-inducing situations (see also Kruglanski & Klar, 1987). Further research continues to explore the parameters of contextual influence on such evaluations, and also the degree to which such influence compares and contrasts to the contextual dependence of explicit measures.

Compatibility of automatically versus deliberately activated evaluations

The issue of whether automatic evaluations correspond with those that are generated more deliberately has been of central interest since the advent of implicit attitude measures. As mentioned in the earlier section on contextual dependence, initial findings suggested a dissociation between attitudes assessed by these two types of measurements, particularly in the area of prejudice (e.g., Banaji & Hardin, 1996; Blair & Banaji, 1996; Bosson, Swann, & Pennebaker, 2000; Devine, 1998; Dovidio et al., 1997; Fazio et al., 1995; Hofman et al., in press; Greenwald et al., 1998). This dissociation was originally interpreted as possible evidence that implicit measures were capturing the negative feelings that participants were not willing to express on explicit measures (e.g., Fazio et al., 1995). Since this work, however, research suggests that these two types of measures can sometimes provide compatible attitudes, especially if the error inherent in each type of measure is controlled (Cunningham, Nezlek, Banaji, 2004; Cunningham, Preacher, & Banaji, 2001).

Moreover, instead of asking *whether* explicit attitudes are consistent with implicit attitudes, the focus of research has moved toward asking *when* such consistency will emerge (see Devine, 2001; Fazio & Olson, 2003). To this end, researchers have identified moderators of the relationship between explicitly and implicitly measured attitudes (e.g., see Dovidio et al., 1997; Fazio et al., 1995; Hofman et al., in press; Nosek, in press). The social stigmatization of the

objects seems to be one important moderator, which is why the evidence of dissociation has emerged mostly in the area of prejudice. When respondents are concerned about the social desirability of their attitudes, they may edit their explicitly reported attitudes, which might result in the dissociation between those attitudes and their automatically activated ones (e.g., Fazio et al., 1995; Greenwald et al., 1998).

The strength of the attitude is also assumed to moderate whether explicit and implicit measures will be in agreement (e.g., Fazio, 1986, 1995; Krosnick & Petty, 1995). When the association between the object and its respective evaluation is strong, the attitude should be more likely to be automatically activated on the mere perception of the object. This should mean that the attitude will be more likely (than a weak attitude) to become spontaneously activated and influential on explicit judgments and behavior, and this in turn should lead to greater correspondence between implicit and explicit measures of that attitude (see also Fazio, 2001).

Nosek (in press) recently reported two additional moderators of the relationship between attitudes assessed by these two types of measurements. One concerns the dimensionality of the respective attitude objects, or, in other words, whether the object is considered to be bipolar or unipolar (e.g., see Pratkanis, 1989). A dimension is referred to as bipolar when one end of the dimension represents the acceptance of the object and the other represents the rejection of the object. For example, attitudes toward the death penalty presumably lay along a bipolar continuum in that acceptance (or support) of the death penalty necessarily means that the person does not reject (or disagree with) the death penalty. A dimension that is unipolar, on the other hand, represents simply more or less positivity toward the object, and can be orthogonal to one's negativity toward the object (see Cacioppo & Gardner, 1999; Larsen, McGraw, & Cacioppo, 2001). Nosek (in press) proposed that because bipolar attitudes tend to be more consistent across

time and structured in a simpler manner compared with unipolar attitudes (see Judd and Kulik, 1980), there should be more consistent explicit-implicit relations for bipolar versus unipolar attitudes, and the data supported this prediction.

Additionally, Nosek (in press) proposed that the distinctiveness of the attitude objects should moderate the relations between implicit and explicit attitude measures. Distinctiveness here refers to the degree to which one's own attitude differs from the perceived norm. To the extent that one's attitude is highly distinctive (e.g., one might not like chocolate even though the vast majority of people do), she or he is likely to be more aware of the attitude, and may therefore express it more consistently. As such, highly distinctive attitudes should show greater consistency across implicit and explicit measures, and this hypothesis was also supported by Nosek's findings.

The research in this area also addresses the question of whether attitudes that are measured implicitly are substantively distinct from those measured explicitly. Researchers differ in terms of whether they believe that inconsistency in implicit-explicit relations suggests two different underlying constructs (e.g., Doviodio et al., 1997; Greenwald & Banaji, 1995; Wilson et al., 2000), two related but distinct constructs (Cunningham et al., 2001; Gawronski et al., in press; Greenwald & Farnham, 2000; Nosek, in press), or one construct that can be measured in different manners (Fazio & Olson, 2003). Whereas the first two views assume that implicit and explicit attitudes are based on non-overlapping or partially overlapping evaluative information, the one-construct perspective assumes that only the expression of the construct differs depending on the type of measurement. One may modify one's (automatic) attitude if he or she has the motivation and opportunity to do so, but should be drawing on approximately the same evaluative information in either case (see Fazio & Olson, 2003). The question of whether

inconsistency between implicit and explicit measures reflects differences in representational format necessarily depends on one's assumptions about the cognitive architecture that might enable evaluative processing. This issue is discussed in more detail in a subsequent section.

Development of automatic evaluations

Is it possible to automatically evaluate an object that one has never seen before (and cannot easily categorize)? Although the development of automatic evaluations has generally been understudied (though see Betsch, Plessner, Schwieren, & Gutig, 2001; Duckworth et al., 2002; Olson & Fazio, 2001), theory concerning automatic evaluations as well as notions about information processing in general assume that underlying responses must be repeatedly enacted over time in order to operate automatically (Bargh, 1984, 1990; Fazio, 1986; Logan, 1980; Shiffrin & Dumais, 1981; Smith & Lerner, 1986). In a classic example, Shiffrin and Schneider (1977) demonstrated that participants had to be trained on over 1,000 trials in order to be able to automatically perceive a given target (e.g., "C") in differently sized arrays of distracters (but see Smith & Lerner, 1986).

The assumption that automatic evaluations in particular require some experience with the respective objects is inherent in the dominant view of attitude representation and organization in memory (Fazio 1986, 1990, 2001). Fazio and others have asserted that an attitude consists of evaluative information (positivity or negativity) associated with the object representation in memory within an associative network organization (see Eagly & Chiaken, 1993; Fazio, 1986, 1990, 2001). From this perspective, attitude strength is a function of the strength of the association between the evaluative information and the object representation. A strong attitude will have a stronger link, or association, with the object, compared with a weaker attitude. Fazio and colleagues have argued that only those objects with strong, associated attitudes should be

automatically evaluated (Fazio et al., 1986). This assertion implies that only objects that have been strongly (repeatedly) evaluated in the past will be able to be evaluated automatically, and in accord with this Fazio and colleagues have manipulated the strength of attitudes in experiments by varying how often participants repeat their evaluations of the objects (e.g., Fazio & Williams, 1986; Roskos-Ewoldsen & Fazio, 1992; Smith, Fazio, & Cejka, 1996). The assumption that automatic evaluations require some previous experience with the respective objects implies that people might be unable to automatically evaluate novel or unfamiliar attitude objects. That is, if people are encountering an object for the first time (and the object is not easily categorizable), they have likely not yet developed an association between the object representation (which may itself be only partially formed) and evaluative information (i.e., an attitude).

The claim that automatic evaluations are slow to develop is also in accord with the assertion that they are slow to change in the face of new object information or experience. For example, researchers have argued that the evaluations or attitudes that are activated automatically from memory are slow to form and change, and are generally resistant to persuasive efforts (e.g., Wilson & Hodges, 1992; Wilson, Lindsey, & Schooler, 2000). In their model of dual attitudes (i.e., implicit and explicit), Wilson et al. (2000) argued that although the generation and expression of explicit attitudes are highly contingent on the context at the time of measurement (e.g., recently activated memories, the perceiver's goals and strategic concerns, etc.), implicit attitudes are assumed to be unchanging and resistant. For instance, they asserted that "Explicit attitudes change relatively more easily whereas implicit attitudes, like old habits, change more slowly. Attitude change techniques often change explicit but not implicit attitudes" (p. 4). This view is also consistent with efforts at changing automatic attitudes through extensive

training regimens (e.g., Karpinski & Hilton, 2001; Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000; Rudman, Asmore, & Gary, 2001).

The view that automatic evaluations might be restricted to familiar objects, and resistant to new information about the respective objects, is also in accord with the contention in social and cognitive psychology that the integration of recently acquired evaluative information requires conscious effort (e.g., Bargh, 1984, 1989, 1990; Jastrow, 1906; Schneider & Fisk, 1982; Shiffrin & Dumais, 1981; Shiffrin & Schneider, 1977; Smith & Lerner, 1986; but see Dijksterhuis, 2004; Gollwitzer, 1996, 1999). Assuming that the evaluation of unfamiliar objects requires some (novel) integration of evaluative information, this perspective implies that people should be unable to evaluate such objects without conscious awareness and deliberation (e.g., Wilson & Hodges, 1992; cf. Bargh, 1999; Duckworth et al., 2002). Explicit attitude measures, on the other hand, ostensibly allow more complex integrations of object-relevant and novel information.

There have been several articles over the last five years that directly address assumptions regarding the development of automatic evaluations. For instance, Duckworth et al. (2002) found that participants were able to automatically evaluate unfamiliar abstract art drawings and novel "pseudo-Turkish" words. Participants automatically evaluated as positive the stimuli that had been explicitly classified as positive by a separate group of participants, and their automatic evaluations also tended to correlate with the explicit ratings of the negative stimuli. Importantly, these findings demonstrate that people can evaluate objects for which they do not have pre-existing object representations in memory, and can do so "on the fly," or within a fraction of a second after perceiving them.

Olson and Fazio (2001) also addressed the development of implicit attitudes, and examined whether they might emerge as a result of the implicit detection of co-variation between a novel attitude object and valenced stimuli. Participants watched a series of images and words appear in a seemingly random order on a computer screen, and were asked to press a key whenever a particular object appeared. Embedded in the images that appeared were pairings of a novel attitude object (a Pokemon character) with either positively (e.g., picture of a sundae) or negatively (e.g., picture of a cockroach) valenced stimuli. (Participants did not have to respond to any of these stimuli.) Results demonstrate that participants were unaware of the co-variation, and reported (explicit) attitudes in line with the conditioning procedure. In a different experiment that included the same conditioning procedure, participants also exhibited implicit attitudes as measured by the IAT in line with the conditioning procedure. This research demonstrates that people unintentionally and seemingly nonconsciously can implicitly form both explicit and automatic attitudes (see also De Houwer, Baeyens, & Field, 2005; De Houwer, Thomas, & Baeyens, 2001; Martin & Levy, 1978; Walther, Nagengast, & Trasselli, 2005).

In an intriguing series of experiments, Betsch et al. (2001) addressed a similar question, and examined whether people's evaluations are sensitive to recently encountered valenced information about the corresponding objects, even when that information is learned passively and without the intention to evaluate the objects. Betsch et al. also predicted that participants' evaluations would reflect the *sum* of the valenced information rather than the *average* (see Anderson, 1971, 1981). Participants watched a series of ads appear on a computer screen while saying aloud the share returns from five companies that scrolled across the top of the screen. Participants' focal task was to remember the information in the ads and they were led to believe that the share information was meant as a distraction from the focal task. They were presented

with 75 total pieces of return information (15 per share) and watched a total of 60 ads. At the end of the task, they were asked to evaluate the shares, and their evaluations did in fact reflect the summed returns, rather than the average. This research demonstrates that people absorb even complex evaluative information about objects without intending to do so. It also speaks to the mechanism of this process by showing that people are sensitive to the sum rather than the average of the information, which would be predicted by some models of cognitive integration (Anderson, 1971, 1981; see also Fredrickson & Kahneman, 1993; Redelmeier & Kahneman, 1996).

Recent work by Castelli and colleagues (Castelli et al., 2004) suggests that people can implicitly evaluate recently encountered stimuli even when they cannot remember the details of the stimuli that led them to their initial evaluations. Participants who learned about target persons described as child molesters later implicitly evaluated those people as negative even when they could not accurately recall whether the people were described as child molesters or as belonging to a more benign category (e.g., teachers). Castelli et al. discuss how this work challenges the assumption that implicit evaluations reflect summary indices associated with object representations in memory. Instead, they suggest that implicit evaluations might be better understood as being generated by a connectionist system that naturally incorporates contextual information in an integrative and online fashion, and can handle dissociations between explicit and implicit memory (see also Bassili & Brown, 2005; Ferguson & Bargh, 2003; Mitchell et al., 2003). This issue is revisited later in the chapter.

The notion that people can quickly and spontaneously generate evaluative assessments of stimuli in their environment is consistent with a long history of research in experimental learning where a wide variety of animals are able to learn fear responses after one exposure to the

threatening stimulus (e.g., for a review see LeDoux, 2000). For example, after rats are shocked a single time while exploring a particular caged environment, they subsequently exhibit fear responses when placed back into the same cage. Such one-trial fear learning has obvious adaptive advantages and has been demonstrated in worms, flies, rats, monkey, and humans. Animals that are able to avoid making the same mistake twice, with other animals, plants, or dangerous elements of any kind, clearly possess an evolutionary advantage. This research is consistent with work showing that people are able to evaluate novel objects on the basis of little information, and with relative ease and little deliberation.

Also in accord with the recent findings by Castelli et al. (2004), work in cognitive science on implicit memory also suggests that people can retain implicit memories of negative or threatening people even if their explicit memory fails them (e.g., Squire & Kandel, 1999; Squire, 1992). For example, Squire and Kandel (1999) described the classic example of a person with anterograde amnesia who is introduced to a new person who unexpectedly shocks the amnesiac's hand with a hand buzzer as they are shaking hands. After leaving the room, several minutes later the amnesiac reports that she has no memory of the shocker. However, the amnesiac demonstrates that she must have retained implicit evaluative memory when she later refuses to shake the shocker's hand, presumably on the basis of the earlier negative experience (see also Gazzaniga & Heatherton, 1992). This type of work coincides with much research on implicit memory (e.g., Squire, 1992), and suggests that people might be able to implicitly evaluate stimuli for which there is no (accessible) explicit memory or even very limited previous experience, contrary to long-standing notions about the repetition needed for automatic responses to develop (Bargh, 1984, 1990; Fazio, 1986; Shiffrin & Dumais, 1981; Shiffrin & Schneider, 1977; Smith & Lerner, 1986).

Consequences of automatic evaluation

In addition to questions concerning the contextual dependence and development of automatic attitudes, a fundamental question in this area concerns the downstream consequences of such attitudes for thinking, feeling, and acting. What are the consequences of automatic evaluations for the evaluated stimuli themselves as well as for unrelated stimuli? This question has been of utmost importance in this area of research for two central reasons. The first is that the way in which automatic evaluations influence processing of subsequently encountered stimuli might suggest important constraints on the underlying architecture (see below).

The second reason is that, as previously mentioned, researchers have sought substantive evidence that implicitly measured attitudes or evaluations represent hypothetical constructs that influence behavior. Indeed, the area of explicit attitudes underwent its most intense period of scrutiny and criticism after Wicker (1969) published his now famous paper criticizing the lack of correspondence between explicit attitudes and the actual behaviors the attitudes were supposed to predict (see also LaPiere, 1934; Schuman & Johnson, 1976; Thurstone, 1928). As a result, researchers have closely attended to this issue with regard to automatic evaluations, and there has been considerable work to this end over the last five years (see Fazio & Olson, 2003).

There are two streams of research concerning the consequences of automatic evaluations. The first concerns the immediate implications of an evaluative act itself for the subsequent processing of stimuli that are related or unrelated to the originally evaluated object. That is, what are the implications of such fleeting generations of positivity or negativity, if any? The second stream of research concerns the degree of predictive utility of these kinds of evaluations. In other words, do such evaluations possess criterion and predictive validity such that a positive

automatic evaluation of a stimulus predicts approach behaviors toward that stimulus in other situations?

Immediate consequences

For evaluated objects themselves. Researchers have argued that automatic evaluations are functional because they quickly afford the perceiver with valuable information about how to behave toward the evaluated objects (e.g., Chen & Bargh, 1999; Duckworth et al., 2002; Fazio, 1989; Ferguson & Bargh, 2002, 2004; Lang, Bradley, & Cuthbert, 1990; LeDoux, 1996; Öhman, 1986; Pratkanis, Breckler, & Greenwald, 1989; Roskos-Ewoldsen & Fazio, 1992). This functional argument presupposes a strong link between the evaluation of an object and immediate approach versus avoidance behavioral tendencies toward that object, and several lines of research support this supposition. For example, a series of studies suggests that the automatic evaluation of an object influences extension or flexion arm movements in line with the nature of the particular evaluation. Chen and Bargh (1999) asked participants to either push a lever away from them (extension) when they saw a word appear on the computer screen in front of them, or pull the lever toward them (flexion). When participants were pushing the lever away from them, they were significantly faster to do so when the stimuli on the screen were negative in valence (e.g., *poison*) compared to positive in valence (e.g., *puppy*). Participants were also significantly faster to pull the lever toward them when the stimuli were positive versus negative. In line with a functional view of automatic evaluations, this suggests that positivity is associated with approach arm movements whereas negativity is associated with avoidance arm movements (see also Duckworth et al., 2002).

There is a large body of work outside of the automatic evaluation literature that suggests that the quick and unintentional evaluation of stimuli has implications for judgment and

decision-making concerning those stimuli. Although these studies do not typically use implicit evaluation measures, they nevertheless suggest that the spontaneous affective assessment of an object has a range of consequences for how the perceiver reacts toward that object. For example, recent work suggests that the way in which a perceiver spontaneously evaluates an object influences her or his recognition of the stimulus (Monin, 2003). Monin argued that people misattribute the positivity of a stimulus as evidence for its familiarity, and thereby commit the so-called *Warm Glow* heuristic. Across a series of experiments, people were significantly more likely to falsely recognize novel stimuli if the stimuli were attractive or positive in valence versus unattractive, neutral, or negative in valence. Although Monin did not ensure that the evaluations of the stimuli were automatic in nature, they were unprompted and thus the findings constitute preliminary evidence that spontaneous evaluations of stimuli can influence unrelated recognition judgments about the stimuli themselves.

Other findings provide additional evidence that automatic evaluations of stimuli influence "real world" judgment and decision-making concerning those stimuli (e.g., Damasio, 1999, 2001; Epley & Caruso, 2004) as well as moral judgments about events related to the respective stimuli (e.g., Haidt, 2001, 2003). Work by Damasio and colleagues, for instance, suggests that the way in which people affectively and immediately react to stimuli serves as an important signal for how they should behave toward the stimuli (see Damasio, 1999). According to Damasio's framework, after people experience the affective consequences of certain stimuli (e.g., eating sugar feels good), they learn to anticipate the emotional consequences of stimuli without necessarily re-experiencing those emotional consequences. Once people have learned that eating sugar is pleasurable, for instance, stimuli that have previously preceded or co-varied with that feeling (e.g., candy) become spontaneously and quickly evaluated as positive. This

evaluative reaction allows the perceiver to react to the stimulus (in this case, probably approach it) without having to first re-experience the positive emotion. Such pre-emptive appraisals afford people extra time to prepare to secure or avoid the stimulus, and such preparation undoubtedly leads to greater success in obtaining or avoiding it. From Damasio's perspective then, contrary to the traditional viewpoint of affect biasing or interfering with otherwise rational judgment and decision making, quick affective reactions provide information that is essential for adaptively making decisions and performing actions toward those stimuli (see Damasio, 1994, 1999, 2001; Fazio, Blascovich, & Driscoll, 1992; Frank, 1988, 2003, Zajonc, 1980).

In a related line of work, Haidt and colleagues (Greene & Haidt, 2002; Haidt, 2001, 2003) have argued that the immediate affective assessment of an event forms the basis for reasoning about the morality of that event. In opposition to classic models of moral reasoning in which affect is incidental, disruptive, or peripheral to the moral judgment, Haidt (2001, 2003) argues that people first experience an inevitable affective reaction to a behavior or action based on learned cultural norms (their "moral intuition"), and then attempt to justify their assessment of the action as moral according to various post-hoc reasons. In this model, people's immediate affective reactions to events and behaviors and people form the basis upon which subsequent judgments are formed, even outside of people's intention and awareness. For example, Haidt has argued that people who are given the hypothetical scenario of a brother and sister having sex together under select circumstances base their moral judgment of the scenario on the way in which they (usually negatively) affectively react to it, and try to generate various deliberate reasons that will explain their judgment (Haidt, 2001; see also Nisbett & Wilson, 1977). According to this framework, the reasons for a moral judgment are post-hoc, and derive from the immediate affective reactions to the event under consideration.

Together, the research described above suggests that the way in which people automatically evaluate objects in their environment has consequences for how they subsequently judge and behave toward those objects. This recent work suggests, for instance, that when people spontaneously evaluate a person as negative, they may be more likely to engage in subtle avoidance behaviors and judge the person along more negative dimensions than if they had first evaluated the person in a more positive manner. Clearly, even though automatic evaluative acts last only a fraction of a second and occur without the person's intention or awareness, they have a host of downstream outcomes that can influence how people see and understand the world around them.

For unrelated objects. Does an evaluative act have consequences for how people process subsequent, unrelated stimuli? The literature on evaluative priming suggests that the automatic evaluation of a stimulus does enable faster responding to subsequent, similarly valenced stimuli, even if there is otherwise no relation between the evaluated object and the target (e.g., for a review see Musch & Klauer, 2003). But beyond effects on the speed of responding to subsequent stimuli, does the automatic evaluation of an object influence the way in which people interpret and understand completely unrelated objects?

The automatic evaluation of clearly valenced stimuli seems to influence evaluative judgments about different, unrelated stimuli (e.g., Muphy & Zajonc, 1993; Niedenthal, 1990; Stapel, Koomen, & Ruijs, 2002). For instance, Murphy and Zajonc (1993) demonstrated that subliminal presentations of smiling or frowning cartoon faces led to respectively more positive or negative evaluations of novel Chinese characters. Stapel et al. (2002) recently expanded on this research by showing that the effects of subliminally presented cartoon face stimuli on judgments of affect in subsequent target cartoon faces depended on the presentation time of the

subliminally presented prime stimuli. The effects were diffuse when the prime stimulus (e.g., a smiling man) was presented for 30 ms ("positive"), but more circumscribed when it was presented for 100 ms ("a smiling man"). This research together suggests that valenced stimuli such as smiling faces can influence people's subsequent perception of emotion as well as their liking judgments of novel stimuli. But are effects of automatic evaluation for unrelated stimuli limited to liking judgments?

Early research on impression formation argues against any effect of evaluative primes on non-liking judgments about unrelated stimuli. Research by Higgins, Rholes, and Jones (1977) on the consequences of trait priming for person judgment, for example, demonstrated that primes only influenced judgments when they were applicable -- that is, related to the behavior that was relevant to the judgments. For example, participants primed with the trait prime *neat* were <u>not</u> more likely to later judge Donald's ambiguously reckless behaviors as adventurous, even though *neat* and *adventurous* share the same (positive) valence. This led Higgins et al. to conclude that inapplicable primes do not influence non-liking judgments, a sentiment that has been reinforced by numerous studies since then (e.g., Bargh, Bond, Lombardi, & Tota, 1986; Devine, 1989; Erdley & D'Agostino, 1988; Higgins, 1996; Sedikides, 1990), and referred to by some as an axiom in priming research (see Stapel and Koomen, 2000).

However, recent studies suggest that there are circumstances under which the evaluation of inapplicable trait primes does influence person judgments. Croizet and Fiske (2001) conducted a conceptual replication of Higgins et al. (1977) but induced some participants to feel like experts in social judgment. Croizet and Fiske (2001) replicated the findings of Higgins et al. only with those participants not induced to feel like experts, and found an effect of the inapplicable primes for those who did feel like experts. Croizet and Fiske (2001) speculated that

the feeling of expertise might have lowered the threshold of usability of activated knowledge for these participants relative to the default situation (see Higgins, 1996).

Stapel and Koomen (2000) have also conducted research concerning this issue and found that trait primes that are extremely valenced and broad (e.g., *bad*) can influence judgments about unrelated behavior while moderately valenced and or narrow trait terms (e.g., *frugal*) do not. This suggests that certain traits might be sufficiently strong and valenced to influence subsequent unrelated trait judgments. For example, while the prime *good* should lead people to interpret an ambiguously adventurous behavior as adventurous (rather than reckless), the prime *frugal* should have no effect.

Other recent work suggests that automatic evaluations have even broader implications for subsequently encountered, unrelated stimuli. Whereas research by Stapel and Koomen (2000) and Criozet and Fiske (2001) suggests that certain trait terms can influence subsequent unrelated person judgments if participants have a certain goal or if the trait terms are broad and evaluatively extreme, recent findings suggest that the automatic evaluation of positive and negative everyday nouns (e.g., movies, germs) can influence a range of judgments about unrelated stimuli (Ferguson et al., 2005). In a series of experiments, the automatic evaluation of subliminally presented nouns influenced how participants subsequently interpreted words, categorized people and objects, and rendered explicit personality judgments about people. For example, participants were asked to classify a series of people (e.g., Bill Clinton; Mike Tyson) and items (e.g., chocolate; vodka) according to the word that seemed to best capture the object. Each object was presented with two possible words that differed in valence; for example, vodka was presented with the words “hangover” and “party” and Bill Clinton was presented with the words “politician” and “adulterer”. The results demonstrated that the valence of the subliminally

presented (yet unrelated) primes influenced the descriptions that participants chose as best capturing the objects. Those primed with positive primes were significantly more likely to associate the subsequent objects with the positive descriptions compared to those primed with control primes and negative primes.

The above findings suggest that automatic evaluations have potentially long-ranging implications because they have the ability to influence the explicit interpretations and judgments about subsequently encountered but semantically unrelated stimuli. This means that quick evaluative acts can influence the way in which people see, understand, and act toward the evaluated stimuli, as well as toward completely unrelated but temporally close stimuli. Any ambiguous object that is encountered immediately after an evaluative act can be potentially disambiguated in line with the valence of the initial evaluated object, even when there is no relation between the two.

General processing effects. Some recent findings suggest that the automatic evaluation of stimuli can also influence people's mood, which can then influence stereotyping effects in accord with how explicitly manipulated mood typically does so. Chartrand, Bargh, & van Baaren (2004) demonstrated that participants who were subliminally primed with positive words reported significantly more positive affect than those primed with negative words. In addition, those who reported a positive mood were significantly more likely to stereotype on an implicit measure. These findings suggest that a series of automatic evaluative acts of the same valence can induce mood states, with all the downstream implications of such an affective state (e.g., Forgas, 2000, 2001; Martin, & Clore, 2001; Ortony, Clore, & Collins, 1988). Furthermore, because automatic evaluations occur without the perceiver's awareness or intention, the reasons for such moods would be difficult for the perceiver to identify.

Correspondence

The second area of research that speaks to the consequences of automatic evaluations concerns their predictive validity. As previously mentioned, a traditionally central research question concerning automatic evaluations is their ability to predict people's behavior, either toward the evaluated stimuli themselves, or toward similar stimuli (e.g., Arkes & Tetlock, 2004; Banaji, 2001; Banaji & Nosek, 2004; Dovidio, Kawakami, & Gaertner, 2002; Fazio & Olson, 2003; Kawakami & Dovidio, 2001; Lambert et al., 2005). From this perspective, people's automatic evaluations can be conceptualized as indices of how they will behave toward some related object in the future. One important distinction between the previous section on consequences and the current one is that whereas the former reviewed research that demonstrated the causal force of automatic evaluations on subsequent judgment and behavior, this area summarizes research on the correlational relationship between automatic evaluations and related judgment and behavior.

One of the first articles to establish the predictive validity of automatic evaluations was conducted by Fazio and colleagues (Fazio et al., 1995). The findings showed correspondence between people's automatic evaluations of Black faces, and behaviors toward a Black experimenter. Across a series of studies, the negativity of participants' automatic evaluations of Blacks predicted participants' subsequent, nonverbal behaviors with a Black experimenter (as judged by the experimenter). In addition, participants' automatic evaluations also predicted some of their opinions about the well-publicized (at the time) trial of police brutality against Rodney King. Subsequent research has confirmed the correspondence (McConnell & Liebold, 2001) between people's automatic evaluations of Blacks as measured by the IAT and their nonverbal behavior toward a Black person.

Research by Nosek et al. (2002) also supports some degree of correspondence between people's automatic evaluations of objects and their performance in domains related to those objects. Participants' implicitly measured attitudes toward math, as assessed by the IAT, predicted their scores on the scholastic aptitude test (SAT). These findings strongly suggest that the degree to which people implicitly associate a certain object with positive versus negative words indicates their behaviors toward that object at different times and in unrelated circumstances. This work suggests the criterion validity of such attitudes, at least in some domains.

Researchers in this area have speculated that whereas explicitly measured attitudes might best predict deliberate and conscious behaviors, implicitly measured attitudes might ultimately predict those behaviors that are primarily driven by automatic processing (e.g., Blair, 2002; Dovidio, Kawakami, & Gaertner, 2002; Dovidio et al., 1997; Kawakami & Dovidio, 2001; Lambert et al., 2005). For example, researchers have frequently assessed the extent to which implicitly measured attitudes predict nonverbal (i.e., difficult to control; see Fazio & Olson, 2003) behavior toward a Black target person (though see Lambert et al., 2005), and have sometimes found that implicit attitudes predict such behaviors better than explicitly measured constructs.

Researchers are continuing to examine the predictive capacity of automatic evaluations, and also the boundary conditions for any effects (e.g., Lambert et al., 2005). For example, some have noted that the predictive validity of automatically activated attitudes might depend on various explicit motivational variables under some circumstances (Fazio & Olson; 2003; Nosek, in press), such as the motivation to modify or edit responses on explicit measures (e.g., Dunton & Fazio, 1997; Plant & Devine, 1998). The automatic attitudes of those who report low

motivation to avoid prejudice might predict a variety more overt and deliberate behaviors compared with the attitudes of those who report higher motivation. From this perspective, the predictive validity of automatic attitudes might ultimately depend both on the motivations of the perceiver and the obviousness of the behaviors.

The generation and representation of automatic evaluations

The previous review of findings on automatic evaluation suggests several constraints for the presumed representation and generation of evaluative information in memory. The findings suggest that automatic evaluations are contextually dependent, correspond with explicit attitudes under some circumstances, can be generated in response to unfamiliar stimuli, and have a host of implications for subsequent thinking, feeling, and action. The dominant perspective of how attitudes are generated and represented in memory is first described, and then an alternative perspective is considered. There is then a consideration of whether the recent findings can differentiate between the dominant and alternative views of how evaluations are represented and generated.

Single-tag perspective

How are automatic evaluations generated? What do implicit attitude measures capture in terms of what is stored in memory? The predominant theoretical model over the last two decades of how attitudes are represented and generated has been articulated by Fazio (1986, 1989, 1995, 2001; see also Pavelchak & Fiske, 1986). Fazio and others have asserted that the attitudes that are captured by evaluative priming paradigms reflect evaluative summary information that is associated in memory with the corresponding object representation (e.g., Bargh et al., 1992; Eagly & Chaiken, 1993; Fazio, 2000; Fazio et al., 1995; Fazio et al., 1986; Fiske & Pavelchak, 1986; Wilson et al., 2000). This summary evaluative index is ostensibly

formed by repeated experience with the object (see Shiffrin & Schneider, 1977), and the strength of the attitude is proportional to the strength of the association between the evaluative information and the object representation. In line with typical instantiations of associative networks (see Quillian, 1968; Collins and Loftus, 1975), when the object is perceived, its corresponding representation is activated, and activation then spreads along the associative links to semantically related information, including evaluative information. Much of the research on the automatic activation of evaluations is based on the supposition that many objects are associated in memory with a positive or negative summary index, or "tag" (e.g., Bargh et al., 1992; Bargh et al., 1996; Fazio et al., 1986; Fiske & Pavelchak, 1986).

From the single–tag perspective, a given object is associated in memory with a stored, summary evaluation of that object and, importantly, this is what is measured in an attitude implicit measure (plus measurement error). In other words, this suggests that what is measured by an implicit attitude measure is a 1-to-1 mapping of the observed response to the stored attitude toward that object in memory (plus measurement error). According to this framework, assuming a perfect measure, it is possible to directly measure the evaluative summary index associated in memory with a given attitude object.

This perspective is most closely aligned with classic, so-called localist, symbolic models of memory that often presuppose associative networks in which isolated nodes represent individual constructs, exemplars, or features of an object (e.g., Anderson, 1983; Anderson & Bower, 1973; Collins & Loftus, 1975). These nodes are interconnected according to the degree of (semantic) relation between the nodes, with activation spreading along these links automatically on perception of an object (e.g., Meyer & Schvaneveldt, 1971; Neely, 1976, 1977; Posner & Snyder, 1975; Shiffrin & Schneider, 1977). Such models have been referred to as

"file-drawer" models of memory, and have guided research and theory in social psychology for decades (e.g., see Carlston & Smith, 1996; Fazio, 1986; Hastie, 1980; Smith, 1996, 1998; Srull, 1981, 1983).

The single-tag perspective has been unquestionably essential in guiding a social cognitive analysis of how evaluative information can be automatically activated in memory on perception of the corresponding objects. Furthermore, this perspective, with its emphasis on the importance of the strength of the association between the object representation and the attitude, paved the way for process-oriented research on the mediating role of attitude accessibility in attitude-behavior relationships (e.g., Bassili, 1996; Fazio et al., 1986; Fazio & Williams, 1986; Roskos-Ewoldsen & Fazio, 1992; Smith, Fazio, & Cejka, 1996).

Constructive perspective

In recent years, a variety of alternatives to the single-tag perspective have emerged (e.g., Bassili & Brown, 2005; Duckworth, Bargh, Garcia, & Chaiken, 2003; Ferguson & Bargh, 2003; Gawronski et al., in press; Greenwald, 1990; Livingston & Brewer, 2002; Mitchell et al., 2003; E. R. Smith, 1997, 2000). These approaches can all be considered "constructive" in that they assume that any given evaluation measured by an implicit test signifies a composite of evaluative information that is built on-the-fly, rather than a single, stored, pre-existing summary index. In this way, although such a composite would necessarily rely on stored evaluative information as the elements of the computation, any given evaluation of an object would be constructed online, across multiple sources of information that are contextually specific. Numerous researchers have argued that explicitly rendered evaluations are constructed in this fashion (e.g., Rosenberg, 1956; Salancik & Conway, 1975; Schwarz & Bohner, 2001; Tesser, 1978; Zaller, 1992), and this possibility has begun to be applied to implicitly measured evaluations (Blair, 2002; Bassili &

Brown, 2005; Duckworth et al., 2002; Ferguson & Bargh, 2003; Livingston & Brewer, 2002; Mitchell et al., 2003).

This constructive view is consistent with the notion that most attitude objects are associated in memory with a complex array of different kinds of information (e.g., Abelson, 1976, 1981; Barsalou, 1992; Bower, 1981; Carlston, 1994; Fishbein & Ajzen, 1975; Fiske & Pavelchak, 1986; Schank & Abelson, 1977; Smith, 1992; Smith & Zarate, 1992). With the additional assumption that some of the associated object memories can differ in their evaluative implications, this notion suggests that explicitly and implicitly measured attitudes will vary depending on the chronic and temporary factors in place at the time of measurement. Indeed, recent work on the contextual dependence of automatic evaluations suggests that given such an organization of evaluative information, evaluations that are measured implicitly are contingent on the evaluative profile of the most accessible attitude object memories (e.g., Dasgupta & Greenwald, 2001; Wittenbrink et al., 2001).

According to this view, a 1-to-1 mapping of the stored evaluative information in memory to the observed response (e.g., in an implicit measure) is impossible, even assuming a flawless measurement technique. Instead, any given object is evaluated based on an integration of evaluative information across multiple sources, regardless of how the object is ultimately categorized. In this way, the observed response on an implicit attitude measure does not reflect a stored, pre-existing evaluative tag associated with a category or object (see Livingston & Brewer, 2002), but instead reflects a computation performed by an evaluative system, across numerous representations such as multiple categories and exemplars that relate to the object in various ways (see Bassili & Brown, 2005; Castelli et al., 2004; Duckworth et al., 2002; Ferguson & Bargh, 2003; Fiedler, 1996; Schwarz & Bohner, 2001; Smith, 1997, 2000).

What kind of architecture would underlie a constructive process? In principle, a constructive process is consistent with the traditional "file-drawer" model of memory in which concepts are represented by nodes, and the nodes are interconnected according to semantic similarity (e.g., see Smith, 1996). In these types of models, representations are inert and static when not activated, and are periodically manipulated by processes such as encoding, storage, and retrieval (see Anderson, 1983). It would be theoretically possible for an integrative or constructive process to take place across numerous single-tag representations, although the process by which such a computation would occur is not clear, and would require additional assumptions from the single-tag perspective.

In contrast, the possibility of integrative processing across various sources of evaluative information is directly predicted by parallel distributed processing models of memory (e.g., Anderson & Rosenfeld, 1988; Bassili & Brown, 2005; Bechtel & Abrahamsen, 1991; Carlston and Smith, 1996; Masson, 1991, 1995; Smith, 1996, 1997; Smith & DeCoster, 1999). Models of connectionist systems include the assertion that every observable response (explicit or implicit) is the result of a transitive state of the mind, wherein all representations are potentially implicated or contributive (e.g., Smith, 1996; Smolensky, 1989). This would suggest that an attitude reflects the current state of activation within a connectionist system (e.g., Gawronski et al., in press; Smith, 1996), and would be influenced by multiple elements of information pertaining to the current physical and psychological circumstances of the perceiver. For instance, in order to provide evidence that (explicit) attitudes are sometimes sensitive to the context, Wilson et al. (2000) stated that "Most parallel distributed processing models assume that mental representations are highly sensitive to the current context, because aspects of the context always influence the pattern of activation that determines mental representation" (pp. 3; also see

Arieli, Sterkin, Grinvald, & Aertsen, 1996). A construction across multiple sources of evaluative information would be highly compatible with the assumptions of connectionist systems (see also Eiser, Fazio, Stafford, & Prescott, 2003; Fiedler, 1996; Mitchell et al., 2003).

Do the recent findings suggest a single-tag or constructive process?

Evidence of contextual dependence. To what degree does the recent evidence for the contextual dependence of automatic evaluations support the single-tag versus constructive perspective? According to the single-tag perspective, variability in the automatic evaluation of an object across contexts could be due to the object being implicitly categorized in different ways across situations. The assumption here would be that almost every category is associated with an evaluative tag, and the evaluation of the object would depend on how it is categorized. This would be consistent with the single-tag claim that a stored, unitary summary index is what is being captured by the implicit measure (plus measurement error).

One critical problem with the single-tag perspective, however, is that it does not address or explain inhibitory processes at work during automatic evaluative processing. Although the vast majority of research on automatic evaluation has addressed only the degree to which facilitatory processes underlie such evaluations, recent work on contextual dependence suggests that both inhibitory and facilitatory processing can be involved (Ferguson & Bargh, 2004; Maddux et al., 2005). For example, in research by Maddux et al. (2005), participants high in the motivation to avoid prejudice exhibited significantly more inhibition of negative information when perceiving an African-American in a threatening context. And, in work by Ferguson and Bargh (2004), the effect of a current goal state on automatic evaluations sometimes involved the inhibition of negativity associated with the object. These findings are difficult to reconcile with a model that posits that a single summary tag is activated in response to the perception of the

object. Namely, the possibility that positive and negative evaluative information associated with the object in memory can be facilitated and inhibited simultaneously (depending on the valence of the information) suggests that objects are associated with more than just one summary tag (see also Cacioppo & Gardner, 1999; Chaiken & Bargh, 1993; Larsen et al., 2001).

In contrast to the single-tag perspective, the findings regarding the situational influences on automatic evaluation are directly consistent with a constructive view of how evaluations are generated. A constructive model would naturally allow for contextual influence – the essence of such a model is that the evaluation of any given object depends on the evaluative profile of the object memories, and other relevant memories, activated at the time of encountering the object. This means that the evaluation of an object might vary across time and contexts according to the degree to which the object is stored in memory with information that is evaluatively complex (i.e., not uniform), and the circumstances of the situation (Cacioppo & Gardner, 1999; Chaiken & Bargh, 1993; Larsen et al., 2001).

Furthermore, the findings regarding inhibitory processing in automatic evaluation could be accounted for by a perspective that assumes a constructive, integrative process across multiple sources of information. Specifically, such a perspective would assume that different sources of evaluative information might be implicated on perception of the respective object – while some types of information might be made more accessible, other types might become less accessible (i.e., inhibited). These various effects could influence the computation that ultimately determines the final positive or negative (approach or avoid) response to the object.

It should be noted that a constructive approach would not necessarily preclude high stability of implicit attitudes across time and situations (cf. Wilson et al., 2000). Assuming that an implicit attitude represents a combination of pieces of evaluative information related to the

object and context, stability should emerge to the degree that the object is associated with mostly univalent memories, and is associated also with univalent contextual memories. For example, one's memories concerning puppies might be almost uniformly positive, and the contextual memories that are associated with puppies might also tend to be positive. If this is the case, then one might expect the automatic reaction to the word *puppy* to be relatively stable across time, assuming little measurement error (e.g., Bargh et al., 1996; Bargh et al., 1992; Fazio et al., 1986). This approach would assume that stability should decrease to the extent that the object is associated with differently valenced object and contextual memories.

It is important to note that even though the evidence of contextual dependence seems to support a constructive approach, it does not necessarily speak to the issue of the type of architecture that would underlie such a model. Again, although connectionist models have been more often formalized and modeled to account for specific phenomena compared to associative networks (see Smith, 1998), and also routinely predict inhibitory processing, it might be the case that a symbolic associative network could be bootstrapped in terms of accounting for the recent findings in this area.

Evidence of compatibility between explicit and implicit measures. Findings concerning the consistency of explicit and implicit attitudes are not exclusively supportive of either the single-tag or constructive approaches to attitude representation and generation. For example, findings suggesting a dissociation could be explained by the single-tag perspective based on the notion that a different attitude is being measured in each case (see Wilson et al., 2000). With additional time and deliberation, a person who is automatically classified (and evaluated) solely by race, for example, might upon further reflection be classified and evaluated according to profession, or personality (see Brewer, 1988; Fiske & Neuberg, 1990; Kunda, 1999). The

categories might be associated with different summary tags, which would lead to the dissociation between the attitude measures (Wilson et al., 2000).

Findings suggesting a dissociation between implicit and explicit attitudes could also be easily explained by the constructive approach. Just as the additional deliberation in explicit measures might lead to a different categorization than during an implicit measure, the same additional deliberation might influence the way in which evaluative information is constructed. For example, the reflection and introspection that is possible during an explicit measure would undoubtedly allow the activation of additional (or different) evaluative information about the object itself as well as the context in which the object is encountered. This additional reflection could change the array of information contributing to the current pattern of activation within a connectionist system.

Although the findings regarding implicit-explicit relations do not discriminate between the single-tag and constructive approaches, they have inspired researchers to claim that different processes underlie explicit versus implicit measures (Gawronski & Bodenhausen, 2004; Gawronski et al., in press; Strack & Deutsch, 2004). In line with dual-process models of social information processing (e.g., Chaiken & Trope, 1999; Lieberman, Gaunt, Gilbert, & Trope, 2002; Kahneman, 2003; Sloman, 1996; Smith & DeCoster, 2000), Gawronski and colleagues have argued that while implicit attitudes are based on associative evaluations, explicit attitudes are based on propositional evaluations. Associative evaluations are those that are automatically activated in response to the perception of a stimulus, and do not reflect the person's endorsement of the evaluation. Such evaluations can be generated via connectionist frameworks, and as such, naturally incorporate contextual information. Explicit evaluations, on the other hand, reflect more deliberate, rule-based processing that is itself based on momentarily activated associative

evaluations. The rule-based processing allows for the assessment of the "truth value" of the evaluative information – propositional evaluations only emerge if they are formally endorsed by the perceiver. In line with this perspective, research is actively examining the degree to which explicit and implicit attitudes differ in terms of whether they reflect endorsed evaluations (e.g., Olson & Fazio, 2004).

Evidence of evaluations of novelty. What does the evidence that automatic evaluation occurs for novel stimuli indicate for the organization and generation of evaluative information in memory? Although the single-tag perspective does not formally predict such effects (see Fazio, 1989, 2001; Wilson et al., 2000), it is not in principle contradictory with such findings. For example, one could surmise that novel objects are quickly classified into the most appropriate, pre-existing category (e.g., an unfamiliar abstract art image might be classified into the pre-existing category "colorful abstract art designs") and then acquire whatever stored, summary evaluative tag is associated with that category. Although this model would break down for objects that are not immediately classifiable, one could argue that there are no objects in existence that would not be immediately categorized, albeit in a perhaps mistaken or temporary manner (see Kunda, 1999).

The evidence concerning novelty is also consistent with a constructive account (e.g., Duckworth et al., 2002; Ferguson & Bargh, 2003). Findings showing that people can automatically evaluate novel or unfamiliar stimuli suggest the ability to integrate evaluative information from numerous, familiar aspects or features in order to generate an evaluative reaction to the unfamiliar object. In this way then, a person who encounters a strange animal while on vacation (e.g., a quakka on Rottnest Island in Western Australia), for example, might effortlessly and unintentionally integrate evaluative information from familiar features (e.g.,

looks like a big rat, but seems gentle like a bunny, and no growling or bared teeth that would suggest an imminent attack), and evaluate the novel animal as relatively positive and safe to approach.

In sum, although researchers have argued that the automatic evaluation of novel objects indicates that people integrate evaluative information across familiar features (e.g., Duckworth et al., 2002), these recent findings cannot ultimately discriminate between the possibility that novel objects are categorized into familiar groups and then acquire the stored, summary tag associated with that group, versus are evaluated according to a computation across multiple relevant exemplars and categories. Both perspectives remain valid at this point.

Evidence of consequences of automatic evaluations. Research concerning the consequences of automatic evaluations has traditionally informed discussions of the ways in which evaluative information might be represented in memory (for a review see Musch & Klauer, 2003). In particular, findings on the immediate consequences of automatic evaluation for both related and unrelated stimuli directly address a debate about the viability of spreading activation versus response competition as underlying mechanisms. Whereas some researchers have asserted that the automatic evaluation of an object leads to the increased accessibility of a valence construct (positive or negative; see Fazio 2001) or all similarly valenced memories ("spreading activation" accounts; Bargh et al., 1996; Ferguson et al., 2005), others have argued that such accessibility effects do not occur (response competition accounts; e.g., see Klauer & Musch, 2003; Klinger, Burton, & Pitts, 2000; Wentura, 1999). These two perspectives on what happens after an automatic evaluative act necessarily presuppose different organizations of evaluative memory. Whereas spreading activation accounts assume that evaluative information

is interconnected just on the basis of valence (though perhaps not literally as in an associative network), response competition accounts would make no such assertion.

This debate is ongoing, and in fact there is evidence that supports both spreading activation accounts (e.g., Ferguson et al., 2005) as well as response competition accounts (e.g., see Klauer & Musch, 2003; Klinger, Burton, & Pitts, 2000; Wentura, 1999). This suggests that both accounts might ultimately hold some explanatory power under various circumstances, and possibly for data gathered from particular paradigms (see Fazio, 2001). Further research on the varied consequences of an automatic evaluative act on subsequent judgment, feeling, and behavior will undoubtedly proffer implications for how evaluative information is represented and organized in memory.

Reconciliation. Based on evidence gathered so far, it is not yet possible to definitively resolve the arguably most central issues concerning automatic evaluation. The first of course is how an evaluation is generated, and whether it reflects the activation of a stored, unitary summary index associated with the object representation (or category), or a constructive process across multiple sources of evaluative information. Although the evidence concerning contextual dependence and novelty seems most consistent with the idea of a constructive model, it is not definitive.

The second issue is much broader, and concerns the question of whether memory involves localist, symbolic systems, connectionist systems, or some combination of both (e.g, see Marcus, 2001). This tension between possible cognitive architectures is driving a considerable amount of research in contemporary cognitive science, and the jury is still out on which framework is best supported by data and theory (see Fodor, 2000; Marcus, 2001; Pinker, 1999; Plotkin, 1999). At the least, there is evidence that the associative network models that have

dominated theory and research in social cognition for the last 30 years are beginning to wane in popularity (see Carlston & Smith, 1996; Smith, 1996, 1998, 2000), and some researchers in social psychology have begun to assert the advantages of connectionist frameworks for explaining social psychological phenomena (e.g., Bassili & Brown, 2005; Eiser et al., 2003; Ferguson & Bargh, 2003, 2004; Mitchell et al., 2003).

Conclusion

Research in social psychology over the last three decades, and especially over the last five to ten years, has provided considerable insight into the nature of automatic evaluation. This work suggests that people tend to evaluate a wide array of both novel and familiar stimuli, without necessarily intending to or being aware of such evaluations. Evaluations that are automatically generated tend to be highly sensitive to the context in which the person encounters the respective stimuli, and also seem to be distinct from more deliberate, conscious appraisals. Finally, even though the positivity or negativity that is activated in response to an object is fleeting, it nevertheless has various consequences for subsequent judgment, emotion, and behavior. Although these characteristics have considerably expanded an understanding of automatic evaluation, many unanswered questions remain. The most important among these would seem to be the way in which evaluative information is represented in memory and generated on perception of a given object. Future research on the boundary conditions and moderating variables underlying contextual-dependence, implicit-explicit relations, novel evaluations, and criterion and predictive validity will undoubtedly shape the answers to such questions over the coming years.

Footnotes

[1] Recent research has suggested that the amygdala is more in response to positive versus neutral stimuli (e.g., Hamann & Mao, 2002; Liberzon, Phan, Decker, & Taylor, 2003; Zald, 2003), and thus might be specialized for processing intense stimuli rather than solely negative stimuli (see Cunningham, Raye, & Johnson, 2004).

REFERENCES

Abelson, R. P. (1976). Script processing in attitude formation and decision making. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior* (pp. 33-45). Hillsdale, NJ: Erlbaum.

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist, 36,* 715-729.

Albarracín, D., Johnson, B. T., & Zanna, M. P. (2005). *The Handbook of Attitudes*. Mahwah, NJ: Erlbaum.

Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *Handbook of social psychology* (pp. 798-844). Worcester, MA: Clark University Press.

Amodio, D. M., Harmon-Jones, E., & Devine, P. G. (2003). Individual differences in the activation and control of affective race bias as assessed by startle eyeblink responses and self-report. *Journal of Personality and Social Psychology, 84*, 738-753.

Anderson, J. A., & Rosenfeld, E. (1988). *Neurocomputing: Foundations of research*. Cambridge, MA: MIT Press.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington, D.C.: Winston and Sons.

Anderson, N. H. (1974). Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 1-101). New York: Academic Press.

Arieli, A., Sterkin, A., Grinvald, A., & Aertsen, A. (1996). Dynamics of ongoing activity: Explanation of the large variability in evoked cortical responses. *Science, 273,* 1868-1871.

Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry, 15*, 257-278.

Ashburn-Nardo, L., Voils, C. I., & Monteith, M. J. (2001). Implicit associations as the seeds of intergroup bias: How easily do they take root? *Journal of Personality and Social Psychology, 81*, 789-799.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, Cognition, 12*, 336-345.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neither, & A. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 117-150). Washington, DC: American Psychological Association.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. *Psychological Science, 7(3)*, 136-141.

Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2004). No place for nostalgia in science: A response to Arkes and Tetlock. *Psychological Inquiry, 15*, 279-310.

Bargh, J. A. (1984). Automatic and conscious processing of social information. In R. S. Wyer Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 1-43). Hillsdale, NJ: Erlbaum.

Bargh, J. A. (1990). Auto-motives: Preconscious determinants of social interaction. In E. T. Higgins, & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition*: *Foundations of social behavior* (Vol. 2., pp. 93-130). New York, NY, USA: The Guilford Press.

Bargh, J. A. (1994). The Four Horsemen of automaticity: Awareness, efficiency, intention, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., pp. 1-40). Hillsdale, NJ: Erlbaum.

Bargh, J. A. (1999). The cognitive monster. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 361-382). New York: Guilford.

Bargh, J. A., Bond, R. N., Lombardi, W. J., & Tota, M. E. (1986). The additive nature of chronic and temporary sources of construct accessibility. *Journal of Personalty and Social Psychology, 50*, 869-878.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62,* 893-912.

Bargh, J. A., 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.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin, 126,* 925-945.

Barsalou, L. W. (1992). *Cognitive psychology: An overview for cognitive scientists*. Hillsdale, NJ: Erlbaum.

Barden, J., Maddux, W. W., Petty, R.E., & Brewer, M.B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87*, 5-22.

Bassili, J. N. (1996) The "How" and "Why" of Response Latency Measurement in Survey Research. In N. Schwarz & S. Sudman (Eds.), *Answering Questions: Methodology for determining cognitive and communicative processes in survey research* (pp.319-346). San Francisco: Jossey-Bass Publishers.

Bassili, J. N., & Brown, R. D. (2005). Implicit and explicit attitudes: Research, Challenges, and theory. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 543-574). Mahwah, NJ: Erlbaum.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks.* Oxford, England UK: Basil Blackwell, Inc.

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.

Betsch, T., Plessner, H., Schwieren, C., & Gütig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin, 27*, 242-253.

Blair, I. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychological Review, 6*, 242-261.

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

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

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

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129-148.

Brauer, M., Wasel, W., Niedenthal, P. M. (2000). Implicit and explicit components of prejudice. *Review of General Psychology, 4*, 79-101.

Brendl, C. M., & Higgins, E. T. (1996). Principles of judging valence: What makes events positive or negative? *Advances in Experimental Social Psychology, 28,* 95-160.

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 760-773.

Brewer, M. B. (1988). A dual process model of impression formation. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 1, pp. 1-36). Hillsdale, NJ: Erlbaum Associates.

Brown, J. D. (1998). *The self*. New York: McGraw-Hill.

Cacioppo, J. T., & Gardner, W. L. (1999). Emotion. *Annual Review of Psychology, 50*, 191-214.

Cacioppo, J. T., Priester, J. R., Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology, 65,* 5-17.

Carlston, D. E. (1994). Associated systems theory: A systematic approach to the cognitive representation of persons and events. In R. S. Wyer (Ed.) *Advances in Social Cognition: Vol. 7. Associated Systems Theory* (pp. 1-78). Hillsdale, NY: Erlbaum.

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, NY: Guilford Press.

Castelli, L., Zogmaister, C., Smith, E.R., Arcuri, L. (2004). On the automatic evaluation of social exemplars. *Journal of Personality and Social Psychology, 86*, 373-387.

Chaiken, S., Bargh, J. A. (1993). Occurrence versus moderation of the automatic attitude activation effect: Reply to Fazio. *Journal of Personality & Social Psychology, 64*, 759-765.

Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.

Chartrand, T.L., Bargh, J.A., & van Baaren, R. (2004). Linking Automatic Evaluation to Mood and Information Processing Style: Consequences for Experienced Affect, Information Processing, and Stereotyping, *Journal of Experimental Psychology: General*.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate Behavioral predispositions to approach and avoid the stimulus. *Personality and Social Psychology Bulletin, 25,* 215-224.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (in press). Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*.

Croizet, J. C., & Fiske, S. T. (2000). Moderation of priming by goals: Feeling entitled to judge increases judged usability of evaluative primes. *Journal of Experimental Social Psychology, 36,* 155-181.

Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural components of social evaluation. *Journal of Personality and Social Psychology, 85*, 639-649.

Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of Black and White Faces. *Psychological Science, 15*, 806-813.

Cunningham, W. A., Nezlek, J. B., & Banaji, M. R. (2004). Implicit and explicit ethnocentrism: Revisiting the ideologies of prejudice. *Personality & Social Psychology Bulletin, 30*, 1332-1346.

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

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience, 16*, 1717-1729.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. Orlando, FL: Harcourt.

Damasio, A. R. (2001). Fundamental Feelings. *Nature, 413*, 781.

Damasio, A. R. (1994). *Descartes' Error: Emotion, Reason and the Human Brain*. Putnam: New York.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of liked and disliked individuals. *Journal of Personality and Social Psychology, 81,* 800-814.

De Houwer, J. (in press). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction*. Thousand Oaks, CA: Sage Publishers.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77-85.

De Houwer, J., Baeyens, F., & Field, A. P. (Eds.) (2005). *Associative learning of likes and dislikes*. Hove: Psychology Press.

De Houwer, J., Crombez, G., Baeyens, F., & Hermans, D. (2001). On the generality of the affective Simon effect. *Social Cognition, 15,* 189-206.

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion, 8*, 45-61.

De Houwer, J., Thomas, S., & Baeyens, F. (2001) Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853–869.

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

Devine, P. G. (2001). Implicit prejudice and stereotyping: How automatic are they? Introduction to the special section. *Journal of Personality and Social Psychology, 81,* 757-759.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345-355.

Doob, L.W. (1947), The behavior of attitudes, *Psychological Review, 51*, 135-156.

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* (Vol. 4, Intergroup Relations, pp. 175-197). Oxford, U.K.: Blackwell.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology, 82*, 62-68.

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

Dovidio, J. F., Mann, J., & Gaertner, S. L. (1989). Resistance to affirmative action: The implications of aversive racism. In F. A. Blanchard & F. J. Crosby (Eds.), *Affirmative action in perspective* (pp. 83-102). New York: SpringerVerlag.

Duckworth, K. L., Bargh, J. A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science, 13*, 513 - 519.

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*. Fort Worth, TX: Harcourt Brace Jovanovich College.

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.*

Epley, N. & Caruso, E. M. (2004). Egocentric ethics. *Social Justice Research, 17*, 171-187.

Erdley, C. A., & D'Agostino, P. R.(1988). Cognitive and affective components of automatic priming effects. *Journal of Personality and Social Psychology, 54*, 741-747.

Fazio, R. H. (1986). How do attitudes guide behavior? In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 204-243). New York: Guilford.

Fazio, R. H. (1989). On the power and functionality of attitudes: The role of attitude accessibility. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp.153-179). Hillsdale, N.J.: Erlbaum.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

Fazio, R. H. (1993). Variability in the likelihood of automatic attitude activation: Data re-analysis and commentary on Bargh, Chaiken, Govender, and Pratto (1992). *Journal of Personality and Social Psychology, 64*, 753-758, 764-765.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion, 14,* 1-27.

Fazio, R. H., Blascovich, J., & Driscoll, D. M. (1992). On the functional value of attitudes: The influence of accessible attitudes upon the ease and quality of decision making. *Personality and Social Psychology Bulletin, 18*, 388-401.

Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency and the strength of the object-evaluation association. *Journal of Experimental Social Psychology, 18*, 339-357.

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 use. *Annual Review of Psychology, 54*, 297-327.

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., & 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.

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review, 103*, 193-214.

Ferguson, M. J., & Bargh, J. A. (2002). Sensitivity and flexibility: Exploring the knowledge function of automatic attitudes. In L. F. Barrett & P. Salovey (Eds.), *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 383-405). Guilford Press.

Ferguson, M. J., & Bargh, J. A. (2003). The constructive nature of automatic evaluation. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp.169-188). NJ: Erlbaum.

Ferguson, M. J., & Bargh, J. A. (2004). Liking is for doing: Effects of goal pursuit on automatic evaluation. *Journal of Personality and Social Psychology, 88*, 557-572.

Ferguson, M. J., Bargh, J. A., & Nayak, D. (2005). After-affects: How automatic evaluations influence the interpretation of unrelated, subsequent stimuli. *Journal of Experimental Social Psychology, 41*, 182-191.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Festinger, L. & Carlsmith, J.M. (1959). Cognitive Consquences of Forced Compliance. *Journal of Abnormal and Social Psychology, 58*, 203-210.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Fiske, S. T. (1982). Schema-triggered affect: Applications to social perception. In M. S. Clark & S. T. Fiske (Eds.), *Affect and cognition: The 17th annual Carnegie symposium on cognition* (pp. 55-78). Hillsdale, N. J.: Erlbaum.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1-74). New York: Academic Press.

Fiske, S. T., & Pavelchak, M. A. (1986). Category-based versus piecemeal-based affective responses: Development in schema-triggered affect. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (pp. 167-203). New York: Guilford.

Forgas, J.P. (Ed.) (2000). *Feeling and thinking: Affective influences on social cognition*. New York: Cambridge University Press.

Forgas, J.P. (Ed.) (2001). *The handbook of affect and social cognition*. Mahwah, New Jersey: Lawrence Erlbaum.

Frank, R. H. (1988). *Passions Within Reason: The Strategic Role of the Emotions*. W.W. Norton.

Frank, R. H. (2003). Introducing Moral Emotions into Models of Rational Choice. In A. S. R. Manstead, N. H. Frijda & A. H. Fischer (Eds.), *Feelings and Emotions: The Amsterdam Symposium*. New York, Cambridge University Press.

Gawronski, B., & Strack, F. (2004). On the propositional nature of cognitive consistency: Dissonance changes explicit, but not implicit attitudes. *Journal of Experimental Social Psychology, 40*, 535-542.

Gawronski, B., Strack, F., & Bodenhausen, G. V. (in press). Attitudes and cognitive consistency: The role of associative and propositional processes. To appear in R. E. Petty, R. H. Fazio, & P. Briñol (Eds.), *Attitudes: Insights from the new wave of implicit measures*.

Giner-Sorolla, R., Garcia, M. T., & Bargh, J. A. (1999). The automatic evaluation of pictures. *Social Cognition, 17,* 76-96.

Glaser, J. & Banaji, M.R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality and Social Psychology, 77*, 669-687.

Gollwitzer, P. M. (1996). The volitional benefits of planning. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 287-312). New York, NY: Guilford Press.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493-503.

Greene, J., & Haidt, J. (2002). How (and where) does moral judgment work? *Trends in Cognitive Science, 6*, 517-523.

Greenwald, A. G. (1990). What cognitive representations underlie social attitudes? *Bulletin of the Psychonomic Society, 28,* 254-260.

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

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

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022-1038.

Greenwald, A. G., Klinger, M. R., & Liu, T. J. (1989). Unconscious processing of dichoptically masked words. *Memory and Cognition, 17,* 35-47.

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

Haidt, J . (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834.

Haidt, J. (2003). The moral emotions. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences*. Oxford: Oxford University Press.(pp. 852-870).

Hamann S, & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport, 13*, 15-19.

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.) (2005). *The new unconscious*. [a sequel to Uleman & Bargh's Unintended Thought, 1989]. New York: Oxford University Press.

Hastie, R. (1980). Memory for behavioral information that confirms or contradicts a personality impression. In R. Hastie, et al. (Eds.), *Person memory: The cognitive basis of social perception* (pp. 155-177). Hillsdale, N.J.: Erlbaum Associates.

Hermans, D., Baeyens, F., & Eelen, P. (1998). Odours as affective processing context for word evaluation: A case of cross-modal affective priming. *Cognition and Emotion, 12*, 601-613.

Hermans, D., Crombez, G., & Eelen, P. (2000). Automatic attitude activation and efficiency: The fourth horseman of automaticity. *Psychologica Belgica, 40*, 3-22.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition and Emotion, 8*, 515-533.

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., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13,* 141-154.

Himmelfarb, S. (1993). The measurement of attitudes. In A. H. Eagly, & S. Chaiken (Eds.), *The psychology of attitudes* (pp. 23-84). Fort Worth, TX: Harcourt Brace Jovanovich College.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (in press). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*.

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology, 36*, 660-676.

Jacoby, L.L. (1992). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513-541.

Jacoby, L.L. (1997). Invariance in automatic influences of memory: Toward a user's guide for the process-dissociation procedure. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24*, 3-26.

Jacoby, L.L., Yonelinas, A.P. & Jennings, J. (1997). The relation between conscious and unconscious (automatic) influences. A declaration of independence. In J. Cohen & J.W. Schooler (Eds.), *Scientific Approaches to Consciousness* (pp.13-47). Mahweh NJ: Erlbaum.

Jastrow, J. (1906). *The subconscious*. Boston, MA: Houghton-Mifflin.

Jones, E. E., & Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76,* 349-364.

Jordan, C. H., Spencer, S. J., & Zanna, M. P. (2003). "I love me … I love me not": Implicit self-esteem, explicit self-esteem, and defensiveness. In S. J. Spencer, S. Fein, M. P. Zanna, & J. M. Olson (Eds.). *Motivated Social Perception: The Ontario Symposium* (pp. 117-145). Vol. 9. Mahwah, NJ: Lawrence Erlbaum Associates.

Judd, C. M., & Kulik, J. A. (1980). Schematic effects of social attitudes on information processing and recall. *Journal of Personality and Social Psychology, 38*, 569-578.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *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.

Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly, 24,* 163-204.

Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology, 55*, 893-905.

Kawakami, K., & Dovidio, J.F. (2001). Implicit stereotyping: How reliable is it? *Personality and Social Psychology Bulletin, 27*, 212-225.

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.

Klauer, K. C., Musch, J. (2003). Affective priming: Findings and theories. In K. C. Klauer, & J. Musch (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7-50). Mahwah, NJ: Erlbaum.

Klauer, K. C., & Stern, E. (1992). How evaluations guide memory-based judgments: A two-process model. *Journal of Experimental Social Psychology, 28,* 186-206.

Klinger, M. R., Burton, P. C., & Pitts, G. S. (2000). Mechanisms of unconscious priming: I. Response competition, not spreading activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 441-455.

Koole, S.K., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: Implicit self-esteem. *Journal of Personality and Social Psychology, 80*, 614-627.

Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). The measurement of attitudes. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The Handbook of Attitudes* (pp. 21-76). Mahwah, NJ: Erlbaum.

Krosnick, J. A., & Petty, R. E. (1995). Attitude strength: An overview. In R. E. Petty and J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences*. Hillsdale, NJ: Erlbaum.

Kruglanski, A.W. and Klar, Y. (1987). A view from a bridge: Synthesizing the consistency and attribution paradigms from a lay epistemic perspective. *European Journal of Social Psychology, 17*, 211-241.

Kunda, Z. (1999). *Social cognition*. Cambridge, MA, MIT Press.

Lambert, A., J., Payne, B. K., Ramsey, S., & Shaffer, L. M. (2005). On the predictive validity of implicit attitude measures: The moderating effect of perceived group variability. *Journal of Experimental Social Psycholog, 41*, 114-128.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990) Emotion, Attention, and the Startle Reflex. *Psychological Review, 97*, 377-395.

LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces, 13,* 230-237.

Larsen, J. T., McGraw, P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology, 81*, 684-696.

LeDoux, J. (1996). *The emotional brain*. New York: Touchstone.

LeDoux, J.E. (2000) Emotion circuits in the brain. *Annual Review Neuroscience, 23*, 155-184.

Levey, A. B., & Martin, I. (1975). Classical conditioning of human "evaluative" responses. *Behaviour Research and Therapy, 13*, 221-226.

Lewin, K. (1935.) *A dynamic theory of personality*. New York: McGraw-Hill.

Liberzon I, Phan KL, Decker LR, Taylor SF (2003). Extended amygdala and emotional salience: A PET investigation of positive and negative affect. *Neuropsychopharmacology, 28*, 726-733.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, Y. (2002). Reflection and reflexion: A social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology, 34*, 199-249.

Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science, 12*, 135-140.

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.

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology, 12*, 523-553.

Logan, G. D. (1985). Skill and automaticity: Relations, implications, and future directions. *Canadian Journal of Psychology, 39,* 367-386.

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

Maddux, W. W., Barden, J., Brewer, M. B., & Petty, R. E. (2005). Saying no to negativity: The effects of context and motivation to control prejudice on automatic evaluative responses. *Journal of Experimental Social Psychology, 41*, 19-35.

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. & Clore, G. L. (Eds). (2001). *Theories of Mood and Cognition: A User's Guidebook*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Martin, I., & Levey, A.B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy, 1*, 57-102.

Masson, M. E. J. (1991). A distributed memory model of context effects in word identfication. In D. Besner, & G. W. Humphreys (Eds.), *Basic processes in reading* (pp. 233-263). Hillsdale, NJ: Lawrence Erlbaum Associates.

Masson, M. E. J. (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(1)*, 3-23.

McConahay, J. B. (1983). Modern racism and modern discrimination: The effects of race, racial attitudes, and context on simulated hiring decisions. *Personality and Social Psychology Bulletin, 9,* 551-558.

McConahay, J. (1986). Modern racism, ambivalence, and the Modern Racism scale. In J. Dovidio (Ed.), *Prejudice, discrimination, and racism* (pp. 91-125). San Diego, CA: Academic Press.

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

McGuire, W. J. (1969). The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (2nd ed., Vol. 3, pp.136-314). Reading, MA: Addison-Wesley.

McGuire, W. J. (1985). Attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 2, pp. 233-346). New York: Random House.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227-234.

Mierke, J., & Klauer, K. C. (2003). Method-Specific Variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*, 1180-1192.

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General, 132*, 455-469.

Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality and Social Psychology, 85(6)*, 1035-1048.

Moors, A., & De Houwer, J. (2001). Automatic Appraisal of Motivational Valence: Motivational Affective Priming and Simon Effects. *Cognition and Emotion, 15*, 749-766.

Moors, A., De Houwer, J., & Eelen, P. (2004). Automatic stimulus-goal comparisons: Evidence from motivational affective priming studies. *Cognition and Emotion, 18*, 29-54.

Morris, J. S, Ohman, A, Dolan, R. J. 1998. Conscious and unconscious emotional learning in the human amygdala. *Nature, 393*, 467-470.

Moskowitz, G. B. (2005). *Social cognition: Understanding self and others*. New York: The Guilford Press.

Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. *Journal of Personality and Social Psychology, 77*, 167-184.

Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: Affective priming with suboptimal and optimal stimuli. *Journal of Personality and Social Psychology, 64*, 723-739.

Musch, J., & Klauer, K. C. (2003). *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*. Mahwah, NJ: Lawrence Erlbaum.

Myers, D. G. (1993). *Social psychology* (4th ed). Columbus, OH: McGraw-Hill.

Neely, J. H. (1976). Semantic priming and retrieval from lexical memory: Evidence for faciliatory and inhibitory processes. *Memory and Cognition, 4,* 648-654.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 225-254.

Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology, 26*, 505-527.

Niedenthal, P. M., & Kitayama, S. (1994). *The heart's eye: Emotional influences in perception and attention*. New York: Academic Press.

Nisbett, R., & Wilson, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231-259.

Nosek, B. A. (in press). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19(6)*, 625-666.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math=Me, Me=Female, therefore Math is not equal to me. *Journal of Personality and Social Psychology, 83*, 44-59.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

Ohman, A. (1986). Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analysis of emotion. *Psychophysiology, 23*, 123-145.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413-417.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*, 653-667.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge, UK: Cambridge University Press.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Chicago: University of Illinois Press.

Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology, 81*, 181-192.

Payne, B.K., Cheng, C.M., Govorun, O., & Stewart, B. (in press). An inkblot for attitudes: Attitude misattribution as implicit measurement. *Journal of Personality and Social Psychology*.

Payne, B.K., Jacoby, L. L., & Lambert, A. J. (2005). Attitudes as accessibility bias: Dissociating automatic and controlled processes. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 393-420). New York: Oxford University Press.

Petty, R. E., & Krosnick, J. A. (1995). *Attitude strength: Antecedents and consequences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*, 729-738.

Phelps, E.A., O'Connor, K.J., Gatenby, J.C., Grillon, C., Gore, J.C. & Davis, M. (2001). Activation of the human amygdala to a cognitive representation of fear. *Nature Neuroscience, 4*, 437-441.

Pinker, S. (1997). *How the mind works.* New York, NY: W. W. Norton and Co, Inc.

Plant, E.A. & Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology, 75*, 811-832.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 55-85). Hillsdale, NJ: Erlbaum.

Pratkanis, A. R. (1989). The cognitive representation of attitudes. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), *Attitude structure and function* (pp. 71-98). New Jersey: Lawrence Erlbaum.

Pratkanis, A. R., Breckler, S. J., & Greenwald, A. G. (1989). *Attitude structure and function.* New Jersey: Erlbaum.

Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic Information Processing* (pp. 216-270). Cambridge, Massachusetts, MIT Press.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45(9)*, 1043–1056.

Rosenberg, M. J. (1956). Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology, 53,* 367-372.

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.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit stereotypes and prejudice. *Journal of Personality and Social Psychology, 81*, 856-868.

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition, 17,* 437-465.

Salancik, G. R., & Conway, M. (1975). Attitude inferences from salient and relevant cognitive content about behavior. *Journal of Personality and Social Psychology, 32,* 829-840.

Sarnoff, I. (1960). Psychoanalytic Theory and Social Attitudes. *Public Opinion Quarterly, 24*, 251-279.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

Schneider, W., & Fisk, A. D. (1982). Degree of consistent training: Improvements in search performance and automatic process development. *Perception & Psychophysics, 31,* 160-168.

Schuman, H., & Johnson, M. P. (1976). Attitudes and behavior. *Annual Review of Sociology, 2,* 161-207.

Schwarz, N., & Bohner, G. (2001). The construction of attitudes. In A. Tesser & N. Schwarz (Eds.), *Blackwell handbook of social psychology: Intraindividual processes* (Vol.1, pp. 436-457). Oxford, UK: Blackwell.

Sherman, J. W. (in press). Building a better process model: Quality over quantity. *Psychological Inquiry*.

Sherman, J. W., & Payne, B. K. (in press). Separating automatic and controlled components of implicit measures: Process dissociation in social cognition. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new wave of implicit measures*.

Sherman, S.J., Presson, C.C., Chassin, L., Rose, J.S., & Koch, K. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology, 22*, 13-39.

Shiffrin, R. M., & Dumais, S. T. (1981). The development of automatism. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp.111-140). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3-22.

Smith, E. R. (1992). The role of exemplars in social judgment. In L. L. Martin and A. Tesser (Eds.), *The construction of social judgment*. Hillsdale, NJ: Erlbaum.

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. S. Wyer (Ed.), *Advances in social cognition* (Vol. 10, pp. 181-202). Mahway, NJ: Erlbaum.

Smith, E. R. (2000, February). *Connectionist representation of evaluation*. Paper presented at the meeting of the Society for Personality and Social Psychology, Nashville, TN.

Smith, E. R., & DeCoster, J. (1999). Associative and rule-based processing: A connectionist interpretation of dual-process models. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 323-336). New York, NY: Guilford Press.

Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible evaluations influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology, 71*, 888-898.

Smith, E. R., & Lerner, M. (1986). Development of automatism of social judgments. *Journal of Personality and Social Psychology, 50,* 246-259.

Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99,* 3-21.

Smith, M. B., Bruner, J. S., & White, R. W. (1956). *Opinions and personality*. New York: Wiley.

Smolensky, P. (1989). Connectionist modeling: Neural computation/mental connections. In L. Nadel (Ed.), P. Culicover, L. A. Cooper, R. M. Harnish (Assoc. Eds.), *Neural connections, mental computation*. Cambridge, MA: MIT Press/Bradford. 49-67.

Squire, L.R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99*, 195-231.

Squire, L. R., & Kandel, E. R. (1999). *Memory: From mind to molecules*. New York: Scientific American Library.

Srull, T. K. (1981). Person memory: Some tests of associative storage and retrieval models. *Journal of Experimental Psychology: Human Learning and Memory, 7*, 440-463.

Srull, T. K. (1983). The Role of Prior Knowledge in the Acquisition, Retension, and Use of New Information. In R. P. Bagozzi & A. M. Tybout (Eds.), *Advances in Consumer Research* (Vol 10, pp. 572-576). Ann Arbor, MI: Association for Consumer Research.

Stapel, D.A., & Koomen, W. (2000). How far do we go beyond the information given?: The impact of knowledge activation on interpretation and inference. *Journal of Personality and Social Psychology, 78*, 19-37.

Stapel, D. A., Koomen, W., & Ruijs, K. (2002). The effects of diffuse and distinct affect. *Journal of Personality and Social Psychology, 83*, 60-74.

Strack, F., & Deutsch, R. (2004). Reflection and impulse as determinants of "conscious" and "unconscious" motivation. In J. P. Forgas, K. Williams, & S. Laham (Eds.), *Social motivation:Conscious and Unconscious Processes*. Cambridge, UK: Cambridge University Press.

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

Tesser, A. (1978). Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11, pp. 289-338). New York: Academic Press.

Tesser, A., & Martin, L. (1996). The psychology of evaluation. In E. T. Higgins, & A. W. Kruglanski (Eds). *Social psychology: Handbook of basic principles* (pp. 400-432). New York: The Guilford Press.

Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33,* 529-554.

Thurstone, L.L. (1931). Measurement of social attitudes. *Journal of Abnormal and Social Psychology, 26*, 249-269.

Tulving, E. & Craik, F.I.M. (2000). *Handbook of Memory*. Oxford: Oxford University Press.

von Hippel, W., Sekaquaptewa, D., & Vargas, P. (1995). On the role of encoding processes in stereotype maintenance. *Advances in Experimental Social Psychology, 27*, 177-254.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading evaluation effect. *Journal of Personality and Social Psychology, 82*, 919-924.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion, 19*, 175-196.

Wentura, D. (1999). Activation and inhibition of affective information: Evidence for negative priming in the evaluation task. *Cognition and Emotion, 13,* 65-91.

Wentura, D. (2000). Dissociative affective and associative priming effects in the lexical decision task: *Yes* versus *no* responses to word targets reveal evaluative judgmental tendencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 456-469.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology, 78,* 1024-1037.

Wicker, A. W. (1969). Attitude versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues, 25(4),* 41-78.

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In A. Tesser & L. martin (Eds.), *The construction of social judgment* (pp. 37-65). Hillsdale, NJ: Erlbaum.

Wilson, T. D., Lindsey, S., Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107,* 101-126.

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

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.

Zajonc, R.B. (1980). Feeling and thinking. Preferences need no inferences. *American Psychologist, 35*, 151–175.

Zajonc, R. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10*, 224-228.

Zald, D. H. (2003). The human amygdala and the emotional evaluation of sensory stimuli. *Brain Research Reviews, 41*, 88-123.

Zald, D.H. and Pardo, J.V. (1997). Olfaction, emotion and the human amygdala: Amygdalar activation during aversive olfactory stimulation. *Proceedings of the National Academy of Sciences, 94*, 4119-4124.

Zaller, J. (1992). *The Nature and Origins of Mass Opinion*. Cambridge, UK: Cambridge University.