

On the Automatic Evaluation of Social Exemplars

Luigi Castelli and Cristina Zogmaister  
Università di Padova

Eliot R. Smith  
Indiana University

Luciano Arcuri  
Università di Padova

The present article focuses on the automatic evaluation of exemplars whose category membership has been learned in the past. Studies 1 and 2 confirmed the hypothesis that once an exemplar has been encoded as a member of a given group, at a later encounter the evaluation associated with the group will be unintentionally retrieved from memory, even when no perceptual cue indicates the exemplar's category membership. Study 3 extended the results to the domain of in-group/out-group differentiation. In addition, Studies 4 and 5 confirmed the hypothesis that stored evaluations can be retrieved and affect responses even when the semantic information on which the evaluations were originally based is no longer available for retrieval. Finally, Study 6 investigated spontaneous approach-avoidance behavior tendencies. Overall, results demonstrate the pervasive effects of person-based representations, and they are discussed in terms of recent models of person perception and out-group discrimination.

One of the most intriguing aspects of mental life is how people make use of previous experiences to deal with current situations (see Smith, 1990). From a disappointing experience with a particular brand of tuna to a pleasing interaction with a clerk, episodic information about likes and dislikes is constantly accumulated in memory. Such an articulated database may then become a helpful tool to make future decisions and plan behaviors.

Of course, current evaluative responses can be produced through a deliberate and effortful process (see Ajzen, 2000; Wilson & Hodges, 1992), but they can also be automatically evoked (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Indeed, in the past 2 decades, a multitude of experiments have repeatedly shown that individuals automatically evaluate the objects they encounter. The mere presentation of an attitude object seems to be the minimal condition required to instigate an automatic evaluation process (Bargh, Chaiken, Govender, & Pratto, 1992; Bargh, Chaiken, Raymond, & Hymes, 1996; Ito & Cacioppo, 2000; but see Fazio, 1993; Fazio et al., 1986). This is particularly worth noting when

the target is a member of a social group. For instance, when faced with a member of a stigmatized group (e.g., a Black man), the evaluation associated with that specific social group is spontaneously and unintentionally retrieved from memory (Fazio, Jackson, Dunton, & Williams, 1995; Greenwald, McGhee, & Schwartz, 1998). Fiske and Pavelchak (1986) have argued that category schemas are directly connected to evaluations and affective reactions. Hence, the mere presence of an exemplar triggering the schema activates the related evaluation without any need to retrieve all stored information that could be invoked to support such an evaluation, like stereotype-consistent traits, personal experiences, and so on. A general affective tag is encoded as the average of the affective valence associated with the low-level attributes of the schema and is automatically brought to mind by the mere presence of a category exemplar or a symbolic equivalent. This ubiquitous autoevaluation effect is assumed to be one of the most powerful determinants of intergroup behaviors as well as of the ordinary activities in which people continuously engage (Fazio, 2000). Perception, attention, memory, and behavior are all shaped by the initial automatic evaluation of the attitude objects (Niedenthal & Kitayama, 1994).

A closer look at the social-cognitive literature on automatic evaluation, however, indicates that the targets used in the experiments are easily categorizable (e.g., Black and White people). In other words, there are unequivocal perceptual cues that enable people to categorize the targets into a certain social category. As a consequence, implicit attitude measures, such as the Implicit Association Test (IAT; Greenwald et al., 1998) or the affective priming procedure (Fazio et al., 1986; Klauer, 1998), are usually

---

Luigi Castelli, Cristina Zogmaister, and Luciano Arcuri, Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova; Eliot R. Smith, Department of Psychology, Indiana University.

We thank Licia Langella and Lisa Del Gatto for helping to collect the data.

Correspondence concerning this article should be addressed to Luigi Castelli, Dipartimento di Psicologia dello Sviluppo e della Socializzazione, Università di Padova, via Venezia 8, 35131 Padova, Italy. E-mail: castelli.luigi@unipd.it

used to assess the evaluation subjectively associated with social categories. Many times, however, social exemplars may be difficult to categorize on the basis of their physical appearance. For instance, political affiliations, sexual and religious proclivities, or national identity often cannot be inferred by a simple look at the target. In these cases, affective responses can be produced only on the basis of exemplar-based representations stored in memory (Smith & Zárate, 1990; 1992). For instance, if one is warned that a target is a drug pusher, at a later encounter his physical appearance cannot signal his category membership, and one's current reactions must necessarily be based on a person representation stored in memory.

The important issue examined in the present article has to do with the processes involved in managing these person-based representations stored in memory. As said above, these representations may be intentionally retrieved and used (Wilson & Hodges, 1992). However, they may also automatically lead to affective responses. Recent lines of research support this latter claim (De Houwer, Hermans, & Eelen, 1998; Hermans, De Houwer, & Eelen, 1994; see also Dagenbach, Horst, & Carr, 1990; Graf & Schacter, 1985; McKoon & Ratcliff, 1986; Pecher & Raaijmakers, 1999). For instance, De Houwer et al. (1998) showed that non-words that had been associated with affective stimuli automatically primed the congruent affective response. Hence, when the non-word *yalan* is associated with the word *war*, that nonword subsequently primes negative affective responses and inhibits positive affective responses. The experiments adopting this episodic affective priming procedure have demonstrated that a strong stimulus-valence association in memory is not a necessary condition for the emergence of automatic evaluation processes. In this light, De Houwer et al. (1998) concluded that "the newly learned affective meaning of a previously unknown word can be retrieved automatically" (p. 164).

Moving from the linguistic domain to person-based representations, it might be reasonably expected that similar processes regulate the encoding and retrieval of affective information about specific exemplars. Exemplar-based representations may give rise to automatic processes in the same way that abstract knowledge does (Baldwin, Carrell, & Lopez, 1990; Dijksterhuis et al., 1998; Macrae, Bodenhausen, Milne, Castelli, Schloerscheidt, & Greco, 1998; see also Logan, 1988). Macrae et al. (1998), adopting a flanker task, found that the gender categorization of a first name was influenced by an irrelevant surname that was simultaneously presented. For instance, participants were faster in responding that *Mary* was a female name when it was paired with the surname *Thatcher* rather than *Clinton*. Because participants automatically activated the gender associated with the famous exemplar, this facilitated the categorization of gender-congruent names and inhibited the categorization of gender-incongruent names. These experiments have demonstrated that person-based representations can be automatically activated similarly to abstract knowledge. However, it still remains unclear whether the effects are confined to well-known exemplars. Most individuals have been repeatedly exposed to exemplars like Bill Clinton, Margaret Thatcher, and the Pope. Hence, they hold easily accessible representations of those exemplars. What happens when people have been exposed only once, or just a few times, to a given exemplar?

In line with both the episodic priming literature and the recent findings about the automaticity of exemplar activation, in the

present article we examine whether previously categorized exemplars automatically activate congruent affective responses even when no perceptual cue indicates the exemplar's category membership. Paralleling what happens in the case of social schemas (Fiske & Pavelchak, 1986), we expect that once an exemplar is categorized into a given group, the representation of that specific exemplar is tagged with the congruent evaluation. Hence, if we learn that someone is a member of a negative social group—because, for example, we see her or him participating in a neo-Nazi meeting—the tag associated with the group as a whole also becomes associated with the representation of that specific group member. It is expected that at any later encounter with the exemplar, the affective tag will be automatically activated.

In the first two experiments, we tested this hypothesis. In particular, we focused on a specific component of automaticity (Bargh, 1989, 1994), namely the unintentional nature of the process. To this end, we adopted the affective priming procedure (Study 1) and a modified version of the IAT (Study 2). The affective priming procedure is a well-established methodology for the assessment of spontaneous affective reactions (see Klauer, 1998), and it thus probably constitutes the most reliable tool to test the aforementioned hypothesis. In addition, we used a modified version of the IAT to obtain converging evidence about the formulated hypothesis. Despite the ongoing debates about the mechanisms underlying the responses in the IAT, this experimental procedure has proved to be an effective instrument for assessing implicit preferences toward a variety of social groups (see Dasgupta, McGhee, Greenwald, & Banaji, 2000; Greenwald et al., 1998; Ottaway, Hayden, & Oakes, 2001; Rudman, Greenwald, Mellott, & Schwartz, 1999). The IAT is conceptually similar to evaluative priming (see Cunningham, Preacher, & Banaji, 2001; Dasgupta & Greenwald, 2001) and predicts spontaneous nonverbal behaviors (McConnell & Leibold, 2001). Most importantly, recent data have demonstrated that IAT-measured attitudes strongly correlate both with activation of the amygdala measured by functional magnetic resonance imaging and with eyeblink startle responses (Phelps et al., 2000). Because these two psychophysiological measures are considered good indexes of automatic emotional response, the IAT should also capture automatic affective reactions.

## Study 1

The goal of this study was to investigate the possibility that exemplar-based representations lead to automatic evaluation effects. The experiment was structured in two phases. Initially, participants were presented with a series of photographs portraying White men, half of which were labeled as "child molesters" and the other half as "child counselors." Next, these same pictures were used as primes in an evaluative priming task, and they were expected to automatically activate the congruent evaluative response.

## Method

*Participants.* Forty-five female and male psychology students at the University of Padova took part in the experiment in exchange for course credits. Data from 1 participant were discarded because of an extremely high number of errors (over 25% of the trials).

*Material.* We selected 15 photographs (6.3 cm × 7.1 cm) portraying head-and-shoulder images of White male adults. In addition, 30 words (15

positive: *paradise, freedom, honesty, loyalty, fortune, love, peace, friendship, joy, pleasure, trust, liking, hug, kiss, caress*; 15 negative: *devil, bomb, vomit, disgust, poverty, disease, disaster, homicide, cancer, tragedy, pain, agony, violence, threat, poison*) were selected from an analysis of previous studies.<sup>1</sup>

**Procedure.** During the first phase, participants were presented with 10 pictures in a random order. Pictures remained on the screen for 5 s, and the interstimulus interval (ISI) was 1 s. Below each picture, a label indicated the category membership of the person portrayed. Five pictures were presented as child molesters and 5 pictures as child counselors. The association between photographs and labels was balanced between participants. Next, after a couple of minutes, participants went through an evaluative priming task. On each trial, a fixation point (an asterisk) remained visible for 500 ms and was followed by blank screen for 100 ms, which in turn was followed by a prime. As primes we used the 10 pictures presented in the previous phase and 5 new ones. The prime remained on the screen for 400 ms and was followed after 100 ms by a target word. Participants' task was to respond as quickly and accurately as possible whether the target word was positive or negative by pressing one of two labeled keys on the computer keyboard. Overall, participants went through 120 trials (divided in two blocks) so that each of the 15 primes was followed by 4 positive and 4 negative target words. The presentation order of the trials was randomized. Finally, participants were thanked and debriefed.

## Results and Discussion

Overall, 4.1% of the responses have been excluded from the analyses either because they were incorrect (3.2% of the trials) or were longer than 2,500 ms (0.8% of the trials). For each participant, we separately calculated the mean response latencies to positive words when they were preceded either by a picture of a child molester (i.e., negative group) or by a picture of a child counselor (i.e., positive group). Similarly, we calculated two mean response latencies for negative words depending on the valence of the preceding prime. These values were then submitted to a 2 (prime: child molester vs. child counselor)  $\times$  2 (target word: positive vs. negative) analysis of variance (ANOVA), with both factors within participants. Results showed a significant effect of the valence of the target word,  $F(1, 43) = 12.45, p < .001$ ; responses to positive words were significantly faster than responses to negative words ( $M_s = 645$  ms and 676 ms, respectively). More importantly, the interaction was also significant,  $F(1, 43) = 7.4, p < .01$ . As can be seen in Table 1, participants were faster in reacting to positive words when these were preceded by a picture of a child counselor than by a picture of a child molester,

Table 1  
Mean Response Latencies (in Milliseconds) Depending on the Valence of the Target Word and the Type of Prime (Study 1)

Target words	Target picture		
	Child molester	Child counselor	New
Positive			
<i>M</i>	650	639	652
<i>SD</i>	143	136	145
Negative			
<i>M</i>	669	682	679
<i>SD</i>	129	145	129

$t(43) = 1.69, p < .05$ , one-tailed, whereas the opposite was true for negative words,  $t(43) = 2.17, p < .05$ .<sup>2</sup>

These results clearly demonstrate the effects of knowing the category membership of an exemplar on spontaneous affective responses. Paralleling the results obtained by De Houwer et al. (1998), it is here demonstrated that a single exposure to an exemplar labeled either as a child molester or as a child counselor is sufficient to elicit spontaneous affective reactions at later encounters, even though no perceptual information signals the category membership of the target.

Additional exploratory analyses have been carried out including responses after new pictures. As shown in Table 1, the pattern of results suggests that when the valence of the target word is consistent with the valence of the prime, responses are facilitated. In particular, in the case of positive words, responses are slightly faster after a picture of a child counselor rather than a new picture,  $t(43) = 1.58, p = .06$ . In contrast, latencies after new pictures and after incongruent primes are very similar. Overall, it seems that in the present study, facilitation is stronger than inhibition. However, new pictures can hardly be considered a good baseline because they differ from the ones that have been previously seen, not only on the basis of valence but also, by definition, in terms of familiarity. Hence, conclusions on the relative impact of facilitation and inhibition effects must be drawn with caution.

## Study 2

The main aim of this study was to provide converging evidence in support of the results obtained in Study 1. In particular, we used another well-known procedure to assess implicit attitudes, namely the IAT. As in Study 1, the experiment was structured into two phases. In the first phase, participants learned the category membership of a series of exemplars. In the second phase, they went through a modified version of the IAT. More specifically, they were administered two double-categorization tasks, in which clearly valenced words had to be categorized as "positive" or "negative," and photographs had to be categorized as "old" (i.e., seen in the previous phase) or "new." In one of the double-categorization tasks, the same response key had to be used for positive words and old photos; in the other double-categorization task, participants had to use the same response key for negative words and old photos. The dependent variables were the number of errors and the mean latencies of responses for the categorization of the photographs of molesters and counselors in the two tasks. We hypothesized a facilitation in responding (i.e., lower number of errors and faster responses) to the molesters when the old response key was also used to categorize negative words. The opposite was expected in the case of counselors.

There were three main differences from the traditional IAT. First, only two tasks—instead of five—were administered. In

<sup>1</sup> Participants were actually presented with the Italian equivalents of the reported words in all the studies presented here.

<sup>2</sup> Responses to positive and negative words were also compared separately after the positive group prime and after the negative group prime. In both cases, we found that responses were faster for positive words than negative words:  $t(43) = 4.15, p < .001$ , and  $t(43) = 2.07, p < .05$ , respectively. These results, however, are biased by the general slowness in responding to negative targets.

addition, one categorization task was not related to the category membership of the target but resembled a memory task. In other words, participants did not have to retrieve the category membership of the targets seen in the initial phase in order to perform the task. It was simply asked whether the target had already been seen and not what category he belonged to. However, we hypothesized that the valence attached to this category membership would nonetheless influence the task. Finally, the traditional IAT is structured in such a way that all congruent trials are presented in one phase and all incongruent ones in another phase, whereas in this experiment both congruent and incongruent trials were presented in each phase. When the old pictures shared the same key with positive words, the pictures of counselors gave rise to congruent trials, whereas the pictures portraying molesters gave rise to incongruent trials. The opposite was true when negative words and old pictures shared the same response key.

## Method

**Participants.** Sixty female and male introductory psychology students at Purdue University took part in the experiment in exchange for course credits. Data from 6 participants were not recorded because of a technical problem. Data from another participant were discarded because of the high number of errors (over 27% of the trials).

**Material.** We selected 39 photographs (6.3 cm × 7.1 cm) portraying head-and-shoulder images of White male adults. Twelve of these photographs were presented as members of evaluatively connotated groups, 3 as practice material, and the other 24 as foils. In addition, 15 words were selected from the pool reported by Greenwald et al. (1998). Six positive words (i.e., *caress, freedom, health, love, peace, cheer*) and 6 negative words (i.e., *poverty, crash, death, murder, hatred, accident*) were used in the experimental trials. Three words (i.e., *vacation, agony, paradise*) were used in the warm-up trials.

**Procedure.** Participants were initially presented with a series of 12 photographs, one after the other, on the computer monitor. Each picture remained visible for 5 s with an ISI of 500 ms. Five hundred milliseconds after its onset, each photograph was accompanied by a label that indicated the category membership of the person portrayed: "child molester" for 6 photographs and "child counselor" for the other 6 photographs. The association between photographs and labels was balanced between participants, and the order of presentation of the photographs was randomized. The participants' task consisted of looking at the photographs and reading the category labels aloud.

Next, participants went through an intervening lexical decision task that took about 10 min, during which words related neither to child molesters nor child counselors were presented. Afterward, participants were administered a modified version of the IAT. This test consisted of two sequential double-classification tasks. In each of these two tasks, two types of stimuli appeared on the monitor, in spaced order: photographs (without any category label) and words. Participants had to categorize the photographs as being either old or new, and the words as being either positive or negative. Only two response keys were available for the answers: a green key on the left of the keyboard and a blue key on the right of the keyboard. Hence, in one case participants had to press the same response key both for photographs that had been seen in the initial phase (both molesters and counselors) and for negative words, whereas the other response key was used both for new photographs and positive words. In the other case, participants had to press the same response key for old photographs and positive words. The other response key was used both for new photographs and negative words. The order in which participants received these two possible combinations was counterbalanced between subjects. In one task, participants were presented with 6 of the old pictures (3 molesters and 3 counselors) and 12 new pictures. In the other task, participants saw the

remaining 6 old pictures and 12 completely new pictures. In each of these tasks, the old pictures were presented twice, so that participants responded to 24 experimental trials.

After completing the second double-categorization task, participants went through an active debriefing procedure, during which they saw again the 12 photographs they had seen during the learning phase. This time, however, the association between photographs and labels was reversed (i.e., during this debriefing, photos that were presented as child molesters during the first phase were presented as child counselors, and photos previously seen as child counselors were now presented as child molesters). The aim of this debriefing was twofold: (a) to let the participants directly see the meaning of *material balancing* in a psychological experiment and (b) to counteract possible effects of implicit memory traces for the exemplars portrayed in the photographs in the real lives of our participants.

## Results

**Response latencies.** Results are based on the latencies of the correct recognitions (i.e., responses to pictures that participants had seen in the learning phase and accurately categorized as old during the IAT). Response latencies longer than 3,000 ms and shorter than 300 ms were excluded from the analyses. Overall, for each participant, four mean latencies were computed. Two were related to the categorization of child molesters: one for correct answers to photographs of child molesters when the response key was shared with negative words and one for the responses to child molesters when the response key was shared with positive words. Similarly, two mean latencies were computed for child counselors.

A 2 (target: child molester vs. child counselor) × 2 (other meaning of the response key: positive vs. negative) × 2 (order of presentation of the tasks) mixed-model ANOVA was carried out on the four mean latencies. The first two factors were within participants, whereas the second was manipulated between participants. No significant main or interaction effect involving the order of presentation of the two tasks emerged, so we ignore this factor here. Most importantly, the analysis revealed a main effect of the meaning of the response key. In general, participants were faster in categorizing the pictures when the same response key was used to categorize negative words rather than positive words,  $F(1, 51) = 4.29, p < .05$  ( $M_s = 960$  ms vs. 1,031 ms). This main effect, however, was qualified by the expected significant two-way interaction between the category membership of the target (i.e., molester vs. counselor) and the other meaning associated with the response key,  $F(1, 51) = 4.99, p = .03$ . As Figure 1 clearly shows, when the old response key was also used to respond to negative words, categorization was faster for pictures of child molesters than for pictures of child counselors, whereas the opposite was true when the response key was shared with positive words, even though the simple main effects were only marginally significant,  $t(52) = 1.66, p = .103$ , and  $t(52) = 1.89, p = .065$ , respectively. In general, responses to child molesters were faster when the response key was shared with negative than positive words,  $t(52) = 3.35, p < .002$ , whereas responses to child counselors were not affected,  $t(52) < 1$ . However, the significant main effect of the other meaning of the response key undermines the relevance of these latter comparisons.

**Errors.** The overall number of errors was very low (with the exception of the aforementioned participant whose data were excluded from the analyses), with a mean number of 1.7 errors per participant. An ANOVA on errors, identical to the one conducted

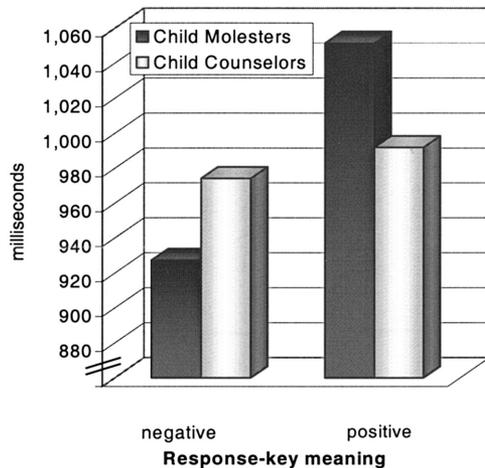


Figure 1. Interaction between category membership of the target and the additional meaning associated with the response key (Study 2).

on the mean latencies, was conducted. This analysis revealed no main or interaction effect of any of the considered factors.

### Discussion

The response latency data clearly demonstrate that participants could not avoid the effects of the targets' original category when asked to provide very simple memory judgments, which required them only to decide whether the target had already been seen. The task in itself was very easy because of the restricted number of pictures presented (i.e., 12), and the low number of errors that participants made supports this conclusion. Nonetheless, responses were deeply affected by the knowledge about who the target was, either a molester or a counselor. This result supports the hypothesis that person-based representations may lead to automatic evaluation effects. In particular, it demonstrates that the affective component associated with each exemplar is activated and influences the responses, even though the task does not require the retrieval of this affective component in order to be performed. Hence, intentionality is not a necessary requirement. In addition, the effects emerge despite the fact that exemplars were not known to participants beforehand. As in Study 1, the association between each exemplar and an evaluation was constructed in the course of the experimental session through a single presentation of each exemplar accompanied by a category label. The demonstration that very limited learning experiences are sufficient for the emergence of automatic evaluation of social targets is consistent with the affective episodic priming literature (De Houwer et al., 1998; Hermans et al., 1994) and indicates that these effects are extremely pervasive and involved in person perception as well as in the processing of linguistic stimuli.

### Study 3

In the previous studies, targets were categorized as members of blatantly negative or positive social groups. In the present study, we investigated whether similar effects emerge when the exemplars are in-group and out-group members. Indeed, despite the actual behaviors that members engage in, individuals tend to have

more positive reactions to members of their own group than to members of the out-group (Fazio et al., 1995; Maass & Schaller, 1991; Perdue, Dovidio, Gurtman, & Tyler, 1990). Perdue et al. (1990, Studies 2 and 3) demonstrated that the pronouns used to designate the in-group activate positive responses, whereas pronouns used to designate the out-group activate neither positive nor negative responses. The in-group appears to be strongly associated with positive reactions. Recent research has demonstrated that this tendency emerges also in minimal group settings. Otten and Moskowitz (1999) and Otten and Wentura (1999) showed that participants spontaneously considered their minimal in-group in a positive way and differentiated it from the out-group.

If the in-group is so consistently preferred over the out-group, it is also expected that individuals who are presented as in-group members will automatically activate more positive reactions at later encounters. Perdue et al. (1990, Study 1) provided evidence that supports this hypothesis. In their experiment, nonsense syllables paired with in-group designators were preferred over nonsense syllables paired with out-group designators. It appears that newly encountered stimuli associated with the in-group receive more positive evaluations. However, judgments were collected on semantic differential scales, and no conclusion can be drawn about the automaticity of the process. In our third experiment, we tested the hypothesis that individuals known to be members of one's own minimal group are automatically evaluated more positively in comparison with out-group members.

### Method

**Participants.** Fifty-eight female and male psychology students at the University of Padova took part in the experiment in exchange for course credits. Data from 1 participant were not recorded because of a computer failure.

**Procedure and material.** Participants initially performed a perceptual task in order to allegedly establish whether they had a "Type A" style of processing or a "Type B" style of processing. Participants were quickly presented with pairs of polygons, and they were asked to respond whether the two polygons were the same or not. At the end, the computer program processed the responses and communicated to each participant that he or she belonged to the group of people characterized by a Type A style of processing. Next, participants were shown the pictures of 10 males. Five of them were presented as belonging to the Type A group, whereas the other 5 were presented as belonging to the Type B group. The procedure closely resembled the one adopted in Study 1, and the material was also the same. The only difference was in the category labels that were used. The subsequent evaluative priming phase was also identical to the one used in Study 1.

### Results and Discussion

Overall, 2.9% of the responses have been excluded from the analysis, either because they were incorrect (2.2% of the trials) or were longer than 2,500 ms (0.7% of the trials).

As in Study 1, we calculated four means for each participant: the mean response latency to positive words after an in-group exemplar, the mean response latency to positive words after an out-group exemplar, and two means for negative words. These values were then submitted to a 2 (prime: in-group vs. out-group)  $\times$  2 (target word: positive vs. negative) ANOVA, with both factors within participants. Results showed a significant effect of the valence of the target word,  $F(1, 56) = 4.01, p = .05$ . In general,

responses to positive words were significantly faster than responses to negative words ( $M_s = 623$  ms and 638 ms, respectively). This effect, however, was qualified by a significant interaction,  $F(1, 56) = 8.92, p < .005$  (see Table 2). As expected, responses to positive words were faster after a picture of an in-group member than after a picture of an out-group member,  $t(56) = 1.81, p < .05$ , one-tailed, whereas the opposite was true for negative words,  $t(56) = 1.89, p < .05$ , one-tailed.<sup>3</sup>

As in Study 1, we also performed exploratory analyses including responses after new pictures. In the case of positive words, responses were somewhat faster after a congruent prime than after the new pictures,  $t(56) = 1.7, p = .087$  ( $M_s = 618$  vs. 627 ms), whereas there was no sign of inhibition ( $M_s = 628$  vs. 627 ms). In the case of negative words, latencies after new pictures did not differ either from latencies after a congruent or after an incongruent prime (all  $p_s > .44$ ). Hence, only in one case did data suggest the presence of facilitation rather than inhibition. As discussed before, however, responses after new primes do not provide an optimal baseline to test the strength of facilitation and inhibition effects.

Overall, the results from the present study demonstrate that after a single exposure to exemplars that are presented either as in-group members or as out-group members, an affective response is spontaneously activated at later encounters; this evaluative response is more positive in the case of in-group members and more negative in the case of out-group members. Of course, there was no previous history of differentiation between the groups, because they were constructed in the course of the experiment and were also based on very limited information (i.e., a difference in perceptual processing style). Nonetheless, this was sufficient to elicit differential evaluative responses. These results are in line with the findings reported by Otten and Wentura (1999) and extend them by demonstrating that in minimal group settings, automatic evaluations occur not only in response to group labels but also to the actual members of the groups.

#### Study 4

In the three previous experiments, we investigated the automatic retrieval of affective responses associated with exemplar-based representations of individuals that are not well known. However, it should be noted that especially when there has been minimal learning, it may be difficult to consciously retrieve a person-based representation. Many aspects of what has been learned, such as an exemplar's group membership, are likely to become progressively

unavailable for conscious retrieval over time. For instance, after a couple of weeks, one may have forgotten the reason why the interaction with a given person was particularly unpleasant, whether because he tripped over us or because he made rude remarks. Nonetheless, at a later encounter with the same person, a negative affective response may still be present and lead to avoidance of further interactions. This is the second crucial issue to be examined in the present article.

This hypothesis is sustained by various theoretical approaches. First, research on implicit memory (Jacoby, Kelley, Brown, & Jasechko, 1989; Schacter, 1987) and on source-monitoring processes (M. K. Johnson, Hashtroudi, & Lindsay, 1993) has demonstrated that current feelings may exist without an awareness of the actual explanations of such feelings. Just as another person may be perceived as familiar without one knowing whether this is due to fame or to a trivial previous encounter (Banaji & Greenwald, 1995; Jacoby, Kelley, et al., 1989), another person may be liked or disliked without one knowing the exact reason for this feeling. Similarly, it has been shown that memory and judgments often do not correlate (Hamilton & Sherman, 1996; Hastie & Park, 1986). What could be retrieved from memory does not necessarily go hand in hand with evaluative judgments. As a matter of fact, people use factual information (e.g., "He makes fun of disabled people") to build up an impression (e.g., "I don't really like that guy"), but this global evaluation may later gain its own independence and become unrelated to the actual information that led to such a judgment. Finally, Niedenthal, Halberstadt, and Innes-Ker (1999) have recently argued that individuals categorize objects, events, and persons not only on the basis of semantic criteria but also on the basis of the affective responses that they are able to evoke. Hence, on the basis of these affective categorizations, negative information on the one side and positive information on the other should be clustered together. This implies that it will later be easier to make within-valence errors rather than between-valence errors. That is, it will be possible to identify the superordinate cluster to which an instance belongs (i.e., positive or negative category) without being able to recall why it has been included in that cluster (i.e., the specific semantic category). All these different approaches lead to a specific prediction: Exemplars that have been categorized as members of a positive or negative social group will later be favored or discriminated against, even when the original category membership is no longer recalled.

In Study 4, we sought to test whether the evaluative reaction toward an exemplar may still be present even when the actual exemplar's category membership is not further recalled. To this end, we used a procedure in which the effects of conscious memory are expected to counteract the effects of unconscious influences (see Jacoby, Yonelinas, & Jennings, 1997). Initially, participants learned the category membership of a series of exemplars. The exemplars belonged to one of three different social categories: two positive and one negative, or one positive and two negative. Later, participants were asked to identify only the mem-

Table 2  
*Mean Response Latencies (in Milliseconds) Depending on the Valence of the Target Word and the Type of Prime (Study 3)*

Target words	Target picture		
	In-group member	Out-group member	New
Positive			
<i>M</i>	618	628	627
<i>SD</i>	104	110	104
Negative			
<i>M</i>	642	633	636
<i>SD</i>	112	103	105

<sup>3</sup> After pictures of in-group members, responses were faster for positive than negative words,  $t(56) = 3.09, p < .005$ , whereas no difference emerged after pictures of out-group members ( $p > .50$ ). Again, as in Study 1, note that responses to negative words were generally slower than responses to positive words.

bers of one category. We hypothesized that there would be a higher number of intrusions from the evaluatively congruent group than from the evaluatively incongruent group. For instance, after being presented with child molesters, drug pushers, and child counselors and later asked to identify the child molesters, we predicted that participants would be more likely to mistakenly include in the target group members of the group of drug pushers rather than child counselors. The benefit of this experimental procedure is that we can contrast the effect of explicit memory with the implicit evaluative effects. Indeed, in the above example, if participants correctly recalled the category membership of the drug pusher, this would have led them to correctly reject the inclusion into the group of child molesters. In contrast, if participants had a negative feeling toward the target without being able to recall his actual category membership, this would have led to an incorrect inclusion. The effects of the correct recall of the category membership of a certain target and the implicit effects of the evaluation associated with this same target are thus posed in opposition (see Jacoby et al., 1997; Jacoby, Kelley, et al., 1989; Jacoby, Woloshyn, & Kelley, 1989).

### Method

*Participants.* Seventy-two students at the University of Padova took part in the experiment on a voluntary basis.

*Procedure.* In the first experimental phase, participants were presented with a series of 40 pictures. Each of these pictures portrayed a male face. Twenty exemplars were said to be “child molesters” and the other twenty were presented as “child counselors.” Participants’ task was simply to look carefully at each picture and read the category label indicating the group to which the person belonged. Each picture remained on the screen for 3 s, and the presentation order was randomized. The association between each set of faces and each social group was counterbalanced between subjects.

Next, participants were presented with 20 members of a third group. For half of the participants it was a positive social group, either “firemen” or “missionaries”; for the other half of the participants it was a negative social group, either “drug addicted persons” or “pushers.” Once again, participants had to look carefully at the pictures and read the category labels.

After a 10-min interval, participants were presented with all 60 pictures, one at a time, in a randomized order. This time, however, there was no category label, and participants’ task was to decide whether the person was a member of the third group that had previously been presented to them. Responses were provided by pressing one of two labeled keys. Participants were given as much time as they needed to answer. We stressed the importance of correct responses and that we were not interested in fast responses. In so doing, we were confident that errors were actually due to a lack of explicit recollection. However, response latencies were recorded even though participants were not aware of this. Finally, participants were thanked and fully debriefed.

*Dependent measure and hypothesis.* The main dependent variable of this study is the proportion of members of the two groups seen in the first experimental phase that are later erroneously included in the third group. The hypothesis is that errors will be more likely between two evaluatively congruent groups (e.g., from child molesters to pushers) rather than between two evaluatively incongruent groups (e.g., from child molesters to firemen). Paralleling Jacoby and colleagues’ procedure (Jacoby, Kelley, et al., 1989; Jacoby, Woloshyn, & Kelley, 1989), the underlying rationale is that although a correct recollection should lead to avoiding the inclusion, when recollection is at least partially impaired, the implicit effects of the affective component should bias responses. The effects of explicit and implicit memory are thus posed in opposition.

### Results

*Recognition data.* Overall, correct inclusions proved to be higher than erroneous inclusions,  $t(71) = 12.58, p < .001$ . Moreover, correct recognition rate was equal for each target group,  $F(3, 71) < 1$ : drug abusers, .54; pushers, .56; firemen, .57; missionaries, .53. An identical pattern emerged when the  $d'$  and beta scores from signal detection analysis were considered.<sup>4</sup>

*Miscategorizations.* A 2 (original group membership)  $\times$  2 (target group valence) mixed-model ANOVA was carried out on the proportion of exemplars erroneously included in the target group. No significant main effect emerged (both  $ps > .5$ ). As expected, however, the interaction was significant,  $F(1, 70) = 8.26, p = .005$  (see Figure 2). The analysis of the simple effects shows that participants were more likely to erroneously include members of a positive social group rather than members of a negative social group in another positive social group,  $t(35) = 2.18, p < .04$ . The opposite, even though the effect is only marginally significant, turned out to be true in the case of negative social groups,  $t(35) = 1.87, p = .07$ . That is, participants were more likely to erroneously include a member of a negative social group rather than a member of a positive social group in a negative target group.<sup>5</sup>

Further analyses were performed to test whether the specific target group had any effect. Two separate ANOVAs were carried out, isolating the conditions in which the target group was either positive or negative and entering the specific target group (i.e., missionaries vs. firemen and drug abusers vs. pushers) as a factor. The target group did not lead to any main effect or interaction (all  $Fs < 1$ ), suggesting that the obtained effects were not limited to one specific target group.

*Response latencies.* Response latencies were also analyzed. Correct recognitions proved faster for exemplars of positive than negative groups (2,601 ms vs. 3,303 ms);  $F(1, 70) = 4.25, p < .05$ . This result is in line with the literature suggesting that negative social information receives greater attention and is processed more carefully than positive social information (see Fiske, 1980; Pratto & John, 1991). Response latencies for wrong inclusions were not affected by the original group membership and the target group or by their interaction. Paralleling the results for correct recognitions, correct exclusions were affected by the valence associated with the original group membership. Responses were indeed faster for exemplars belonging to a positive social group than to a negative

<sup>4</sup> A signal detection analysis was conducted on the recognition data, and the  $d'$  and beta parameters were submitted to ANOVAs to investigate whether criteria for recognition of the target changed depending on the target’s membership (i.e., firemen, missionaries, drug addicted persons, pushers). Results of these analyses showed that neither  $d'$  nor beta differed significantly depending on the category of the target:  $F(3, 68) = 1.11, p > .35$ ;  $F(3, 68) = 0.31, p > .80$ , respectively.

<sup>5</sup> A different way to approach the data is to verify whether actual members of a positive group are more likely erroneously included in a positive rather than negative target group, and members of a negative group are more likely included in a negative rather than positive target group. In both cases, data are in the expected direction, but the effect reached the conventional significance level only in the case of members of a negative group,  $t(70) = 2.02, p < .05$ .

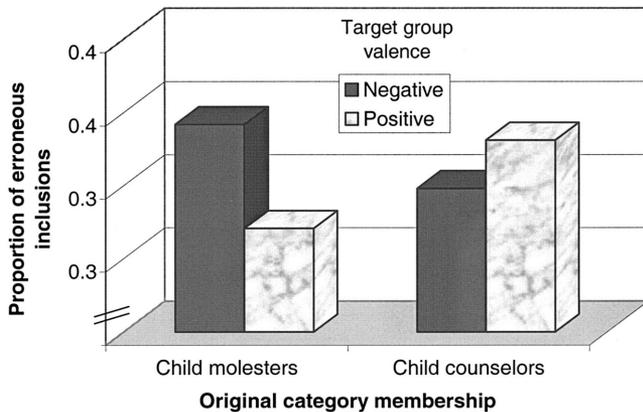


Figure 2. Miscategorizations as a function of the valence of the group target and the valence of the original group to which the exemplars belonged (Study 4).

social group (2,635 ms vs. 3,158 ms);  $F(1, 70) = 4.68, p < .05$ . No other effect was significant.

### Discussion

The pattern of memory errors demonstrates that participants relied on their affective feelings when trying to retrieve the actual category membership of a target. When the category membership was no longer recalled, participants were likely to base their responses on a mere affective feeling in order to guess the correct category membership. Hence, original members of a negative group were more likely than members of a positive group to be included in another negative group. The opposite was true for members of positive groups. These results are in line with the frequently reported finding that judgments (i.e., summary evaluations) do not always correlate with the average evaluation of the single pieces of information that can be retrieved from memory. The affective tag that is formed when receiving the information—built up as a summary evaluation—may gain its own independence and be accessed even if the actual information can no longer be retrieved. Whereas previous studies supported this argument by showing nonsignificant correlations (i.e., accepting the null hypothesis), in the present study the obtained results demonstrate that the research hypothesis of an independence between evaluation and memory holds true (see also Yazuv, 1963). Similarly, these data are strongly consistent with Niedenthal and colleagues' view about categorization processes (see Niedenthal et al., 1999; Niedenthal & Halberstadt, 2000; see also Bruner, Goodnow, & Austin, 1956). According to this view, information is organized in memory not only on the basis of semantic criteria but also on the basis of the affective responses that the information is able to evoke, and hence, confusions between evaluatively congruent items arise, as we found.

### Study 5

Results from Study 3 demonstrate that in-group members spontaneously lead to more positive evaluative responses in comparison with out-group members. Hence, we could predict that in the recognition memory paradigm adopted in Study 4, in-group mem-

bers that are no longer remembered should be more easily included in a positive social group than members of an out-group. This hypothesis was tested in the present experiment.

### Method

**Participants.** Overall, 333 female and male psychology students attending two separate 1st-year courses at the University of Padova took part in the study. One class of students ( $n = 165$ ) was assigned to one condition and the other class ( $n = 168$ ) to the other condition.

**Procedure.** Participants were initially presented with a series of 28 pictures portraying White women, one after the other. Half of the pictures were said to be former psychology students at the University of Padova, whereas the other half were said to be former psychology students at the University of Rome. Each picture was projected on a large screen and remained visible for 5 s (ISI = 1 s), and the label "Padova" or "Rome" placed just below the picture indicated where the person portrayed had received her degree in psychology. The assignment of each picture to the two groups was randomly determined.

Next, participants were presented with an additional 16 pictures portraying White women. In one class, it was said that those women had been charged with abuses against children (i.e., were child molesters). In the other class, it was said that those women had been working to help poor people in the Third World (i.e., were missionaries). Again, these pictures remained visible for 5 s (ISI = 1 s). Finally, after about 5 min, we presented all 44 pictures again. This time, however, there was no category label, and participants were asked to identify all the child molesters or the missionaries according to the condition. At the end, participants were fully debriefed, and in the course of a subsequent class they were presented with the results of this study.

### Results and Discussion

The main dependent variables of this study were the number of in-group members and out-group members who were erroneously included in the target group. Hence, we performed a 2 (original group membership: Padova vs. Rome)  $\times$  2 (target group: child molesters vs. missionaries) ANOVA, with the first factor manipulated within participants and the second between participants. Results showed a highly significant effect of the original group membership,  $F(1, 331) = 94.77, p < .001$ , showing that errors were more likely in the case of exemplars that were originally presented as members of the out-group. Differently stated, this result suggests that information about in-group members is better encoded in memory (see Rogers, Kuiper, & Kircher, 1977). This main effect was qualified by the predicted two-way interaction,  $F(1, 331) = 20.51, p < .001$ . As can be seen in Figure 3, in-group members were more likely included in the positive target group than in the negative target group,  $F(1, 332) = 17.57, p < .001$ . In contrast, in the case of out-group members the effect was nonsignificant, and, if anything, out-group members were included in the negative social group more than in the positive one.<sup>6</sup> This result is consistent with previous literature suggesting that in-group favoritism is far more prevalent than out-group derogation (Brewer,

<sup>6</sup> Comparisons within the two experimental conditions showed that both in the case of a positive and a negative target group, out-group members rather than in-group members were more likely included,  $t(167) = 3.64, p < .001$ , and  $t(167) = 10.20, p < .001$ , respectively. However, these comparisons are strongly biased by the better memory for in-group members, which decreases their incorrect inclusions in any target group.

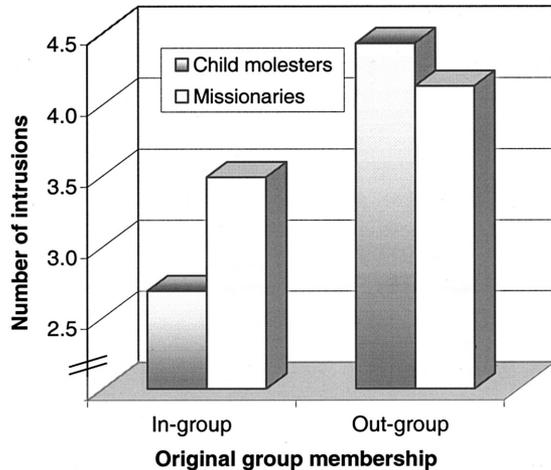


Figure 3. Miscategorizations as a function of the group target (i.e., child molesters vs. missionaries) and the group membership of the exemplars (i.e., in-group vs. out-group) in Study 5.

1979, 1999; Perdue et al., 1990). In addition, this specific out-group shared an important feature with the in-group, namely, the fact that they both were psychology students, and this might have reduced discrimination (see Gaertner & Dovidio, 2000). The effect of the target group was not significant.

These results provide converging evidence for the operation of evaluative responses toward social exemplars even though their actual category membership is not correctly recalled. Applying the logic of an opposition paradigm, correct recognition for in-group and out-group members would lead to a correct response of exclusion, whereas the reliance on affective reactions selectively induces participants to include actual in-group members in a positive group.

### Study 6

Thus far, we have examined the effects of learning the category membership of an exemplar on spontaneous affective responses (Studies 1–3) and on reconstructive memory processes (Studies 4 and 5). Affective information is also highly relevant for the production of appropriate behaviors. Indeed, it would be of little help to produce an automatic affective response when seeing a dangerous insect if we did not use this affective reaction to guide adaptive behavior. In the present study, we tested the hypothesis that at a very basic level, exemplars that are known to belong to positive social groups will be associated with an approach tendency, whereas exemplars that are known to belong to negative social groups will be associated with an avoidance tendency. To this aim, we used a procedure developed in our laboratory (see Castelli & Paladino, 2003). The procedure is based on the fact that in order to provide some responses, participants have to move their arm forward (i.e., approach movement), whereas they have to move the arm backward in order to provide other types of responses (i.e., avoidance movement). For instance, in one study it was shown that Italian participants were faster in categorizing Italian names when moving forward, whereas the categorization of Slavic names was faster when responses were given by moving backward. The task

was framed in such a way that participants did not realize that one movement was actually an approach behavior and the other movement an avoidance behavior. The results from the aforementioned study have been consistently replicated with a variety of different social groups, and the emerging picture shows that people tend to spontaneously approach what is liked and to spontaneously avoid what is disliked (see also Vaes, Paladino, Castelli, Leyens, & Giovanazzi, 2003). In the present study, participants were shown photographs of members of a positive and a negative social group. Next, they were asked to categorize pictures as already seen (“old”) or “new” (see Study 2). In one block, the old responses were provided by moving the arm forward, and it was predicted that in this case the responses to members of the positive group would be facilitated. In another block of trials, the old responses were provided by moving the arm backward, and it was predicted that this movement would facilitate the responses to members of the negative group.

### Method

**Participants.** Forty-five students at the University of Padova participated in the experiment as a partial course requirement.

**Procedure.** Participants were initially presented with a series of 12 pictures portraying White male faces. Each picture remained on the screen for 4 s (ISI = 1,500 ms) and was accompanied by a category label indicating the group membership of the person portrayed. Six targets were presented as child molesters and 6 as child counselors. The association between photographs and labels was counterbalanced between participants. Next, participants were required to perform a memory task. They were asked to discriminate the pictures seen in the initial phase from new pictures of White males. Responses were provided through a modified standard computer keyboard in which only three big buttons were present. The keyboard was placed at the right or at left of the participants according to their dominant hand, and it was rotated 90° (see Figure 4). The task was structured into two blocks. In one block, participants were instructed to remain with their hand on the middle key and, as quickly as possible, to

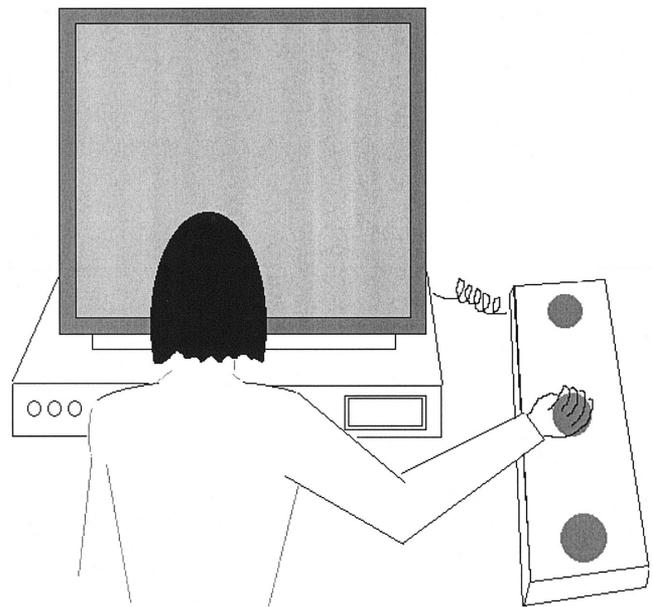


Figure 4. Schematic example of the devices used in Study 6.

press the forward key any time an old picture appeared and to press the backward key any time a new picture appeared. After each response, participants had to again place their hand on the middle key. There were 36 trials (3 child counselors, 3 child molesters, and 6 new pictures, each presented three times) that were followed by a short break. In the other block, participants performed the same task, but this time they had to press the forward key for new pictures and the backward key for old pictures. Again, participants went through 36 trials. The relative order of the two blocks was counterbalanced between participants. Finally, participants were thanked and debriefed.

### Results and Discussion

We considered latencies of responses to old targets. In these trials, error rate was quite high (16.4%), and latencies of wrong answers were not included in the analyses. In addition, 0.5% of the trials were classified as outliers (either longer than 4,000 ms or shorter than 250 ms) and were excluded from the analyses.

The response latencies of correct classifications of targets as already seen in the initial phase were submitted to a 2 (category membership: child molester vs. child counselor)  $\times$  2 (movement: forward vs. backward) ANOVA. A significant main effect of the target's category membership emerged,  $F(1, 44) = 5.57, p < .05$ , showing that responses to members of the negative social group were slower than responses to members of the positive social group ( $M_s = 1,247$  ms and 1,203 ms, respectively). In addition, in accordance with the hypothesis, a significant interaction emerged,  $F(1, 44) = 4.34, p < .05$  (see Table 3). Indeed, when having to move forward (i.e., approach), participants were faster in the case of the members of the positive group,  $t(44) = 3.06, p < .005$ , whereas there was no difference when moving backward (i.e., avoidance;  $p > .6$ ). Comparisons made keeping the category membership of the target constant showed that in the case of child molesters, avoidance movements were slightly faster than approach movements (1,238 ms vs. 1,256 ms), whereas in the case of child counselors, approach movements tended to be performed faster than avoidance movements,  $t(44) = 1.7, p < .05$ , one-tailed (1,180 ms vs. 1,227 ms). Hence, even though participants were not required to report the category membership of the target, such category membership automatically influenced motor responses. More specifically, it appears that exemplars that belong either to positive or negative social groups spontaneously lead to differential behavior tendencies. It was thus demonstrated that behaviors as well as affective responses are automatically influenced by person-based representations stored in memory. It must be noted, however, that in the present experiment, the effect seemed to be mainly driven by a tendency to approach members of the positive group. It is likely that stronger avoidance tendencies may be found in the case of "dangerous" group members (e.g., murderers, meningitis-infected persons), that is, individuals with whom interaction can be perceived as potentially harmful (see Wentura, Rothermund, & Bak, 2000). Future research will have to examine whether a spontaneous tendency to avoid members of negative and stigmatized groups emerges in such cases.

### Meta-Analysis of the Six Studies

In the six studies presented in this article, we used four different paradigms (evaluative priming, a modified version of the IAT, recognition memory, and an approach-avoidance task) to measure

Table 3  
*Mean Response Latencies (in Milliseconds) According to the Category Membership of the Target and the Movement Performed (Study 6)*

Movement	Child molester	Child counselor
Approach		
<i>M</i>	1,256	1,180
<i>SD</i>	255	235
Avoidance		
<i>M</i>	1,238	1,227
<i>SD</i>	259	217

the impact of previous experiences with members of valenced groups (clearly positive vs. negative group members; in-group vs. out-group members). As expected, all these studies showed that people react to members of positive groups in a different way than to members of negative ones, even without an extensive learning experience with these exemplars. We conducted a meta-analysis to summarize our results. Computations were performed using B. T. Johnson's (1989) DSTAT computer program, which applies the approach proposed by Hedges and Olkin (1985). As a first step, for each experiment, two separate effect sizes were calculated: One of them was related to positive exemplars (or in-group members), and the other was related to negative exemplars (or out-group members). Each effect size was calculated as the standardized difference between mean responses for incongruent and congruent trials (i.e., higher expected values minus lower expected values). Effect sizes were computed as Cohen's  $d$ , where a positive effect size indicates a hypothesis-consistent result, and a negative effect size indicates a hypothesis-inconsistent result. Further, for each experiment, we also computed a summary effect size applying the same logic described above. In Table 4, these three effect sizes are reported for each study.

Finally, the overall effect size (Cohen's  $d$ ) of the meta-analysis of the six presented studies was calculated using the unweighted integration method because of the large differences in the number of participants across the studies. The value of this index is .25, with a 95% confidence interval ranging from .17 to .33. The magnitude of the effect size, though, differs from one experiment to the other, ranging from .09 to .35, as can be seen in Table 4. Furthermore, in some experiments, positive exemplars seem to have a stronger impact than negative ones, whereas in other experiments the contrary seems to be true. The overall effect size for positive exemplars is .33 (95% confidence interval from .22 to .44), whereas for negative exemplars it is .17 (95% confidence interval from .06 to .28). Both these effect sizes are statistically significant ( $p < .00001$  for positive groups, and  $p < .005$  for negative groups). The difference in the magnitude of the effect size for positive and negative exemplars seems to be at least partially due to the presence of significant main effects in many of the reported studies. Hence, we believe that no clear conclusion can be drawn from the differences in the overall effect sizes for positive and negative exemplars. The safest conclusion seems to be that using different methodologies and target groups, we found small (according to Cohen's standards; Cohen, 1988) but reliable effects of group membership on automatic affective reactions for both positive and negative group members.

Table 4  
*Summary of the Results From the Six Studies*

Study	Dependent measure	Type of groups	Main effects	Effect size		Summary
				Positive exemplars and in-group members	Negative exemplars and out-group members	
1	RT in evaluative priming	Clearly valenced groups	Faster reactions to positive words	.31	-.14	.09
2	RT in IAT	Clearly valenced groups	Faster reactions with "negative meaning" key	-.06	.44	.19
3	RT in evaluative priming	Minimal groups	Faster reactions to positive words	.22	-.05	.09
4	Memory misattributions	Clearly valenced groups	None	.23	.48	.36
5	Memory misattributions	In-group/out-group	Better memory for in-group exemplars	.46	.16	.30
6	RT in approach-avoidance task	Clearly valenced groups	Faster reactions to members of the positive group	.21	.07	.14
Overall effect size				.33	.17	.25

*Note.* A positive effect size represents a hypothesis-consistent result, and a negative effect size represents a hypothesis-inconsistent result. RT = reaction time; IAT = Implicit Association Test.

### General Discussion

In line with an increasing number of studies, the present work further demonstrates the pervasive effects of person-based representations on information processing (Andersen & Cole, 1990; Bodenhausen, Schwarz, Bless, & Wänke, 1995; Castelli & Zogmaister, 2000; Castelli, Zogmaister, & Arcuri, 2001; Lewicki, 1985; Macrae, Bodenhausen, Milne, & Castelli, 1999; Macrae et al., 1998; Smith, 1992; Smith & Zárate, 1990, 1992; Stapel, Koomen, & van der Plight, 1997). In particular, our data corroborate the findings reported by Macrae et al. (1998) about the unintentional nature of exemplar activation. In their experiments, it was shown that the gender information associated with an exemplar-based representation is automatically activated on encountering a cue that might trigger the process (i.e., the surname). In the present article, we argue that similar effects emerge in the case of affective information associated with an exemplar-based representation. Hence, autoevaluation phenomena (see Bargh et al., 1992, 1996; Fazio, 1986) also occur in response to specific exemplars on the basis of what is stored in memory. Most importantly, it is demonstrated that a single exposure to an exemplar is a sufficient condition for this process to occur. Differently stated, the spontaneous retrieval of evaluations occurs not only in the case of famous exemplars such as Mike Tyson or the Pope, but they also emerge for briefly encountered exemplars. As first demonstrated by De Houwer et al. (1998) through the episodic affective priming technique, stimulus-valence associations are quickly developed, and the evaluation can later be automatically accessed at any encounter with the target stimulus.

We believe that the automatic activation of affective responses even for newly learned exemplars is a highly adaptive process. Indeed, there is widespread agreement that automatic evaluation processes facilitate fast behavioral responses that lead to approaching positive target stimuli and avoiding negative target stimuli (Chen & Bargh, 1999; Wentura et al., 2000). In this light, it is important to automatically retrieve the valence of well-known stimuli, such as snakes and potentially harmful out-groups, but also the valence associated with newly learned stimuli. For in-

stance, if one is warned that an apparently innocuous butterfly is actually able to inject a deadly poison, it is critical to automatically avoid it in the future, even though the strength (i.e., accessibility based on the frequency of access) of the association between the stimulus and the valence is weak. Otherwise, one might be unable to develop any future strong association.

Of course, these arguments directly point to the debate on the prerequisites of autoevaluation effects. Fazio (1986; Fazio et al., 1986) claimed that a strong stimulus-valence association in memory is a necessary condition, whereas Bargh et al. (1992, 1996) considered the process as occurring both in strong and weak stimulus-valence associations. Evidence of automatic evaluation in response to newly learned exemplars has supported Bargh et al.'s view (see De Houwer et al., 1998). However, we must stress that the categories used in most of our studies were extremely positive and negative (e.g., child counselors and child molesters). De Houwer et al. (1998) also used highly polarized words as conditioning stimuli. Hence, at this time, we do not know whether the effects are limited to extreme categories or occur also with weakly positive and negative categories. The first hypothesis is more consistent with Fazio's model, whereas the second hypothesis is more consistent with Bargh et al.'s model. Future research will be needed to address this issue, even though the results from the minimal group study suggest that the effects are extremely pervasive, and an intergroup differentiation based on minimal information is sufficient to lead to autoevaluation phenomena.

The demonstrated spontaneous affective response toward specific instances has important consequences for social perception. For instance, perceivers encountering known members of negative social groups will activate negative responses even when no perceptual cue indicates their category membership—such as when these individuals try to obscure their category membership. This means that if an individual possesses a stigma (Crocker, Major, & Steele, 1998), such as belonging to a negative group, even if he or she tries to conceal it, the perceiver will nonetheless activate the associated affective response. At any encounter with the stigmatized individual, neither perceptually evident cues about the stigma

(e.g., the color of the skin) nor well-established memory traces are necessary for the emergence of a spontaneous evaluation. Occasional associations between the target and a stigma are sufficient for subsequent spontaneous evaluation effects. This occurs also for members of an out-group defined according to minimal criteria. Individuals who are known to be similar to us and to belong to our own group are automatically preferred over individuals who differ from us because they belong to a different social group.

### *Relation to the Affective Conditioning Literature*

Of course, the literature on evaluative conditioning is highly relevant for the present studies (Niedenthal, 1990; for an excellent review, see De Houwer, Thomas, & Baeyens, 2001). In our studies, as in evaluative conditioning research, a neutral item (i.e., an unknown face; conditioned stimulus [CS]) was paired with an affective stimulus (i.e., a category label; unconditioned stimulus [US]), so that the CS acquired the evaluative connotation of the US. However, important differences between these paradigms should also be underscored. In most evaluative conditioning research, the CS and US are simply presented in temporal and spatial contiguity, whereas in the present studies the category label defined who the unknown person was. Strictly speaking, all impression formation tasks may be conceptualized in terms of conditioning, but this would be limiting at best. Indeed, category labels, group memberships, or personality traits are perceived as shaping the very essence of the target (Rothbart & Taylor, 1992; Yzerbyt, Rocher, & Schadron, 1997). This difference has meaningful consequences. For instance, it has been shown that counterconditioning occurs when a CS that had been initially paired with an unpleasant US was subsequently paired with a pleasant US, or vice versa (Baeyens, Eelen, Van den Bergh, & Crombez, 1989). A triangle paired with guns and blood can lose its negative connotation if it is subsequently paired with flowers, smiling faces, and nice cars. In contrast, we suspect that child molesters paired with flowers, smiling faces, and nice cars will nonetheless continue to evoke negative reactions. Similarly, within the area of person perception, a reversal in evaluation has been shown for conditioned faces that were later described with oppositely evaluated personality traits (Baeyens, Eelen, Van den Bergh, & Crombez, 1992). The initially acquired evaluation was easily overridden, and the new evaluation based on the trait description proved to be strong and long lasting. In contrast, a long tradition in social psychology shows how it is difficult to modify attitudes and impressions that are based on group membership, suggesting that simple evaluative conditioning would be unable to override an evaluation that was originally produced by a category label.

As we have said, category labels and personality traits define the essence of the target—the way in which it is mentally construed and represented. Category labels are assumed to reflect “natural” and enduring differences between individuals. They direct the organization of information in memory, trigger inferential processes, and so on. It is likely that the resulting processes differ from those that are involved in traditional evaluative conditioning research. The process differences might also account for the strong effects found in the present research, in which a single exposure produced pervasive automatic effects in a variety of domains.

In addition, evaluative conditioning can hardly explain results from the minimal group study (Study 3), where faces were paired

with labels that were originally neutral (i.e., Type A vs. Type B). Hence, the evaluation of the target could not directly follow from the valence of the category label, but it had to be built up by identifying whether the target was an in-group or an out-group member.

To conclude, we believe that the categorization of social targets cannot simply be conceptualized in terms of evaluative conditioning. Even though we cannot rule out the possibility that evaluative conditioning has somehow played a role, evaluative conditioning can hardly be the only process at work in these studies.

### *Affective Responses as Implicit Memory Effects*

The results from Study 1 demonstrate that the affective connotation associated with an exemplar-based representation is unintentionally retrieved. We also argue that the affective connotation “attached” to a person-based representation may indeed influence information processing even though the original causes of such an affective response are forgotten. In this context, the anecdotal case reported by the famous French neuropsychologist Claparede (1911/1951) is highly indicative. Claparede shook hands with one of his amnesic female patients while surreptitiously concealing a pin in his hand. At later encounters, the woman reflexively withdrew and refused to shake hands with Claparede even though she was not able to remember the previous unfortunate experience. Similarly, we have demonstrated that the affective responses toward group members could still remain subtly present when participants can no longer recall the category membership of the target. This result is relevant to two different issues. The first has to do with the organization of information in memory, whereas the second is more directly related to person perception and stereotyping.

As for the organization of information in memory, it has been repeatedly shown that information is encoded at different levels of abstraction, so that when low-level information is lost there is still indication of higher level clustering (Arcuri, 1982; Brewer, Weber, & Carini, 1995; Lingle & Ostrom, 1981; Potts, 1974; Sherman, Castelli, & Hamilton, 2002; Taylor, Fiske, Etcoff, & Ruderman, 1978). Studies 4 and 5 demonstrated that clustering may also occur on the basis of affective information (Niedenthal et al., 1999) and not only on the basis of “cold” conceptual information. As suggested by Niedenthal et al. (1999), “the construction and use of a category of things that have caused the same emotion allow social perceivers to understand the meaning of an object in light of their own personal learning histories and goals” (p. 338). Therefore, members of evaluatively consistent groups are clustered together.

In addition, results provide strong support for the often-found dissociation between evaluative responses and explicit memory. The adoption of a procedure in which the effects of evaluative responses and the effects of explicit memory are posed in opposition has multiple advantages. First, it enables tests of a specific hypothesis rather than grounding the theoretical arguments on a null effect (a null correlation between memory and judgments). In addition, there is no need to explicitly ask participants for an evaluation, but it is possible to verify how spontaneous evaluations influence responses in a memory task. Hence, this paradigm may also be used to assess the implicit evaluation associated with social groups. For instance, the differential intrusions of in-group and

out-group members into blatantly positive and negative groups may be assessed, as demonstrated in Study 5.

As for person perception processes and stereotyping, the results from Studies 4 and 5 suggest that discrimination toward members of stigmatized groups may be manifested in different forms and may take subtle routes. Indeed, it appears that the correct recognition of the target's category membership is not fundamental to the activation of evaluative responses (see Brewer, 1988; Fiske & Neuberg, 1990) and that in contrast, evaluative responses may sometimes affect categorization. This means that members of stigmatized groups may later be more likely to be included in another negative group when their original category membership is not recalled. For instance, if an animal rights activist discovers that someone goes hunting every weekend and later has to judge the performance of that person, it is possible that the activated negative feeling will be (mis)attributed. If the original category membership of the target is not remembered, the negative evaluation may be attributed to a poor performance. Faced with a familiar exemplar whose category membership has been forgotten, the activated feelings may be attributed to other irrelevant cues in the current situation (see Castelli & Zogmaister, 2000). Therefore, members of stigmatized groups may continue to be discriminated against for reasons that have nothing to do with their actual category membership.

## References

- Ajzen, I. (2000). Nature and operation of attitudes. *Annual Review of Psychology*, 52, 27–58.
- Andersen, S. M., & Cole, S. W. (1990). "Do I know you?": The role of significant others in general social perception. *Journal of Personality and Social Psychology*, 59, 384–399.
- Arcuri, L. (1982). Three patterns of social categorization in attribution memory. *European Journal of Social Psychology*, 12, 271–282.
- Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective–evaluative value: Conservative but not unchangeable. *Behaviour Research and Therapy*, 27, 279–287.
- Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning: Acquired valence is sensitive to US-revaluation. *Learning and Motivation*, 23, 200–224.
- Baldwin, M. W., Carrell, S. E., & Lopez, D. F. (1990). Priming relationship schemas: My advisor and the Pope are watching me from the back of my mind. *Journal of Experimental Social Psychology*, 26, 435–454.
- Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology*, 68, 181–198.
- Bargh, J. A. (1989). Conditional automaticity: Varieties of automatic influence in social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought*. New York: Guilford Press.
- Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Erlbaum.
- 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.
- Bodenhausen, G. V., Schwarz, N., Bless, H., & Wänke, M. (1995). Effects of atypical exemplars on racial beliefs: Enlightened racism or generalized appraisals? *Journal of Experimental Social Psychology*, 31, 48–63.
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive–motivational analysis. *Psychological Bulletin*, 86, 307–324.
- Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.
- Brewer, M. B. (1999). The psychology of prejudice: In-group love or out-group hate? *Journal of Social Issues*, 55, 429–444.
- Brewer, M. B., Weber, J. G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology*, 69, 29–40.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Castelli, L., & Paladino, M. P. (2003). *Spontaneous behavioral tendencies in intergroup perception*. Manuscript in preparation.
- Castelli, L., & Zogmaister, C. (2000). The role of familiarity in implicit memory effects: The case of exemplar activation. *European Journal of Social Psychology*, 30, 223–234.
- Castelli, L., Zogmaister, C., & Arcuri, L. (2001). Exemplar activation and interpersonal behavior. *Current Research in Social Psychology*, 6, 33–45.
- Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, 25, 215–224.
- Claparede, E. (1951). Recognition and "me-ness." In D. Rapaport (Ed.), *Organization and pathology of thought* (pp. 58–75). New York: Columbia University Press. (Original work published 1911)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Crocker, J., Major, B., & Steele, C. (1998). Social stigma. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 504–553). New York: McGraw-Hill.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163–170.
- Dagenbach, D., Horst, S., & Carr, T. H. (1990). Adding new information to semantic memory: How much learning is enough to produce semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 581–591.
- Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, 81, 800–814.
- Dasgupta, N., McGhee, D. E., Greenwald, A. D., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, 36, 316–328.
- De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Emotion and Cognition*, 12, 145–169.
- De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin*, 127, 853–869.
- Dijksterhuis, A., Spears, R., Postmes, T., Stapel, D., Koomen, W., van Knippenberg, A., & Scheepers, D. (1998). Seeing one thing and doing another: Contrast effects in automatic behavior. *Journal of Personality and Social Psychology*, 75, 862–871.
- Fazio, R. H. (1986). How do attitudes guide behavior? In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (Vol. 1, pp. 204–243). New York: Guilford Press.
- Fazio, R. H. (1993). Variability in the likelihood of automatic attitude activation: Data reanalysis and commentary on Bargh, Chaiken, Govender, and Pratto (1992). *Journal of Personality and Social Psychology*, 64, 753–758.
- Fazio, R. H. (2000). Accessible attitudes as tools for object appraisal: Their

- costs and benefits. In G. R. Maio & J. M. Olson (Eds.), *Why we evaluate: Functions of attitudes* (pp. 1–36). Mahwah, NJ: Erlbaum.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013–1027.
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229–238.
- Fiske, S. T. (1980). Impression and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, *38*, 889–906.
- 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: Developments in schema-triggered affect. In R. M. Sorrentino & E. T. Higgins (Eds.), *Handbook of motivation and cognition* (pp. 167–203). New York: Guilford Press.
- Gaertner, S. L., & Dovidio, J. F. (2000). *Reducing intergroup bias: The common ingroup identity model*. Philadelphia: Psychology Press.
- Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic patients. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 501–518.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.
- Hamilton, D. L., & Sherman, S. J. (1996). Perceiving persons and groups. *Psychological Review*, *103*, 336–355.
- Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*, 258–268.
- Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of affective information in memory. *Cognition and Emotion*, *8*, 515–533.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- 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., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*, 326–338.
- Jacoby, L. L., Woloshyn, V., & Kelley, C. M. (1989). Becoming famous without being recognized: Unconscious influences of memory produced by dividing attention. *Journal of Experimental Psychology: General*, *118*, 115–125.
- Jacoby, L. L., Yonelinas, A. P., & Jennings, J. M. (1997). The relation between conscious and unconscious (automatic) influences: A declaration of independence. In J. D. Cohen & J. W. Schooler (Eds.), *Scientific approaches to consciousness* (pp. 13–48). Mahwah, NJ: Erlbaum.
- Johnson, B. T. (1989). DSTAT Version 1.00 software for the meta-analytic review of research literature [Computer software]. Mahwah, NJ: Erlbaum.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28.
- Klauer, K. C. (1998). Affective priming. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 8, pp. 67–103). Chichester, England: Wiley.
- Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology*, *48*, 563–574.
- Lingle, J. H., & Ostrom, T. M. (1981). Principles of memory and cognition in attitude formation. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 399–420). Hillsdale, NJ: Erlbaum.
- Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492–527.
- Maass, A., & Schaller, M. (1991). Intergroup biases and the cognitive dynamics of stereotype formation. In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 2, pp. 189–209). Chichester, England: Wiley.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Castelli, L. (1999). On disregarding deviants: Exemplar typicality and person perception. *Current Psychology*, *18*, 47–70.
- Macrae, C. N., Bodenhausen, G. V., Milne, A. B., Castelli, L., Schloerscheidt, A., & Greco, S. (1998). On activating exemplars. *Journal of Experimental Social Psychology*, *34*, 330–354.
- 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.
- McKoon, G., & Ratcliff, R. (1986). Automatic activation of episodic information in a semantic memory task. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 108–115.
- Niedenthal, P. M. (1990). Implicit perception of affective information. *Journal of Experimental Social Psychology*, *26*, 505–527.
- Niedenthal, P. M., & Halberstadt, J. H. (2000). Grounding categories in emotional response. In J. P. Forgas (Ed.), *Feeling and thinking: The role of affect in social cognition* (pp. 357–386). New York: Cambridge University Press.
- Niedenthal, P. M., Halberstadt, J. H., & Innes-Ker, A. H. (1999). Emotional response categorization. *Psychological Review*, *106*, 337–361.
- Niedenthal, P. M., & Kitayama, S. (1994). *The heart's eye: Emotional influences in perception and attention*. New York: Academic Press.
- Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the Implicit Association Test. *Social Cognition*, *19*, 97–144.
- Otten, S., & Moskowitz, G. B. (1999). Evidence for implicit evaluative in-group bias: Affect-biased spontaneous trait inference in a minimal group paradigm. *Journal of Experimental Social Psychology*, *36*, 77–89.
- Otten, S., & Wentura, D. (1999). About the impact of automaticity in the minimal group paradigm: Evidence from affective priming task. *European Journal of Social Psychology*, *29*, 1049–1071.
- Pecher, D., & Raaijmakers, J. G. W. (1999). Automatic priming effects for new associations in lexical decision and perceptual identification. *Quarterly Journal of Experimental Psychology*, *52A*, 593–614.
- Perdue, C. W., Dovidio, J. F., Gurtman, M. B., & Tyler, R. B. (1990). Us and them: Social categorization and the process of intergroup bias. *Journal of Personality and Social Psychology*, *59*, 475–486.
- 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.
- Potts, G. R. (1974). Storing and retrieving information about ordered relationships. *Journal of Experimental Psychology*, *103*, 431–439.
- Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*, 380–391.
- Rogers, T., Kuiper, N., & Kircher, W. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677–688.
- Rothbart, M., & Taylor, M. (1992). Category labels and social reality: Do we view social categories as natural kinds? In G. R. Semin & K. Fiedler (Eds.), *Language, interaction, and social cognition* (pp. 11–36). London: Sage.

- 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.
- Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.
- Sherman, S. J., Castelli, L., & Hamilton, D. L. (2002). The spontaneous use of a group typology as an organizing principle in memory. *Journal of Personality and Social Psychology, 82*, 328–342.
- Smith, E. R. (1990). Content and process specificity in the effects of prior experiences. In T. K. Srull & R. S. Wyer Jr. (Eds.), *Advances in social cognition* (Vol. 3, pp. 1–59). Hillsdale, NJ: Erlbaum.
- Smith, E. R. (1992). The role of exemplars in social judgment. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 107–132). Hillsdale, NJ: Erlbaum.
- Smith, E. R., & Zárate, M. A. (1990). Exemplar and prototype use in social categorization. *Social Cognition, 8*, 243–262.
- Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99*, 3–21.
- Stapel, D. A., Koomen, W., & van der Plight, J. (1997). Categories of category accessibility: The impact of trait versus exemplar priming on person judgments. *Journal of Experimental Social Psychology, 33*, 44–76.
- Taylor, S. E., Fiske, S. T., Etoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36*, 778–793.
- Vaes, J., Paladino, M. P., Castelli, L., Leyens, J. -P., & Giovanazzi, A. (2003). On the behavioral consequences of infrahumanization: The implicit role of uniquely human emotions in intergroup relations. *Journal of Personality and Social Psychology, 85*, 1016–1034.
- 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.
- Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37–65). Hillsdale, NJ: Erlbaum.
- Yazuv, H. S. (1963). The retention of incidentally learned connotative responses. *Journal of Psychology, 63*, 409–418.
- Yzerbyt, V. Y., Rocher, S. J., & Schadron, G. (1997). Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. Oakes, N. Ellemers, & A. Haslam (Eds.), *The psychology of stereotyping and group life* (pp. 20–50). London: Blackwell.

Received September 10, 2001

Revision received October 17, 2003

Accepted October 19, 2003 ■