Journal of Personality and Social Psychology
2001, Vol. 80, No. 4, 669–685

Copyright 2001 by the American Psychological Association, Inc.
0022-3514/01/$5.00 DOI: 10.1037//0022-3514.80.4.669

What's in a Name: Implicit Self-Esteem and the Automatic Self

Sander L. Koole
Free University Amsterdam

Ap Dijksterhuis and Ad van Knippenberg
University of Nijmegen

This article explores the links between implicit self-esteem and the automatic self (D. L. Paulhus, 1993). Across 4 studies, name letter evaluations were positively biased, confirming that implicit self-esteem is generally positive (A. G. Greenwald & M. R. Banaji, 1995). Study 1 found that this name letter bias was stable over a 4-week period. Study 2 found that positive bias for name letters and positive bias for birth date numbers were correlated and that both biases became inhibited when participants were induced to respond in a deliberative manner. Studies 3–4 found that implicit self-evaluations corresponded with self-reported self-evaluations, but only when participants were evaluating themselves very quickly (Study 3) or under cognitive load (Study 4). Together, these findings support the notion that implicit self-esteem phenomena are driven by self-evaluations that are activated automatically and without conscious self-reflection.

Just what are people doing when they are evaluating themselves? Conventional psychological wisdom holds that the self-evaluation process invariably operates through conscious self-reflection (Baumeister, 1998; Brown, 1998; Sedikides & Strube, 1997). Accordingly, individuals engaged in self-evaluation are believed to be "peering inward" (Hixon & Swann, 1993), asking themselves questions about such topics as their self-attributes (Pelham & Swann, 1989; Sedikides, 1993), the causes of their behavior (Bradley, 1978; Sedikides, Campbell, Reeder, & Elliott, 1998), their actions in the past (Brown & Dutton, 1995; Ross, 1989), or their plans for the future (Koole & van't Spijker, 2000; Markus & Nurius, 1986).

In recent years, however, several students of the self have come to question the role of consciousness in the self-evaluation process (Brown, 1993; Epstein & Morling, 1995; Greenwald & Banaji, 1995; Leary & Downs, 1995). In view of evidence that many important social–cognitive processes can function without any need for conscious guidance (Bargh & Chartrand, 1999; Greenwald & Banaji, 1995), these scholars have argued that certain self-evaluations may similarly operate at unconscious levels. Such unconscious self-evaluations are presumably inaccessible to introspection, so that they may only be observed implicitly or indirectly (Greenwald & Banaji, 1995). Consistent with this argument, an accumulation of research has shown that people display a pervasive positive bias in their evaluations of self-associated stimuli, such as name letters (Nuttin, 1985, 1987), personal belongings (Beggan, 1992), and ingroup members (Otten & Wentura, 1999). These and related forms of *implicit self-esteem* (Greenwald & Banaji, 1995) are remarkable for occurring in the absence of any explicit encouragement to engage in self-evaluative activity. Moreover, people are lacking in awareness of exhibiting implicit self-esteem (Nuttin, 1985), suggesting that implicit self-esteem is a form of self-evaluation that occurs in the absence of conscious self-reflection (Greenwald & Banaji, 1995).

Noting the theoretical importance of understanding the role of unconscious processes in self-evaluation, the present research seeks to further illuminate the psychology of implicit self-esteem. In the following paragraphs, we begin by considering a hypothetical account of the development of implicit self-esteem phenomena. We present this account not because we want to test it in its entirety, but to merely clarify the basis for our predictions. Next, we discuss how both implicit and explicit, self-reported forms of self-evaluation may operate within a single self system. Finally, we present four investigations that were designed to test some key hypotheses of our theoretical analysis.

THE AUTOMATION OF SELF-EVALUATION

How can people manage to overlook a certain subset of their self-evaluations? Some useful clues may be found by considering how self-evaluations may change over time. As developmental research has shown, some key components of the self-evaluation process can already be discerned during infancy. For instance, the ability to make affective discriminations (Fernald, 1993; Shapiro, Eppler, Haith, & Reis, 1987, cited in Swann & Schroeder, 1995) and the capacity for self-recognition (Lewis & Brooks-Gunn, 1979) have been observed in infants less than 1 year old. Findings of this sort suggest that even very young children are capable of a rudimentary form of self-evaluation. For at least two reasons, such early self-evaluations may shape the person's subsequent self-evaluations. First, early self-evaluations may act as mental working models that structure the flow of self-relevant cognitions, affects, and behaviors (Bowlby, 1973; Mikulincer, 1995). Early self-evaluations may thus become consolidated into the person's cognitive–affective architecture. Second, repeated activation may render the retrieval of the person's early self-evaluations increas-

Sander L. Koole, Department of Social Psychology, Free University Amsterdam, Amsterdam, the Netherlands; Ap Dijksterhuis and Ad van Knippenberg, Department of Psychology, University of Nijmegen, Nijmegen, the Netherlands.

Correspondence concerning this article should be sent to Sander L. Koole, Department of Social Psychology, Free University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands.

ingly facile, to the point where they can spring to mind unwittingly (Bargh & Barndollar, 1996; Smith & Lerner, 1986). Because of these processes, then, early self-evaluations may become an integral part of the *automatic self,* the self that consists of his or her most well-practiced, and hence, most chronically activated self-evaluations (Paulhus, 1993).

Granting that developmentally early self-evaluations are especially likely to become automated, it becomes important to know what people's earliest self-evaluations are like. Interestingly, research has revealed some clear developmental trends in the contents of people's self-evaluations. In particular, it appears that young children's self-evaluations are often characterized by a rampant preference for positive feedback (Swann & Schroeder, 1995). Already at the age of 5 months, children prefer smiling faces over nonsmiling faces (Shapiro et al., 1987) and begin orienting to voices having the melodic contour of acceptance (Fernald, 1993). This initial tendency toward positivity may reflect young children's primary concern with establishing a secure positioning of the self (Solomon, Greenberg, & Pyszczynski, 1991). Moreover, self-enhancement strivings can be accomplished through computationally simple means. Indeed, a self-enhancing device only needs to follow two simple rules: (a) if the self-reflection information is positive, then accept it; and (b) if the self-relevant information is negative, then reject it (Swann, Hixon, Stein-Seroussi, & Gilbert, 1990). This computational simplicity may further contribute to the early acquisition of self-enhancement skills (Swann et al., 1990). Research indicates that children's tendencies to self-enhance continue to flourish after they become verbal and thus capable of rating themselves (Benenson & Dweck, 1986; Eshel & Klein, 1981; Stipek & Daniels, 1988; see Swann & Schroeder, 1995). In fact, self-positivity does not subside until early adolescence, a period when self-evaluations often become clouded by identity confusion (Baumeister & Tice, 1986; Erikson, 1968). By that time, however, children's early inclinations toward self-positivity may have received ample opportunity to become fully automated (Paulhus, 1993; Swann et al., 1990).

On the basis of the foregoing account, we may sketch a preliminary account of implicit self-esteem effects. When people encounter stimuli that are in some way associated with self, the subset of people's highly overlearned self-evaluations may become automatically activated (Bargh, Chaiken, Raymond, & Hymes, 1996; Fazio, Sanbonmatsu, & Williams, 1986). Because of their unconscious nature, these evaluations will often be diffuse and ill-defined, so that they are easily confused with the evaluations of other stimuli that happen to be in focal attention (Murphy & Zajonc, 1993). Furthermore, given that people's automatic self-evaluations are generally positive (Paulhus, 1993), activation of automatic self-evaluations will typically render positive evaluations more accessible in memory. As a consequence, automatic self-evaluations may give rise to a positive bias in people's evaluations of self-associated stimuli.

THE SELF-EVALUATION PROCESS REVISITED: THE CASE FOR DUAL SELF-EVALUATIONS

Now that we have gained an improved understanding of implicit self-esteem effects, we are in a better position to consider how implicit self-evaluations may be related to self-evaluations that are derived through conscious self-reflection. The foregoing account implies that there exist a host of differences between implicit and explicit, introspectively accessible self-evaluations. In terms of process, implicit self-evaluations are presumably more automatic, meaning that they are relatively more unconscious, unintentional, efficient, and uncontrollable than explicit self-evaluations (Bargh, 1994). In terms of contents, implicit self-evaluations are likely to be more positive than explicit self-evaluations. Indeed, implicit self-evaluations are likely to be produced by rather primitive self-enhancement mechanisms, whereas explicit self-evaluations may involve more sophisticated cognitive judgments of the self (e.g., verifying whether an evaluation matches pre-existing self-knowledge; Swann & Schroeder, 1995).

In light of these considerations, it seems reasonable to conclude that implicit and explicit self-evaluations represent two qualitatively different kinds of self-evaluation. At first glance, this conclusion may seem rather awkward, perhaps because it violates deeply rooted beliefs about the self as a unitary entity (Baumeister, 1998). However, as Greenwald (1982) argued, there is much to be said for the more unorthodox conception of the self as a conglomerate of multiple subsystems. A similar point is made by Epstein's (1994) cognitive-experiential self theory (CEST). Briefly, using CEST, Epstein has argued that people respond to the world on two levels, rational and experiential. These ways of responding are conceived as two distinct mental systems for adapting to reality. The rational system is more duty- than passion-driven. It operates primarily on a conscious level and in the medium of language and requires deliberate effort. The experiential system is motivated by affect. It operates primarily on an unconscious level and processes information holistically, rapidly, and effortlessly. Although the two systems are seen as independent, they are assumed interact with each other, so that behavior is almost always influenced by both systems. In a recent extension of CEST, Epstein and Morling (1995) have argued that the two systems may map on to different kinds of self-evaluation, with the rational system being predominant in people's explicit self-evaluations and the experiential system being predominant in people's important self-evaluations. As such, CEST provides an integrative framework for understanding the relation between implicit and explicit self-evaluation.

Complementing the CEST approach, Wilson, Lindsey, and Schooler (2000) recently proposed a model of dual attitudes, which highlights the complex interplay between implicit and explicit evaluations. According to this model, people may simultaneously possess different evaluations towards the same attitude object. The formation of such dual attitudes typically occurs when people decide to change their attitudes. In such cases, people's newly formed attitudes become represented at an explicit level, where they are readily accessible to self-reporting. However, the activation of these newly formed attitudes has not yet become automatic and, hence, requires a fair amount of cognitive capacity and motivation. At the same time, people's original attitudes are assumed to continue to exist in memory, albeit at an implicit level. This implicit level is presumably more difficult to access through conscious introspection. Because people's implicit evaluations are developmentally older, they are more likely to be overlearned, so that they can be retrieved from memory automatically. Thus, even after attitude change has occurred, people may spontaneously activate their original, implicit attitudes upon encountering the attitude object. When there is sufficient capacity and motivation, people will effortfully retrieve their newly formed, explicit atti-

tudes from memory. By contrast, when capacity and motivation are lacking, people's original, implicit attitudes may reemerge.

Applied to the domain of self-evaluation, the model of dual attitudes implies that implicit self-esteem effects are reflective of self-evaluations that are habitual and automatic. Presumably, these automatic self-evaluations affect uncontrollable responses, or responses that people do not view an expression of their self-evaluations and consequently do not attempt to control. For instance, people may not regard their evaluations of their name letters as indicative of their self-evaluations and thus allow their automatic self-evaluations to contaminate their judgments of these stimuli (Wilson & Brekke, 1994). The model of dual attitudes further implies that whenever people are engaging in conscious self-reflection their implicit self-evaluations may be overruled by more effortful, explicit self-evaluations. Such deliberative overriding, however, should only occur when people possess sufficient capacity and motivation to do so. When sufficient capacity or motivation are unavailable, people's implicit self-evaluations may predominate in their reported self-evaluations. In line with this reasoning, several experiments have shown that depriving people of cognitive resources often induces people to provide more favorable self-evaluations (Paulhus, Graf, & Van Selst, 1989; Paulhus & Levitt, 1987; cf. Paulhus, 1993).[1] It is conceivable that these increases in self-positivity are reflective of primitive self-enhancement tendencies that have continued to exist at an automatic mode of responding. As such, these findings point to the existence of dual self-evaluations.

OVERVIEW OF THE PRESENT RESEARCH

The preceding theoretical analysis led us to make the following set of interrelated predictions. First, to the extent that implicit self-esteem phenomena represent a sediment of people's habitualized, highly overlearned self-evaluations, implicit self-esteem should show at least moderate consistency over time. This prediction was tested in Study 1. Second, given that introspection has been found to disrupt the operation of automatic evaluative processes (Wilson, Dunn, Kraft, & Lisle, 1989; Greenwald & Banaji, 1995), active efforts at introspection may inhibit displays of implicit self-esteem. This prediction was tested in Study 2. Finally, our theoretical account suggests that the relationship between implicit and explicit self-evaluation is complex and highly dynamic. When people possess sufficient capacity and motivation, their explicit self-evaluations may predominantly reflect their more deliberate evaluations of themselves. However, when motivation or capacity are lacking, people's self-evaluations may predominantly reflect their automatic, implicit self-evaluations. The latter predictions were tested in Studies 3 and 4.

In all four studies, we operationalized implicit self-esteem by assessing participants' tendency to overevaluate name letters. As noted earlier, research has shown that people tend to evaluate own name letters more positively than people who do not share their name letters (Greenwald & Banaji, 1995; Hoorens, 1990; Kitayama & Karasawa, 1997; Nuttin, 1985, 1987). The robustness of this effect is suggested by recent findings showing that a positive bias for own name letter affects even important life decisions, such as deciding which place to live, which college to attend, or which career to choose (Pelham, Mirenberg, & Jones, 2000). Moreover, overevaluation of own name letters occurs even when name letters are presented separately and in scrambled order among the other letters of the alphabet (e.g., Nuttin, 1985), so that any explicit reference to the self is avoided. Finally, people are unaware of displaying a preference for name letters (Nuttin, 1985, 1987), and this preference cannot be reduced to methodological artifacts (Nuttin, 1985, 1987), cultural-linguistic idiosyncrasies (Kitayama & Karasawa, 1997; Nuttin, 1987; Hoorens, Nuttin, Erdélyi Herman, & Pavakanun, 1990), or more frequent exposure to name letters than to no-name letters (Greenwald & Banaji, 1995; Hoorens & Nuttin, 1993; Kitayama & Karasawa, 1997). In light of these findings, overevaluation of name letters qualified as a valid indicator of implicit self-esteem.

STUDY 1

Our first investigation set out to achieve two important goals. First, we sought to obtain further evidence for an implicit self-esteem effect in the evaluation of own name letters. Previous work has found the tendency to overevaluate own name letters to be pervasive and highly robust (Kitayama & Karasawa, 1997; Nuttin, 1985, 1987; Pelham et al., 2000; cf. Greenwald & Banaji, 1995; Hoorens, 1990). Nevertheless, we deemed it desirable to replicate this phenomenon. To this end, we asked our participants to provide evaluations of each of the letters of the alphabet. Using these ratings, we then determined whether participants' name letter evaluations were exaggerated in comparison with a corresponding baseline of no-name letter evaluations. The occurrence of a systematic bias in favor of each participants's name letters was taken as evidence for implicit self-esteem (Kitayama & Karasawa, 1997).

Second, and of greater theoretical interest, we sought to gain more insight into the temporal stability of implicit self-esteem, as assessed through relative liking for own name letters. According to our theoretical analysis, implicit self-esteem results from automatic, highly practiced self-evaluations. If this is correct, one may expect a considerable amount of temporal stability in implicit self-esteem, and, consequently, in relative liking for own name letters. Of some relevance here, a series of studies by Hoorens et

[1] Some recent research suggests that depriving people of cognitive resources does not always induce people to provide more favorable self-evaluations. Specifically, Kruger (1999) recently found that cognitive load amplified positivity in self-evaluation in domains in which absolute skills tend to be high, whereas cognitive load reduced positivity in self-evaluation in domains in which absolute skills tend to be low. In agreement with Kruger (1999), we do not regard these findings as inconsistent with the existence of automatic self-enhancement. First, Kruger's (1999) analysis is only applicable to comparative ability judgments. The computational complexity of comparative judgments is markedly higher than that of absolute judgments, so that the former may be less subject to automatic self-enhancement tendencies than the latter. Second, domains in which people's personal skills are low also tend to be low in personal importance (Pelham & Swann, 1989). Accordingly, the type of explicit self-evaluation studied by Kruger (1999) may have possessed too little relevance to the self to arouse the self-enhancement motive (Pelham, 1991). Finally, the current account does not preclude that cognitive load amplifies self-enhancement tendencies more strongly among some individuals than others. Indeed, as we later argue, the effects of cognitive load on explicit self-evaluation may vary meaningfully between individuals with different levels of implicit self-esteem.

al. (1990) found evidence for overevaluation of name letters among second graders (who were between 6 and 7 years old). This evidence is at least consistent with the notion that the self-evaluations underlying overevaluation of name letters are related to developmentally early self-evaluations. As far as we know, however, no previous research has directly examined the temporal stability of enhanced name letter liking. To address this issue, we designed Study 1 to examine the temporal stability of relative liking for name letters across a 4-week interval.

Method

Participants

During the first time of measurement, around 160 completed questionnaires were returned out of the 200 that had been distributed. During the second time of measurement, 93 completed questionnaires were returned. Thus, the overall response rate was close to 50%. Participants were 93 undergraduate students from the University of Nijmegen (16 male and 77 female, average age 20 years), who participated on a voluntary basis.

Procedure and Materials

Administration of the research materials took place in groups of up to 15 persons. After taking classes in an introductory psychology course, participants were asked to fill out a packet of questionnaires. They were informed that the investigators were interested in studying personality differences. The first questionnaire in each packet contained our measure of name letter liking. The remaining questionnaires contained filler items that were irrelevant to the purpose of the current investigation.

Following Nuttin (1985), the first instruction sheet explained that this study was concerned with people's aesthetic judgments of simple stimuli, that is, letters of the alphabet. It was further explained that participants might not be accustomed to evaluating letters, but that previous research had shown that the study of these kinds of judgments can lead to a better understanding of certain aspects of human emotions. As in previous research (Nuttin, 1985), participants were encouraged to rely on their first, intuitive reactions toward the letters. The next pages contained the letters of the alphabet, arranged in a randomized order (for the purpose of the investigation, we created several randomized versions of the questionnaire). Participants were asked to evaluate each letter of the alphabet, by circling the appropriate numbers on 9-point scales (1 = *not at all beautiful,* 9 = *extremely beautiful*). After the name letter evaluation task, participants answered several filler questionnaires and some general background questions. At the end of the session, participants were requested to write down their full names on a consent form. The same procedure was repeated after four weeks. Finally, participants were debriefed and thanked for their participation.

Construction of Dependent Variables

To control for the baseline attractiveness of participants' name letters, we followed the procedure outlined by Kitayama and Karasawa (1997).[2] Accordingly, we first computed a baseline evaluation of each letter for those whose names did not include it. Subsequently, baseline evaluations were aggregated across both conditions for each letter. Next, for each respondent, a difference score was computed between the evaluations of each of the name letters and the corresponding no-name evaluations to yield a relative liking score. Overevaluation of name letters is indicated by positive relative evaluations.[3]

Results and Discussion

On average, participants' relative name letter evaluations were positive at Time 1, $M = .40$, and at Time 2, $M = .64$. Indeed, both means differed reliably from zero, both Fs > 10, both ps $< .001$. Thus, our findings corraborate earlier findings that name letter evaluations are positively biased (e.g., Hoorens et al., 1990; Kitayama & Karasawa, 1997; Nuttin, 1985, 1987). A repeated measures analysis of variance (ANOVA) showed that relative liking for name letters did not differ between the two measurements, $F(1, 92) = 1.69$, $p < .20$. Importantly, a correlational analysis revealed that relative liking for name letters at Time 1 was strongly correlated with relative liking for name letters at Time 2, $r(93) = .62$, $p < .001$. However, it remains possible that this correlation was inflated by a general tendency towards positive responding. If this explanation is correct, the test–retest correlation between name letter during the two measurements should become substantively reduced when variations in no-name letter liking

[2] Because the Kitayama and Karasawa (1997) method corrects for group liking scores, it may lead to an underestimation of name letter bias for participants with well-liked name letters and to an overestimation of name letter bias for participants with unpopular name letters. However, several considerations make it unlikely that differential popularity of name letters was a strong influence in the current findings. First, when we examine group liking scores across our studies, we find that group liking scores generally did not display many extreme values. For instance, in Study 1, group liking scores varied between 4.1 and 6.6 on a 9-point scale, with the mean value being 4.88. Mean group liking scores in the other studies similarly hovered around the conceptual midpoint of the scale. This lack of extreme values may reflect the fact that letters are intrinsically neutral stimuli (at least, when they are not part of a person's name!). In addition, group liking scores represent an average of the evaluations by people who do not have the relevant letter in their names. This procedure causes any idiosyncracies (and, hence, extreme values) in letter evaluations to be averaged out. In sum, the occurrence of extreme group liking scores appears to have been quite low in our research. Second, our name letter measure contained multiple name letters for each participant. With these multiple measurements, chances become reduced that particular letters were overrepresented among our sample. Thus, the influence of extreme group liking scores was further minimized in our studies. Third, the potential influence of extreme group liking scores would have resulted in less power in the current research. As such, it cannot account for any theoretically meaningful relationships that were uncovered in our research.

[3] In the literature on name letter liking, there has been some debate on the question of whether the phenomenon occurs for every name letter, or whether it is restricted to initial name letters (Johnson, 1986, cited in Greenwald & Banaji, 1995; Nuttin, 1987; Kitayama & Karasawa, 1997). In addition, some studies have found differences in overevaluation of first names and family names (Kitayama & Karasawa, 1997). To examine these issues, we computed overevaluation of initial letters and remaining letters separately for first names and family names for all participants of the four studies reported in this article (participants in the reasons condition in Study 2 were excluded from this analysis). Subsequent analyses revealed that overevaluation of initial letters was much stronger than overevaluation of remaining letters. In addition, overevaluation of first names was somewhat stronger than overevaluation of family names. Nevertheless, all types of name letters were reliably overevaluated, all ps $< .001$. In addition, overevaluation of the four types of name letters was intercorrelated (Cronbach's alphas for the three studies in this article were between .56 and .70). Moreover, the main results reported in this article did not vary systematically as a function of name letter type. In light of these results, we are inclined to agree with Nuttin's (1987) assertion that overevaluation of the different types of name letters is driven by the same underlying processes, that is, implicit affection for self. We therefore report only the results for general overevaluation of name letters in the main body of this article.

(which presumably reflect a general tendency towards positive responding) are statistically held constant. To follow up on this possibility, we simultaneously regressed liking for no-name letters at Times 1 and 2, and liking for name letters onto liking for name letters at Time 1. The results revealed a significant effect for no-name letter liking at Time 1, $\beta = .38$, $t(90) = 3.73$, $p < .001$, which suggested that name letter liking at Time 1 was moderately correlated with a general tendency towards positive responding. More important, however, name letter liking at Time 2 remained strongly correlated with name letter liking at Time 1, $\beta = .56$, $t(90) = 6.49$, $p < .001$. Thus, the test–retest correlation for name letter liking could not be explained by a general tendency towards positive responding.

It may still be argued that self-reported self-evaluations frequently attain temporal stabilities that are comparable or higher, but it should be recognized that measurement error in implicit measures is typically higher than in explicit measures (cf. Dovidio, Kawakami, & Beach, in press; McClelland, 1980). The relatively large measurement error in implicit measures may serve to depress these measures' reliability, such as stability over time. Thus, finding any temporal stability in implicit self-esteem may be considered more informative than the level of this temporal stability. Although investigating the temporal stability of implicit self-esteem across longer time periods is desirable, the obtained 4-week stability of enhanced name letter liking is at least consistent with our theorizing that this phenomenon is reflective of automatic, overlearned self-evaluations.

STUDY 2

In Study 2, we sought to test the presumed automaticity of implicit self-esteem in an alternative manner. Several lines of research have shown that deliberative thought is capable of inhibiting automatic affective responding. For instance, Epstein and colleagues have shown that experiential processing may be inhibited by rational processing when cues in the situation indicate a need for rational analysis (Epstein, Donovan, & Denes-Raj, 1999). In a related vein, Wilson and colleagues have demonstrated that asking people to explain their feelings increases the salience of the cognitive component of an attitude, at the expense of the salience of the affective component (Wilson et al., 1989). Finally, Murphy and Zajonc (1993) have found that the affective impact of priming stimuli becomes diluted when the priming procedures allow sufficient time for deliberative processing of the priming stimuli (see also Greenwald & Banaji, 1995).

In light of these previous findings, we reasoned that the automatic self-evaluations that presumably underly implicit self-esteem effects might similarly become inhibited by self-reflective thought. To test this reasoning, we asked participants to complete the same letter evaluation task as in Study 1. During this evaluation task, one half of the participants were encouraged to reason why they felt the way they did about the letter stimuli (Wilson et al., 1989). The remaining participants were instructed to rely on their feelings in evaluating the stimuli, as in Nuttin (1985) and Study 1. If enhanced liking for name letters is indeed caused by people's automatic, intuitive self-evaluations, positive bias for name letters should readily emerge when participants rely on their feelings but less so when they are analyzing reasons. If, on the other hand, enhanced liking for name letters is dependent upon more deliberative processes, this effect should not be obtained or even reversed.

We further examined some of the processes that may mediate the impact of thinking about reasons. One possibility is that thinking about reasons increases participants' awareness of their excessive liking for name letters, and subsequently instigates attempts to suppress or correct this bias. Such a correction account supposes that people (a) possess some awareness of their bias for name letters and (b) have appropriate lay theories about the size and direction of their bias for name letters (Strack & Hannover, 1996; Wegener & Petty, 1997; Wilson & Brekke, 1994). It should be noted that previous findings by Nuttin (1985) cast considerable doubt on the validity of these assumptions. Specifically, Nuttin challenged 100 participants to discover a hidden structure in the stimulus lists of the kind he had previously used to demonstrate a positive bias for name letters. These stimulus lists contained participants' names printed in spelling order from bottom to top. Even though there was no time limit and a very high monetary reward was offered, nobody succeeded in discovering any name structure in the stimulus lists. These findings suggest that people have very little conscious access to their potential bias for own name letters. Notwithstanding these considerations, we decided to test empirically for the possible operation of correction processes in our reasons manipulation. To this end, we checked whether our participants showed any awareness of their possible bias towards own name letters and examined whether this awareness was systematically different between the experimental conditions. In addition, we investigated whether participants' positive bias for name letters became reversed in the reasons condition, because such contrast effects are often tell-tale signs of (overzealous) correction processes (Greenwald & Banaji, 1995; Stapel, Koomen, & Zeelenberg, 1998; Wilson & Brekke, 1994).

Another, more plausible, possibility is that thinking about reasons instigates deliberative overriding of implicit self-esteem effects relatively automatically, in the absence of participants' awareness that they are displaying implicit self-esteem. In their affective priming studies, Murphy and Zajonc (1993) identified a form of deliberative overriding that meets these requirements, which they referred to as *dilution.* As conceived by Murphy and Zajonc (1993), the dilution process consists of simply adding more evaluations to one's initial evaluation. Accordingly, dilution may remove an implicit affective reaction without any awareness of its occurrence, and without a specific motivation to appear unbiased. Although much slower than the automatic evaluation process (which may already unfold within a mere 4 ms; cf. Murphy & Zajonc, 1993), dilution may still occur rather quickly (i.e., within a second; cf. Murphy & Zajonc, 1993). A similar dilution process is highly compatible with Wilson, Lindsey, and Schooler's (2000) notion of automatic overriding through thinking about reasons. According to Wilson et al. (2000), thinking about reasons leads people to generate a variety of evaluations of an attitude object, which then serve to moderate people's initial affect toward the object (Wilson et al., 1989). To explore the possible operation of dilution processes, we examined participants' open-ended responses regarding their stimulus perceptions at the end of the experiment. We predicted that participants in the reasons condition would display more cognitively differentiated appraisals of the stimuli than participants in the feelings condition. Moreover, we hypothesized that these perceptual differences would mediate the

inhibiting impact of the reasons manipulation on enhanced name letter liking.

Finally, Study 2 sought to obtain more evidence for the convergent validity of relative name letter evaluations as a measure of implicit self-esteem. For this purpose, we simultaneously assessed evaluative responses towards name letters and birthdate numbers, so that we were also able to examine the association between overevaluation of name letters and overevaluation of birth date numbers (cf. Kitayama & Karasawa, 1997). If, as we have argued, these phenomena both tap into automatic feelings of self-affection, one would expect them to be strongly associated. This association might be more pronounced for participants in the feelings condition, given that thinking about reasons may distort the association between the two forms of implicit self-esteem. Finally, overevaluation of birthdate numbers was expected to become inhibited among participants who were analyzing reasons in the same way as overevaluation of name letters was affected by this manipulation.

Method

Participants and Design

Forty undergraduate students from the University of Nijmegen (3 male and 37 female, average age 20) were randomly assigned to two experimental conditions (feelings vs. reasons). Participants received 5 guilders (approximately $2) for their participation.

Procedure

As they arrived in the laboratory, participants were led to individual cubicles, each containing an Apple Macintosh computer. The experimenter explained that all the instructions would be administered by a computer-program and left the cubicle. Participants then started the program by pressing a button on the keyboard. The computer randomly assigned participants to one of two experimental conditions.

Participants first received a brief instruction regarding the use of the computer. After answering a personality questionnaire,[4] participants were introduced to the letter evaluation task, which was described as a study of aesthetic preferences. As in the previous studies, participants were asked to evaluate each letter of the alphabet on 5-point scales (1 = *not at all beautiful,* 5 = *extremely beautiful*). Letters of the alphabet appeared separately on screen and were presented in a random order. At this point, induction of a focus on feelings or on reasons was introduced. This manipulation resembled instructions used in previous research (Epstein et al., 1999; Simon et al., 1997; Wilson, Dunn, Bybee, Hyman, & Rotondo, 1984). Participants in the feelings condition were informed that "Previous research has shown that the study of these kinds of judgments can lead to a better understanding of certain aspects of human emotions." In addition, they were encouraged to give their first, intuitive impression of each letter. Participants in the reasons condition were informed that "Previous research has shown that the study of these kinds of judgments can lead to a better understanding of certain aspects of human reasoning processes." In addition, they were encouraged to reason why they found certain letters more beautiful than others and to analyze carefully which features of the letters they liked or did not like. Finally, they were told that the experimenter would later ask them to explain in what manner they had evaluated the stimuli. This latter instruction was added to ensure that participants were motivated to engage in more effortful, deliberative processing (De Dreu, Koole, & Steinel, 2000).

After evaluating all the letters of the alphabet, participants were asked to evaluate the numbers 1 through 50 on 5-point scales (1 = *not at all beautiful,* 5 = *extremely beautiful*). Before evaluating the numbers, the instructions designed to manipulate a focus on feelings or on reasons—adapted to the number evaluation task—were repeated. The numbers 32 through 50 were included to obscure the purpose of the experiment (cf. Kitayama & Karasawa, 1997). As in the letter evaluation task, the numbers appeared separately on screen and were presented in a random order.

On completion of the number evaluation task, participants were asked several general background questions regarding their age, gender, and college major. In addition, they were requested to write down their full names on a consent form to indicate that they agreed that their data be used for scientific purposes. Next, they were asked to leave their cubicle and return to the main room. There, the experimenter asked participants to fill out a questionnaire. This questionnaire consisted of two questions regarding the manner in which participants had judged the letter and number stimuli. The first question was, "In what way did you judge the numbers and letters?" The second question was, "Did you pay attention to any particular characteristics of these numbers/letters? If so, please indicate what they were." After they had completed this questionnaire, participants were debriefed, paid, and dismissed.

Construction of Dependent Variables

Evaluations of Self-Associated Letters and Numbers

Preliminary analyses indicated that baseline no-name-letter and no-birthdate number evaluations did not differ between the two experimental conditions, $ps > .23$. Consequently, base line evaluations were aggregated across both conditions for each respective letter and number. Relative evaluations of name letters and birthdate numbers were again computed according to the Kitayama and Karasawa (1997) procedure.

Open-Ended Responses

Number of words. Participants' handwritten answers were first entered into a word processing program. Using this program, we counted the number of words that each participant had written.

Coding system. To examine participants' qualitative responses, we developed a coding system. This system included two categories for statements concerning the general manner in which the stimuli were evaluated: (a) feelings (statements that one's evaluations were based on feelings or first reactions), and (b) reasons (statements about having thought about the reasons for one's evaluations). These categories were included as a manipulation check. In addition, the coding system included two categories of perceptual details of the stimuli that were mentioned by the participants: (a) mentioning round features of the experimental stimuli, and (b) mentioning straight features of the experimental stimuli (i.e., line pieces or angular shapes). Participants also mentioned other characteristics

[4] During this questionnaire study, participants answered a Dutch translation of the Rational-Experiential Inventory (REI). The REI consists of two unipolar scales. One scale, consisting of 19 items from the Need for Cognition (NFC) scale, measured the tendency to engage in rational thinking. This scale included items like "I would prefer complex to simple problems." The other scale, consisting of 12 items from the Faith in Intuition (FI) scale, measured the tendency to engage in experiential thinking. This scale included items like "I believe in trusting my hunches." The items appeared in a random order on the computer screen and were scored on 11-point scales (1 = *disagree completely,* 11 = *agree completely*). Previous research has shown that the NFC and FI scales of the REI are valid instruments in assessing individual differences in analytical-rational and intuitive-experiential thinking styles (Epstein, Donovan, & Denes-Raj, 1999). Although the NFC and FI scales had adequate reliabilities (their respective alpha's were .89 and .70), subsequent analyses did not reveal any significant effects. These scales were hence excluded from the analyses reported in the main body of this article.

of the experimental stimuli (e.g., sound, symmetry), but the frequencies of statements about these characteristics were too low to merit further analysis (they were mentioned by less than 5% of the participants). The final two categories included statements about having thought about own name letters or own birthdate numbers, or both. Two independent raters rated the presence or absence of each of the five categories in each participants' responses. Both raters were blind to the experimental conditions. Overall interrater agreement was 97%. Differences between raters were resolved through discussion.

Results

Manipulation Check

Most participants (60%) indicated that their feelings had been important in determining their evaluations of the experimental stimuli. Unexpectedly, there was no significant effect of the experimental manipulation, $\chi^2(1) < 1$. Thus, participants in the reasons condition were no less likely to indicate that they based their evaluations on their feelings than participants in the feelings condition.

Furthermore, a minority of the participants (20%) indicated that they had been thinking about reasons in determining their evaluations. A one-way between-subjects ANOVA showed that participants in the reasons condition were more likely to indicate that they based their evaluations on reasons than participants in the feelings condition, $\chi^2(1, N = 40) = 4.27, p < .04$ (32% vs. 6% of the respective groups). Although this difference was in line with expectations, it should be noted that only a minority of the participants in the reasons condition indicated that they had been thinking about reasons. Conceivably, participants felt little need to refer to their logical thinking style because they knew that the experimenter was already aware of their thinking style (Grice, 1975).

Evaluations of Self-Associated Letters and Numbers

Evaluations of self-associated letters and numbers were analyzed in a 2 (feelings vs. reasons: between-subjects) × 2 (name letters vs. birthdate numbers: within-subject) ANOVA. Relevant means are displayed in Table 1. Consistent with predictions, both name letters and birthdate numbers were evaluated more positively by participants in the feelings condition than by participants in reasons condition, $F(1, 38) = 7.91, p < .01$ (combined Ms = 0.46 against −0.06). The interaction between instruction set and stimulus type was not reliable, $F(1, 38) = 1.14, p = .291$. Separate tests revealed that evaluations of name letters differed significantly from no-name letter baseline evaluations for participants in the feelings condition, $F(1, 17) = 12.09, p < .003$ ($M = 0.42$), but not for participants in the reasons condition, $F(1, 21) < 1$ ($M = 0.06$). Likewise, evaluations of birthdate numbers differed significantly from no-birthdate numbers baseline evaluations for participants in the feelings condition, $F(1, 17) = 4.79, p < .05$ ($M = 0.50$), but not for participants in the reasons condition, $F(1, 21) < 1.1$ ($M = -0.17$). It is important to note that participants in the reasons condition did not display a negative bias for name letters and birthdate numbers. Thus, our findings showed no evidence for a contrast effect among participants in the reasons condition.

Table 1
Relative Name Letter and Birthdate Number Evaluations As a Function of Instructional Set (Study 2)

	Instructional set			
	Feelings		Reasons	
Evaluation type	*M*	*SD*	*M*	*SD*
Name letter	0.42	0.51	0.06	0.48
Birthdate number	0.50	0.96	−0.17	0.88

Next, a series of correlational analyses was carried out to examine whether the instructions to focus on feelings or reasons had influenced the relationship between name letter evaluations and birth date number evaluations. Across the two experimental conditions, there was a significant relationship between relative name letter evaluations and relative birth date number evaluations, $r(40) = .35, p < .03$. However, this relationship was markedly stronger for participants in the feelings condition, for whom relative name letter evaluations and relative birthdate number evaluations were strongly correlated, $r(18) = .62, p < .007$. For participants in the reasons condition, no significant relationship between relative name letter evaluations and relative birth date number evaluations was observed, $r(22) = -.08, p > .70$. In addition, no significant correlations were observed between no-name letter evaluations and no-birthdate number evaluations.

As in Study 1, we sought to ascertain whether the correlation between name letter and birthdate number evaluations was inflated by a general tendency towards positive responding. Thus, we simultaneously regressed relative liking for no-name letters, liking for no-name letters, and liking for no-birthdate numbers on relative liking for birth date numbers. The results showed only a significant effect of relative name letter liking, $\beta = .66, t(36) = 3.15, p < .008$. Thus, the correlation between relative liking for name letters and relative liking for birthdate numbers could not be explained by a general tendency toward positive responding.

Open-Ended Responses

Number of Words

The number of words for each participant were analyzed in a oneway (feelings vs. reasons; between-subjects) ANOVA. This analysis revealed that participants in the reasons condition wrote somewhat more words about the manner in which they had evaluated the stimuli than participants in the feelings condition, $F(1, 38) = 3.65, p = .065$. Apparently, participants in the reasons condition had formed somewhat more elaborate verbal accounts of their judgment process than participants in the feelings condition.

Awareness of Name Letters or Birthdates

The majority of participants (93%) did not make any reference to own name letters or birthdate numbers in their protocols. However, 3 participants (7.5%; 2 in the feelings condition, 1 in the reasons condition) mentioned that they had thought of their own name while evaluating letters. Of these three, 1 participant remarked that she had also thought of her birthdate numbers while evaluating numbers. Neither the pattern nor the significance of the reported findings was changed when these 3 participants were removed from the analysis.

Attention to Perceptual Details of Self-Associated Stimuli

Turning to statements regarding perceptual details, a majority of the participants (63%) indicated that they had paid attention to round stimulus features. This percentage did not differ between experimental conditions, $\chi^2(1, N = 40) < 1$. In addition, a minority of the participants (33%) indicated that they had paid attention to straight stimulus features (i.e., line pieces and angles). Importantly, participants in the reasons condition were more likely to indicate that they based their evaluations on straight features than participants in the feelings conditions, $\chi^2(1) = 6.82, p < .01$ (50% vs. 11% of the respective groups). Thus, it appeared that both groups were equally likely to pay attention to round features, whereas participants in the reasons condition were especially likely to pay attention to lines pieces and angles. As such, these findings partially support the prediction that thinking about reasons leads to more differentiated stimulus appraisals.

Mediation Analysis

Finally, we conducted a series of path analyses to see whether enhanced attention to straight lines mediated the effect of thinking about reasons on overevaluation of self-associated stimuli. In this analysis, overevaluation of name letters and overevaluation of birthdate numbers were averaged into a single index (separate analyses yielded similar results). For attention to straight lines to qualify as a mediator, three conditions must be met (Baron & Kenny, 1986). First, the reasons manipulation must have a significant relationship on attention to straight lines. Consistent with this, the relationship between thinking about reasons and attention to straight lines was significant, $\beta = .41, t(38) = -2.80, p < .01$. Second, attention to straight lines must bear a significant relationship to the overevaluation of name letters and birth date numbers. Consistent with this, attention to straight lines was a (negative) predictor of overevaluation of self-associated stimuli, $\beta = -.51$, $t(38) = -3.64, p < .001$. Finally, the effect of thinking about reasons on implicit self-esteem should be eliminated or greatly diminished when attention to straight lines is controlled for. Before controlling for attention to straight lines, the effect of thinking about reasons on overevaluation of name letters and birth date numbers was $\beta = -.41$, $t(38) = -2.81$, $p < .01$. When the attention to straight lines variable was included in the analysis, however, the effect of thinking about reasons became become smaller and no longer significant, $\beta = -.25$, $t(37) = -1.64$, $p = .11$. By contrast, attention to straight lines remained predictive of overevaluation of self-associated stimuli, $\beta = -.41$, $t(37) = -2.71, p < .02$. Using Baron and Kenny's (1986) modification of the Sobel test (see Kenny et al., 1998), this mediation effect was found to be statistically significant, $Z = 1.60, p < .05$ (one tailed). Accordingly, variations in attention to straight lines indeed appear to have mediated the effect of thinking about reasons on implicit self-esteem.

Discussion

In Study 2, we investigated whether thinking about reasons can inhibit the occurrence of implicit self-esteem effects. As predicted, participants encouraged to rely on feelings clearly manifested implicit self-esteem, that is, a positive bias for name letters and birthdate numbers. In contrast, these forms of implicit self-esteem were no longer apparent among participants who were encouraged to reason why they felt the way they did. It thus appears that thinking about reasons inhibits the manifestation of implicit self-esteem effects. As such, these findings suggest that implicit self-esteem effects may be understood as automatic evaluations that can be overruled by more deliberative forms of processing (Wilson et al., 2000).

It is interesting to consider whether previous evidence for implicit self-esteem phenomena occurred similarly in the relative absence of analytic thought. From the perspective of CEST, this seems plausible because relying on one's initial affective reactions tends to be the default option in evaluating stimuli (Wilson et al., 1989). Moreover, it is possible that previous researchers inadvertently primed a focus on feelings by interacting in an informal manner with their participants (Simon et al., 1997). It is also noteworthy that Nuttin, who first documented people's tendency to overevaluate name letters, explicitly encouraged his participants to rely on their feelings in evaluating letters (Nuttin, 1985, p. 356). Thus, there are some indications that previously reported implicit self-esteem effects occurred similarly in the relative absence of deliberative processing.

Given that thinking about reasons was found to inhibit the expression of implicit self-esteem in participants' judgments of self-associated stimuli, it is important to determine which processes may have mediated this effect. On the basis of our theoretical analysis, we hypothesized that thinking about reasons might inhibit the expression of implicit self-esteem through a dilution process, that is, by causing participants to consider the experimental stimuli in a more differentiated manner (Murphy & Zajonc, 1993; Wilson et al., 2000). In line with this hypothesis, our protocol data showed that in the reasons condition, participants were more likely to consider to lines pieces and angular stimulus features in the experimental stimuli than in feelings condition. At the same time, however, the protocols showed no effects of thinking about reasons on the consideration of round stimulus features. Perhaps, attention to line pieces and angles is reflective of analytic stimulus appraisals, whereas attention to round features is reflective of a more holistic, Gestalt-like focus. This tentative suggestion must await further research. In any case, mediation analyses showed that greater consideration of line pieces and angles was able to account for the inhibition of implicit self-esteem effects among participants in the reasons condition. Accordingly, these results support the idea that thinking about reasons reduced implicit self-esteem effects through a dilution process, during which more reasoned stimulus appraisals were added to participants' implicit feelings of self-affection.

Could it still be that some of the effects of thinking about reasons were the result of a correction process? Several findings seem to argue against such an interpretation. First, there was no evidence for a contrast effect in participants' evaluations in the reasons condition. Given that such contrast effects are often an important signature of correction processes (Stapel et al., 1998; Wegener & Petty, 1997; Wilson & Brekke, 1994), the lack of such effects is suggestive of a relative absence of correction. In addition, our protocol data revealed little evidence that participants were aware of any relation between the experimental stimuli and the self. A critic might argue that participants may have been reluctant to reveal their awareness of this relation. However, the

previously described findings by Nuttin (1985) render this argument much less credible, considering that Nuttin's participants were unable to report any relation between the experimental stimuli and self even when they were given a strong incentive to come up with such an answer. Finally, a correction account does not predict a priori that more differentiated stimulus appraisals should mediate the effects of the reasons manipulation. Taken together, it seems unlikely that our reasons manipulation instigated correction among our participants. It may be noted that this conclusion is consistent with theorizing by other researchers who have studied the effects of thinking about reasons (see Wilson et al., 2000).

Finally, a high correlation between overevaluation of name letters and of birthdate numbers was obtained among participants in the feelings condition. This finding is of substantial theoretical interest, because they provide further evidence that relative liking for name letters possess convergent validity as a measure of implicit self-esteem. Previous work has found that, among bilinguals, overevaluation of name letters is significantly correlated between different alphabets (Hoorens et al., 1990; Hoorens & Todorova, 1988). The present results extend these findings in an important way by showing that implicit self-esteem in evaluations of different self-attributes (i.e., name letters and birthdate numbers) is substantively correlated. Thus, our confidence is bolstered that overevaluation of name letters is not merely reflective of an evaluative bias for a single self-attribute but driven by more general feelings of self-affection.

STUDY 3

In Study 3, we sought to extend our analysis to the relation between implicit self-esteem phenomena and explicit self-evaluations. According to the current theorizing, this relation should be rather complex and highly dynamic. In line with the model of dual attitudes (Wilson et al., 2000), we assume that people's explicit self-evaluations are influenced by the self-evaluations that are most accessible at the time of reporting. Upon encountering a cue that is associated with self, people may automatically activate their implicit self-evaluations, given that these evaluations are well-practiced and overlearned. During subsequent stages, people may effortfully retrieve their explicit self-evaluations that are complex and highly differentiated (Hixon & Swann, 1993; Swann et al., 1990). Consistent with the model of dual attitudes, these explicit self-evaluations may then override people's more automatic, implicit self-evaluations. However, this entire sequence may be truncated whenever people lack the capacity or motivation to engage in conscious self-reflection. In the latter type of situation, the person's effortful, explicit self-evaluations will not be retrieved from memory, allowing the person's implicit self-evaluations to gain access to phenomenal experience (cf. Wilson et al., 2000).

To test these ideas, we examined the degree of congruence between implicit self-esteem and self-reported self-evaluation as a function of the ability to engage in conscious self-reflection. As before, we assessed participants' implicit self-esteem by determining their affective bias towards own name letters (Greenwald & Banaji, 1995; Kitayama & Karasawa, 1997; Nuttin, 1985, 1987). Explicit self-evaluation was operationalized as the degree of positivity of participants' dichotomous (i.e., applies to me or not) trait endorsements. The latter operationalization was chosen because dichotomous decisions allow for variation in cognitive elaboration in self-evaluation than conventional rating scales (Paulhus et al., 1989; Paulhus & Levitt, 1987). To study variations in the ability to engage in deliberative self-analysis, we capitalized on naturally occurring differences between participants in the speed of supplying self-evaluations. Previous work has shown that deliberative processing is more time consuming than automatic processing (Bargh, 1994). Thus, quickly rendered self-evaluations are presumably more likely to be influenced by automatic self-evaluations, whereas slowly rendered self-evaluations are presumably more likely to be influenced by deliberative self-evaluations (Paulhus, 1993).

As we noted before, our main interest was in the congruence between implicit and explicit self-esteem. We expected that slow explicit self-evaluations would show low congruence with implicit self-esteem. In contrast, we expected that fast explicit self-evaluations would show high congruence with overevaluation of name letters. Finally, because previous work has shown that self-enhancement tendencies become more apparent under automatic conditions (Paulhus et al., 1989; Paulhus & Levitt, 1987; Swann et al., 1990), we expected quickly rendered self-evaluations to be generally more self-enhancing than slowly rendered self-evaluations.

Method

Participants and Design

Participants were 54 undergraduate students from the University of Nijmegen (23 male and 31 female, average age 22). The experimental design had two between-subject factors, namely overevaluation of own name letters (high vs. low) and response time for positive traits (fast vs. slow). Participants received 5 guilders (approximately $2) for their participation.

Procedure

As participants arrived in the laboratory, they were led to individual cubicles and told that all the instructions would be administered by a computer program. After the experimenter left, participants started the program by pressing a button. Participants were first informed that the investigation would comprise two unrelated studies and received a brief instruction regarding the use of the computer. Participants then moved on to the first study, which consisted of the same letter evaluation task that was used in Study 2. Next, participants began the second task, described as a "study concerning the way in which people judge traits." They were informed that a number of trait words would be presented individually on the computer screen. Participants were to decide as quickly as possible whether they themselves possessed the trait or not. Participants responded by pushing either a "me or a "not me" button on the keyboard and were asked to keep their hands near the buttons throughout the task.

The trait rating task consisted of 30 trials. During each trial, a trait word appeared in the center of the computer screen, which remained there until participants had pushed a response button. Approximately 400 ms later, the next trait word appeared on the screen. In 15 cases, the target word was a positively valenced trait (e.g., *creative*), whereas in the remaining 15 cases, the target word was a neutrally valenced trait word (e.g., *emotional*). These words were selected on the basis of their social desirability ratings from a comprehensive list of 1,203 Dutch trait words (Brokken, 1978) and pretested for word valence in a pilot study. The presentation order of the items was randomized for each participant by means of appropriate software. After completing the trait rating task, participants were asked several

general background questions and were requested to write down their full names on a consent form. Finally, participants were debriefed, paid, and dismissed.

Results

Name Letter Evaluations

Relative name letter evaluations, computed as in Study 1, were positive, $F(1, 53) = 10.04$, $p < .005$ ($M = 0.24$). Thus, participants again displayed a highly reliable bias for own name letters. Participants were classified as having high or low relative name letter evaluations on the basis of a median split on their relative name letter evaluations.[5]

Response Latencies

For each participant, the computer recorded both the response (i.e., applies to me or not) and the latency of the response for each of the presented items. To reduce skewness of the distribution, response latencies longer than 3 standard deviations from the mean were excluded from the analyses. The remaining response times to positive traits were averaged into one index (Cronbach's alpha = .82), as were the response times to neutral traits (Cronbach's alpha = .68). Response times to positive traits were faster than response times to neutral traits, $F(1, 53) = 45.29$, $p < .001$ ($M = 1{,}130$ ms vs. $M = 1{,}299$ ms). In spite of this difference, response times to positive traits were strongly correlated with response times to neutral traits, $r(54) = .76$. Because response times were conceived as a situational variable (i.e., processing pace), and not as a general individual difference variable, response times to positive traits and response times to neutral traits were treated separately in subsequent analyses. Participants were classified as fast or slow positive responders on the basis of a median split on each participants' mean response time to positive traits. Similarly, participants were classified as fast or slow neutral responders on the basis of a median split on each participants' mean response time to neutral traits.

Trait Endorsements

The proportion of positive traits endorsed was computed by dividing the number of positive trait endorsements by the total number of positive traits that were presented. Likewise, proportion of neutral traits endorsed was computed by dividing the number of neutral trait endorsements by the total number of neutral traits that were presented. The resulting values were analyzed in a 2 (response time to positive traits: fast vs. slow) × 2 (relative name letter evaluations: high vs. low) × 2 (trait type: positive or neutral) mixed-model ANOVA with repeated measures on the third factor. The three-way interaction between response time, name letter evaluations, and trait type fell short of significance, $F(1, 50) = 2.32$, $p = .134$. Nonetheless, given the specificity of our hypotheses, we proceeded by analyzing the results for each trait type separately.

First, positive trait endorsements were analyzed in a 2 (relative name letter evaluations: high vs. low) × 2 (response time to positive traits: fast vs. slow) between-subjects ANOVA. Relevant means are displayed in Table 2. The analysis revealed that fast positive responders endorsed more positive traits than slow positive responders, $F(1, 50) = 13.19$, $p < .002$ ($M = 0.80$ vs. $M = 0.67$). In addition, participants with high relative name letter evaluations endorsed more positive traits than participants with low relative name letter evaluations, $F(1, 50) = 4.58$, $p < .04$ ($M = 0.76$ vs. $M = 0.70$). These main effects were, however, qualified by the predicted interaction between relative name letter evaluations and response time, $F(1, 50) = 4.20$, $p < .05$. Simple effects analysis showed that relative name letter evaluations were predictive of the proportion of positive trait endorsements among fast positive responders, $F(1, 50) = 7.62$, $p < .009$, but not among slow positive responders, $F(1, 50) < 1$. This finding was corraborated by a correlational analysis, which showed that relative name letter evaluations were strongly correlated with positive trait endorsements among fast positive responders, $r(25) = .51$, $p < .001$, but not among slow positive responders, $r(29) = -.06$, *ns*. Another way to interpret this interaction is to note that higher speed of positive responding increased the number of positive trait endorsements among participants with high name letter evaluations, $F(1, 50) = 15.04$, $p < .001$, but not among participants with low name letter evaluations, $F(1, 50) = 1.22$, $p = .28$. Analyses of positive trait endorsements using response times to neutral traits as an independent variable revealed no significant effects, all $ps > .13$.

Table 2
Proportion of Positive and Neutral Traits Endorsed as a Function of Relative Name Letter Evaluations and Response Time to Positive Traits (Study 3)

	Response time to positive traits			
	Fast		Slow	
Trait	High NLE	Low NLE	High NLE	Low NLE
Positive	.88	.73	.68	.67
Neutral	.47	.50	.55	.55

Note. NLE = name letter evaluations.

To investigate the specificity of our findings, we conducted the same analyses described above, only this time we substituted no-name letter evaluations for relative name letter evaluations. None of the effects involving no-name letter evaluations were significant, all $ps > .20$. It thus appeared that the obtained pattern of findings was specific to name letter evaluations. We also conducted a series of regression analyses using relative name letter evaluations and response times as continuous variables. The results of these analyses were highly similar to those yielded by the ANOVA approach. Finally, parallel analyses using neutral trait endorsements as a dependent variable yielded no significant results, all $ps > .40$.

[5] In both Studies 3 and 4, relative name letter evaluations were also analyzed as a continuous variable in a series of regression analyses. The results of these analyses strongly converged with the ANOVA results reported in the main text of this article. Specifically, in Study 3, the critical interaction between relative name letter evaluations and response time was significant at $\beta = -.23$, $t = -2.37$, $p < .03$. In Study 4, the critical interaction between relative name letter evaluations and cognitive load was significant at $\beta = .10$, $t = 2.09$, $p < .05$. Thus, our findings did not depend on the specific analytic strategy that was followed.

Discussion

In Study 3, we examined the degree of congruence between implicit and explicit self-evaluation as a function of naturally occurring variations in speed of self-evaluation. As expected, measures of implicit and explicit self-esteem were congruent only for participants who rendered their positive self-evaluations relatively quickly. Quick responders presumably lacked the time to engage in extensive self-reflection, so that they were forced to rely on their more automatic, implicit self-evaluations. In contrast, slow responders may have had ample opportunity to engage in deliberative self-analysis, so that their deliberative self-evaluations were able to override their more automatic, implicit self-evaluations.

It is notable that shorter response latencies for positive self-evaluations were also generally associated with more positive self-evaluations. This positivity effect did not appear to reflect a simple tendency towards more yea-saying among quick evaluators (Knowles & Condon, 1999), because response latencies for positive self-evaluations were uncorrelated with neutral trait endorsements. As such, the obtained positivity effect is in line with earlier findings that reductions in the ability to engage in deliberative self-reflection lead to increased self-positivity (Paulhus, 1993). The observation that effects of speed of self-evaluation parallelled previous effects of cognitive load further supports our use of processing times as an operationalization of ability to engage in deliberative self-evaluation (Dijksterhuis & van Knippenberg, 1995).

Although the findings of Study 3 support our conceptual analysis, some important ambiguities remain. In particular, it is difficult to rule out the possibility that speed of self-evaluation was related to a chronic personality trait rather than to a situation-specific ability to engage in deliberative self-evaluation. Arguing against the former interpretation, the moderating role of evaluation latencies was found to be specific to evaluation times for positive traits. Thus, any personality characteristic that can explain our findings would have to involve a very specific pattern of responding differentially toward positive versus neutral traits. Moreover, it should be noted that our conceptual account is not at odds with the possibility that some individual difference variable (e.g., those captured by Paulhus's, 1994, self-deception scale) is the distal personality variable which causes certain individuals to evaluate themselves less deliberatively than others. Nevertheless, it remains desirable to establish whether varying degrees of deliberative self-evaluation were the critical mediator of our findings.

Another ambiguity is that quick positive responders may have been less motivated to engage in deliberative self-evaluation rather than or in addition to being less able to engage in deliberative processing (Kruglanski & Webster, 1996). Although conceptually distinct from cognitive capacity, lack of motivation to engage in deliberative self-evaluation may be functionally equivalent to a reduction in cognitive capacity by undermining people's tendency to engage in deliberative self-evaluation. Thus, it remains unclear whether reductions in cognitive capacity are sufficient to increase congruence between implicit and explicit self-evaluation. Study 4 was carried out to resolve these ambiguities.

STUDY 4

In Study 4, we sought to replicate and extend Study 3's primary findings, this time using an experimental manipulation of processing capacity. In particular, we hypothesized that depriving participants of their cognitive resources would have similar effects as speed of self-evaluation in Study 3. Thus, participants under high cognitive load were expected to show greater congruence between their implicit and explicit self-evaluations than participants under low cognitive load. Moreover, cognitive load was expected to cause enhanced positivity in participants' self-descriptions (Paulhus, 1993; Swann et al., 1990).

As in Study 3, participants were requested to evaluate themselves on a number of personality traits. To extend the generality of our findings, we used a different set of traits and also included negative traits. Similar to our predictions for positive self-evaluations, we might predict that cognitive load would cause decreased negativity in self-evaluation (Paulhus et al., 1989; Paulhus & Levitt, 1987) and greater congruence between implicit and explicit self-evaluations. It should be noted, however, that past research has shown that the psychological domains of positive and negative valence are not necessarily each other's psychological mirror image (Cacioppo & Berntson, 1994; Taylor, 1991; Watson & Tellegen, 1985). Most relevant here, the existence of positive–negative asymmetries has recently been documented in the area of self-enhancement biases (Hoorens, 1995) and implicit self-esteem (Nuttin, 1987). Thus, the extension of Study 3's findings to negatively valenced self-attributes may be less than straightforward.

In Study 4, we varied participants' cognitive capacity by asking them to hold either an eight-digit number or a single-digit number in memory during the self-evaluation task. Similar tasks have been used successfully in past research to manipulate processing load (Paulhus et al., 1989; Swann et al., 1990).

Method

Participants and Design

Participants were 50 undergraduate students from the University of Nijmegen (17 male and 33 female, average age 21). The experimental design had two between-subject factors, namely overevaluation of own name letters (high vs. low) and cognitive load (high vs. low). Participants received 5 guilders (approximately $2) for their participation.

Procedure

When participants arrived in the laboratory, they were led to individual cubicles and told that all the instructions would be administered by means of a computer program. After the experimenter left, participants started the program by pressing a button. As in Studies 2 and 3, participants were informed that the investigation would contain two unrelated studies, and received a brief instruction regarding the use of the computer. Participants then moved on to the first study, which consisted of a letter evaluation task identical to the one used in Studies 2 and 3.

After this, participants began the second task, described as a "study concerning the way in which people judge traits." Participants were told that, during this study, they would perform two tasks simultaneously. One task consisted of a trait rating task similar to the one used in Study 1. To enhance the generalizability of our findings, we implemented several modifications in the trait rating task. First, a new set of 15 positive traits was used in this study. Second, the trait rating task included a set of 15 negatively valenced traits (e.g., jealous), again selected from Brokken's (1978) trait list. Thus, participants judged 15 positive traits, 15 negative traits, and 15 neutral traits. The presentation order of the items was again randomized for each participant. Participants were further asked to hold a

number in memory while they performed the trait rating task. Cognitive load was manipulated by varying the number of digits that participants were asked to hold in memory. Participants in the high load condition were asked to hold an eight-digit number in memory. Participants in the low load condition were asked to hold a single-digit number in memory. As a means of assessing compliance, participants were asked to type in the number assigned to them at the end of the trait rating task.

When participants completed the trait rating task, they were asked several general background questions and were requested to write down their full names on a consent form. Finally, participants were debriefed, paid, and dismissed.

Results

Name Letter Evaluations

Relative name letter evaluations, computed according to the Kitayama and Karasawa (1997) procedure, were higher than no-name letter baseline evaluations, $F(1, 49) = 9.14$, $p < .005$ ($M = 0.22$). Thus, participants again showed a positive affective bias towards own name letters. Participants were classified as having high or low relative name letter evaluations on the basis of a median split on mean relative name letter evaluations.

Trait Endorsements

As in Study 3, trait endorsements were converted to proportion scores for each trait type. A 2 (cognitive load: high vs. low) × 2 (relative name letter evaluations: high vs. low) × 3 (trait type: positive, negative, or neutral) mixed-model ANOVA with repeated measures on the third factor did not reveal the predicted three-way interaction between cognitive load, name letter evaluations, and trait type, $F < 1$. However, given the specificity of our hypotheses, we proceeded by analyzing the results separately for each trait type. First, the proportions of positive trait endorsements were analyzed in a 2 (cognitive load: high vs. low) × 2 (relative name letter evaluations: high vs. low) between-subjects ANOVA. Relevant means are displayed in Table 3. The analysis showed that participants under high cognitive load endorsed somewhat more positive traits than participants under low cognitive load, $F(1, 46) = 3.64$, $p = .06$ ($M = 0.84$ vs. $M = 0.79$). In addition, the predicted interaction between cognitive load and relative name letter evaluations emerged, $F(1, 46) = 6.55$, $p < .02$. Simple effects analysis showed that relative name letter evaluations were predictive of the proportion of positive trait endorsements for participants under high cognitive load, $F(1, 46) = 5.24$, $p < .03$, but not for participants under low cognitive load, $F(1, 46) < 1.10$. This finding was corroborated by a correlational analysis, which showed that relative name letter evaluations were strongly correlated with positive trait endorsements under high cognitive load, $r(23) = .48$, $p = .001$, but not under low cognitive load, $r(27) = -.15$, *ns*. Another way to interpret this interaction is to note that high cognitive load increased the number of positive trait endorsements among participants with high name letter evaluations, $F(1, 46) = 15.04$, $p < .005$, but not among participants with low name letter evaluations, $F(1, 46) < 1$.

Table 3
Proportion of Positive, Neutral, and Negative Traits Endorsed as a Function of Relative Name Letter Evaluations and Cognitive Load (Study 4)

	Cognitive load			
	High		Low	
Trait	High NLE	Low NLE	High NLE	Low NLE
Positive	.89	.79	.76	.81
Neutral	.56	.59	.55	.65
Negative	.11	.15	.20	.26

Note. NLE = name letter evaluations.

Second, the proportions of negative trait endorsements were analyzed in a 2 (cognitive load: high vs. low) × 2 (relative name letter evaluations: high vs. low) between-subjects ANOVA. This analysis showed that negative traits were endorsed less frequently by participants under high cognitive load than by participants under low cognitive load, $F(1, 46) = 3.98$, $p < .055$, ($M = 0.13$ vs. $M = 0.24$). However, there was no hint of an interaction between cognitive load and name letter evaluations, $F(1, 46) < 1$. Third, parallel analyses using neutral trait endorsements as a dependent variable yielded no significant results, all $ps > .15$.

To investigate the specificity of our findings, we conducted the same analyses as described above, only this time we substituted no-name letter evaluations for relative name letter evaluations. None of the effects involving no-name letter evaluations was significant, all $ps > .20$, suggesting that our findings were indeed specific to evaluations of self-associated stimuli. Finally, analogous regression analyses using relative name letter evaluations as a continuous variable yielded highly similar results.

Discussion

The pattern of findings in Study 4 again confirmed our predictions: implicit self-esteem, (i.e., relative liking for name letters) was predictive of positive self-evaluations under high cognitive load, but not under low cognitive load. As such, these findings are a conceptual replication of Study 3, which used naturally occurring variations in speed of self-evaluation to operationalize processing capacity. Taken together, the results of Study 3 and 4 provide converging support for the idea that congruence between implicit and explicit self-esteem becomes stronger when the capacity to engage in deliberative self-reflection is undermined.

Just as in Study 3, scarcity of processing resources was associated with a general increase in self-positivity. Our findings in Study 4 thus offer further support for previous research that automatic self-evaluation is generally more positive than the controlled evaluation of self (Paulhus, 1993; Swann et al., 1990). Also paralleling the results of Study 3, the positivity effect was qualified by participants' level of implicit self-esteem. That is, cognitive load increased positivity of explicit self-evaluations for participants with high implicit self-esteem, but not for participants with low implicit self-esteem. Thus, the positivity effect was again restricted to participants who felt good about themselves on an implicit level.

Although our predictions were clearly confirmed for positive self-evaluations, results on negative self-evaluations were more ambiguous. As in previous studies (Paulhus et al., 1989), cognitive load caused self-evaluations to become less negative. However, the predicted interaction between implicit self-esteem and cogni-

tive load was not obtained. It is possible that our selection of negative traits possessed certain idiosyncratic features. Alternatively, and on a more theoretical level, our findings may reflect a basic asymmetry between positive and negative evaluations. Previous work suggested that positive and negative evaluations may reflect partially independent psychological systems (Cacioppo & Berntson, 1994; Taylor, 1991; Watson & Tellegen, 1985). In this regard, it is interesting to note that overevaluation of name letters is typically much more robust when it is measured in positive terms (i.e., "How much do you like these letters?") than when it is measured in negative terms (i.e., "How much do you dislike these letters?"; Nuttin, 1987). Moreover, the most common form of implicit self-esteem appears to be enhancement of self-associated attributes rather than deprecation of self-dissociated attributes (Greenwald & Banaji, 1995). Accordingly, it may be that implicit self-esteem is most strongly related to a positive evaluative system. Pending further investigation, however, this suggestion must remain speculative.

GENERAL DISCUSSION

In the present research, we sought to identify some of the links between implicit self-esteem phenomena and the automatic self. Previous work has shown that the automatic self is highly positive (Paulhus, 1993; Swann et al., 1990). In line with this, four separate studies found evidence for exaggerated name letter liking, confirming that implicit self-esteem is generally positive (Greenwald & Banaji, 1995). The automatic self has further been described as the highly practiced self (Paulhus, 1993), implying a certain amount of stability over time. Supporting this, Study 1 showed that enhanced name letter liking is temporally stable, at least across a 4-week period. The automatic self has also been described as occurring without deliberative thought (Paulhus, 1993). Consistent with this, Study 2 showed that bias for self-associated stimuli becomes inhibited when people are induced to respond in a deliberative manner. Finally, the automatic self has been known to emerge when people are giving mindless self-reports or are undergoing stress (Paulhus, 1993). In agreement with this, Studies 3 and 4 found evidence that implicit self-evaluations only correspond to reported self-evaluations when people lack the time or the cognitive resources to engage in deliberative self-reflection. Taken together, these findings support the notion that implicit self-esteem constitutes an integral part of the automatic self.

By linking implicit self-esteem to the automatic self, the present research adds to the growing support for automaticity in self-evaluation (Baldwin, Carrell, & Lopez, 1990; Brown, 1993; Epstein & Morling, 1995; Gilbert, Giesler, & Morris, 1995; Greenwald & Banaji, 1995; Leary & Downs, 1995). As such, classic notions of self-evaluation as a process of effortfully peering inward may be in need of revision. Indeed, when conscious attention is directed elsewhere, people's more implicit, automatic self-evaluations may be activated. Although these self-evaluations may be experienced as vague and irrational, this does not prevent automatic self-evaluations from exerting an important influence on people's thoughts, feelings, and behaviors. In particular, they may lead people to be biased toward objects or people that have become associated with themselves (Greenwald & Banaji, 1995), no matter how trivial these objects or people are. Even beyond this, implicit self-evaluations may influence the subjective experience of the self when people are not actively engaging in self-reflective thought.

It is important to note that we do not suggest that conscious self-reflection never influences people's self-evaluations. To the contrary, the current analysis suggests that conscious self-reflection may exert a profound influence on the self-evaluation process. Specifically, Study 2 showed that introspective efforts may dilute implicit self-esteem in evaluations of self-associated stimuli, even when people are unaware of the influence that their implicit self-esteem exerts on their judgments. Furthermore, Studies 3 and 4 found a dissociation between implicit and explicit self-evaluations under circumstances that permitted deliberative processing. Although these latter studies did not obtain direct measures of the operation of conscious self-reflection, research by Swann, Hixon, and their colleagues has illuminated these processes. In particular, their work has shown that during conscious reflection, self judges may effortfully retrieve specific self-conceptions from memory (Swann et al., 1990) and thus cause self-evaluations to sample from the rich knowledge base of autobiographic memory (Andersen, Glassman, & Gold, 1998; Hixon & Swann, 1993). Thus, we may speculate that during conscious self-reflection, people may access a more inclusive selection of information about the self which subsequently dilutes their implicit self-evaluations that are chronically accessible. The findings of Study 2 fit this notion by demonstrating that explicit knowledge can serve to dilute the influence of implicit self-evaluations in judgments of self-associated stimuli.

We are inclined to regard dilution processes as the main cause of dissociations between implicit and explicit self-evaluations. This is because dilution requires few preconditions other than that judges are able to include evaluative knowledge into their judgments that goes beyond their initial (and more implicit) self-evaluations. Dilution processes thus qualify as a relatively "automatic" form of deliberative overriding, which may unfold already within a single second (Murphy & Zajonc, 1993). However, dilution is by no means the only potential source of dissociations between implicit and explicit self-evaluations. In their model of dual attitudes, Wilson et al. (2000) argued that dissociations between implicit and explicit evaluations may occur when people correct for their implicit evaluations, when they seek to repress them, or when implicit and explicit evaluations function completely independent of each other. In principle, each of these mechanisms may be applicable to the self-evaluation process. For instance, a person may correct his or her highly positive implicit self-evaluations when he or she feels that the situation calls for modesty (Tice, Butler, Muraven, & Stillwell, 1995). Alternatively, a person may resort to repression when implicit self-evaluations are predominantly negative or otherwise threatening (Baumeister & Cairns, 1992; Paulhus, Fridhandler, & Hayes, 1997). Finally, it is possible that certain implicit self-evaluations have become so automated that they are never consciously experienced. Further study into these possible sources of dissociations between implicit and explicit self-evaluations provides an important avenue for future research.

It is interesting to note that the issue of how implicit and explicit self-evaluations come to be dissociated has important implications for how such dissociations are interpreted. In so far as such dissociations result from repression, they may be interpreted as signs of psychological conflict, or even as symptomatic of a

personality disorder. In as far as dissociations result from correction, dissociations may point to the presence of socially undesirable self-evaluations (e.g., excessive egotism). Finally, in so far as dissociations result from functional independence or dilution processes, dissociations may be seen as a natural outcome of the way the self system is designed. Given these various alternative interpretations of such dissociations, it makes little sense to brand dissociations between implicit and explicit self-evaluations as "abnormal" or "biased." At least for the time being, it seems more useful to approach implicit and explicit self-esteem without making a priori assumptions regarding the functionality of specific configurations of the different types of self-evaluation.

Limitations and Future Perspectives

The present research relied mainly on the assessment of overevaluation of name letters as an indicator of implicit self-esteem. Our decision to use this measure was deliberate, because overevaluation of name letters is currently one of the most well-documented forms of implicit self-esteem (Hoorens, 1990; Kitayama & Karasawa, 1997; Nuttin, 1985, 1987; Pelham et al., 2000). The convergent validity of name letter evaluations as a measure of implicit self-esteem was further bolstered by the current findings that this measure is meaningfully related to explicit self-esteem and to an alternative measure of implicit self-esteem, that is, overevaluation of birthdate numbers (Kitayama & Karasawa, 1997). Nevertheless, extending the current findings to alternative measures of implicit self-evaluation would be highly desirable.

Although the measurement of implicit self-evaluation is still in its infancy, useful methodologies are increasingly becoming available. One promising new paradigm is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwarz, 1998). The IAT can be adapted to the study of implicit self-evaluation by examining how well participants are able to pair words related to self or others with pleasant or unpleasant words (Farnham, Greenwald, & Banaji, 1999). Additional measures have been pioneered by Pelham, Hetts, and their colleagues (Hetts, Sakuma, & Pelham, 1999; Pelham & Hetts, 1999). In their studies, participants are first primed with identity related words (e.g., *I, me*) or sentences (e.g., *I am very sensitive to my inner thoughts and feelings*), after which accessibility of positive and negative evaluations is assessed. Using this logic, both word completion and reaction time measures have been shown to posses impressive predictive validity (cf. Hetts et al., 1999; Pelham & Hetts, 1999). Together, these methodologies offer new and flexible ways to investigate implicit self-evaluation.

Alternative methodologies such as the IAT and wordstem completion measures rely on self-related words to prime implicit self-evaluations and thus capitalize on implicit self-evaluations operating at the whole-word level. In contrast, the name letter measure delves into implicit self-evaluations at the subword level, by relying on responses to isolated name letters. Accordingly, the name letter measure may tap into somewhat more primitive implicit self-evaluations than implicit measures that presume whole-word analysis (Abrams & Greenwald, 2000). Nevertheless, the results obtained with whole-word measures of implicit self-esteem are generally consistent with the current framework. In particular, IATs, reaction time, and wordstem completion measures of implicit self-evaluation have shown low congruence with explicit self-evaluation (Farnham et al., 1999; Hetts et al., 1999; Pelham & Hetts, 1999). Considering that these studies obtained measures of explicit self-esteem under conditions that allowed for deliberative self-evaluation, these findings corroborate a key finding in the current research. In addition, Hetts et al. (1999) found evidence that immigrants' implicit self-evaluations may remain consistent with the norms of their native countries even after years of living in a new country, which is consistent with the presumed temporal stability of implicit self-esteem. Finally, Pelham and Hetts (1999) found that birth order (i.e., being a first-born son) and being born in close proximity to a holiday (i.e., Christmas) are systematically related to implicit self-esteem. The latter findings are consistent with our theorizing that implicit self-evaluations are habitualized forms of self-evaluation that may be acquired relatively early in life.

Even if implicit self-esteem may be formed early in life, this does not preclude that implicit self-esteem cannot be responsive to changes in the immediate situation. Indeed, recent research has identified several circumstances that can influence implicit self-esteem on a momentary basis. One recent study found that bias in name letter evaluations was absent after participants had failure feedback on an alleged IQ test and reemerged after subsequent affirmation of a personally important value (Koole, Smeets, van Knippenberg, & Dijksterhuis, 1999). Another relevant set of studies found that participants showed decreased first name letter evaluations following reminders of personal death (Koole, Dechesne, & van Knippenberg, 2000), a manipulation that presumably arouses feelings of threat to the self (Pyszczynski, Greenberg, & Solomon, 1999; Tesser, Martin, & Cornell, 1996). Additional support for the malleability of implicit self-esteem was reported by Pelham and Hetts (1999), who found that implicit self-esteem may be affected by recent affective experiences, such as graduating from college or reflecting on a negative life event. Overall, it appears that implicit self-esteem may be influenced by impactful affective experiences that are relevant to self (see Woike, 1995, for an interesting parallel in the area of implicit motivation). To explain how implicit self-esteem can be both durable and malleable, Pelham and Hetts (1999) suggested that implicit self-esteem may be highly resilient or "rubbery." Like a rubber toy, implicit self-esteem may be affected by immediate pressures in the environment, but return to its original state when these pressures are removed. The hypothesized resilience of implicit self-esteem is consistent with Wilson et al.'s (2000) model of dual attitudes, which states that implicit evaluations are like old habits, which only change after prolonged practice. Understanding the factors that cause long-term change in implicit self-evaluations provides a major challenge for future research.

Concluding Thoughts

People are often surprisingly partial when it comes to evaluating persons or objects that are in some way connected with self. The present studies suggest that the psychological significance of such signs of implicit self-esteem goes beyond their mere entertainment value. Indeed, implicit self-esteem phenomena appear linked to people's automatic self-evaluations, the self-evaluations that are most likely to emerge when people are giving mindless self-reports or are undergoing stress. Consequently, displays of implicit self-

esteem may betray feelings of self-affection that often remain hidden from even the self judge's own inward gaze.

References

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science, 11,* 118–124.

Andersen, S. M., Glassman, N., & Gold, D. A. (1998). Mental representations of the self, significant others, and nonsignificant others: Structure and processing of private and public aspects. *Journal of Personality and Social Psychology, 75,* 845–861.

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.

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.). Hillsdale, NJ: Erlbaum.

Bargh, J. A., & Barndollar, K. (1996). Automaticity in action: The unconscious as repository of chronic goals and motives. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 457–481). New York: Guilford Press.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic attitude activation with a pronunciation task. *Journal of Experimental Social Psychology, 32,* 104–128.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462–479.

Baron, R. M., & Kenny. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Baumeister, R. F. (1998). The self. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 680–740). New York: McGraw-Hill.

Baumeister, R. F., & Cairns, K. J. (1992). Repression and self-presentation: When audiences interfere with self-deceptive strategies. *Journal of Personality and Social Psychology, 62,* 851–862.

Baumeister, R. F., & Tice, D. M. (1986). How adolescence became the struggle for self: A historical transformation of psychological development. In J. Suls & A. G. Greenwald (Eds.), *Psychological perspectives on the self* (pp. 183–201). Hillsdale, NJ: Erlbaum.

Beggan, J. K. (1992). On the social nature of nonsocial perception: The mere ownership effect. *Journal of Personality and Social Psychology, 62,* 229–237.

Benenson, J., & Dweck, C. (1986). The development of trait explanations and self-evaluations in the academic and social domains. *Child Development, 57,* 1179–1187.

Bowlby, J. (1973). *Attachment and loss: Vol. 2. Separation: Anxiety and anger.* New York: Basic Books.

Bradley, G. W. (1978). Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology, 36,* 56–71.

Brokken, F. B. (1978). *The language of personality.* Unpublished doctoral dissertation. University of Groningen, the Netherlands.

Brown, J. D. (1993). Self-esteem and self-evaluation: Feeling is believing. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 4, pp. 27–58). Hillsdale, NJ: Erlbaum.

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

Brown, J. D., & Dutton, K. A. (1995). The thrill of victory, the complexity of defeat: Self-esteem and people's emotional reactions to success and failure. *Journal of Personality and Social Psychology, 68,* 712–722.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with the emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115,* 401–423.

De Dreu, C. K. W., Koole, S. L., & Steinel. (2000). Unfixing the fixed pie: A motivated information processing approach to integrative negotiation. *Journal of Personality and Social Psychology, 79,* 975–987.

Dijksterhuis, A., & van Knippenberg, A. (1995). Memory for stereotype-consistent and stereotype-inconsistent information as a function of processing pace. *European Journal of Social Psychology, 25,* 689–693.

Dovidio, J. F., Kawakami, K., & Beach, K. R. (in press). Implicit and explicit attitudes: Examination of the relationship between measures of intergroup bias. *Blackwell Handbook in Social Psychology: Vol. 4. Intergroup processes.*

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49,* 709–724.

Epstein, S., Donovan, S., & Denes-Raj, V. (1999). The missing link in the paradox of the Linda conjunction problem: Beyond knowing and thinking of the conjunction rule, the intrinsic appeal of heuristic processing. *Personality and Social Psychology Bulletin, 25,* 204–214.

Epstein, S., & Morling, B. (1995). Is the self motivated to do more than enhance and/or verify itself? In M. H. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 9–29). New York: Plenum.

Erikson, E. H. (1968). *Identity youth and crisis.* New York: Norton.

Eshel, Y., & Klein, Z. (1981). Development of academic self-concept of lower-class and middle-class primary school children. *Journal of Educational Psychology, 73,* 287–293.

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

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic evaluation of attitudes. *Journal of Personality and Social Psychology, 50,* 229–238.

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64,* 657–674.

Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. *Journal of Personality and Social Psychology, 69,* 227–236.

Greenwald, A. G. (1982). Is anyone in charge? Personalysis versus the principle of personal unity. In J. Suls (Ed.), *Psychological perspectives on the self* (pp. 151–181). Hillsdale, NJ: Erlbaum.

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

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

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). New York: Academic Press.

Hetts, J. J., Sakuma, M., & Pelham, B. W. (1999). Two roads to positive regard: Implicit and explicit self-evaluation and culture. *Journal of Experimental Social Psychology, 35,* 512–559.

Hixon, J. G., & Swann, W. B., Jr. (1993). When does introspection bear fruit? Self-reflection, self-insight, and interpersonal choices. *Journal of Personality and Social Psychology, 64,* 35–43.

Hoorens, V. (1990). Nuttin's affective selfparticles hypothesis and the name letter effect: A review. *Psychologica Belgica, 30,* 23–48.

Hoorens, V. (1995). Self-favoring biases, self-presentation, and the self-other asymmetry in social comparison. *Journal of Personality, 63,* 793–817.

Hoorens, V., & Nuttin, J. M. (1993). Overevaluation of own attributes: Mere ownership or subjective frequency? *Social Cognition, 11,* 117–200.

Hoorens, V., Nuttin, J. M., Erdélyi Herman, I., & Pavakanun, U. (1990). Mastery pleasure versus mere ownership: A quasi-experimental cross-

cultural and cross-alphabetical test of the name letter effect. *European Journal of Social Psychology, 20,* 181–205.

Hoorens, V., & Todorova, E. (1988). The name letter effect: Attachment to self or primacy of own name writing? *European Journal of Social Psychology, 18,* 365–368.

Johnson, M. M. S. (1986). *The initial letter effect: Ego-attachment or mere exposure?* Unpublished doctoral dissertation, Ohio State University.

Kenny, D. A., Kasher, D. A. & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, S. T. Fiske, & G. Lindsey (Eds.), *Handbook of social psychology* (4th ed., Vol. 1, pp. 233–265). New York: McGraw-Hill.

Kitayama, S., & Karasawa, M. (1997). Implicit self-esteem in Japan: Name letters and birthday numbers. *Personality and Social Psychology Bulletin, 23,* 736–742.

Knowles, E. S., & Condon, C. A. (1999). When people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77,* 379–386.

Koole, S. L., Dechesne, M., & van Knippenberg, A. (2000). *The sting of death: Evidence that reminders of mortality undermine implicit self-esteem.* Unpublished manuscript.

Koole, S. L., Smeets, K., Van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology, 77,* 111–125.

Koole, S. L., & van't Spijker, M. (2000). Overcoming the planning fallacy through willpower: Effects of implementation intentions on actual and predicted task completion times. *European Journal of Social Psychology, 30,* 873–888.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average-effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77,* 221–232.

Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing of the mind: "Seizing" and "freezing." *Psychological Review, 103,* 263–283.

Leary, M. R., & Downs, D. L. (1995). Interpersonal functions of the self-esteem motive: The self-esteem system as a sociometer. In M. Kernis (Ed.), *Efficacy, agency, and self-esteem* (pp. 123–144). London: Plenum.

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self.* New York: Plenum Press.

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist, 41,* 954–969.

McClelland, D. C. (1980). Motive dispositions: The merits of operant and respondent measures. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 1, pp. 10–41). Beverly Hills, CA: Sage.

Mikulincer, M. (1995). Attachment style and the mental representation of the self. *Journal of Personality and Social Psychology, 69,* 1203–1215.

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

Nuttin, M. J. Jr. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology, 15,* 353–361.

Nuttin, M. J. Jr. (1987). Affective consequences of mere ownership: The name letter effect in twelve European languages. *European Journal of Social Psychology, 17,* 381–402.

Otten, S., & Wentura, D. (1999). About the impact of automaticity in the minimal group paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology, 29,* 1049–1071.

Paulhus, D. L. (1993). Bypassing the will: The automatization of affirmations. In D. M. Wegner & J. W. Pennebaker (Eds.), *Handbook of mental control* (pp. 573–587).

Paulhus, D. L., Fridhandler, B., & Hayes, S. (1997). Psychological defense: Contemporary theory and research. In R. Hogan, J. Johnson, & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 543–579). London: Academic Press.

Paulhus, D. L., Graf, P., & van Selst, M. (1989). Attentional load increases the positivity of self-presentation. *Social Cognition, 7,* 389–400.

Paulhus, D. L., & Levitt, K. (1987). Desirable responding triggered by affect: Automatic egotism? *Journal of Personality and Social Psychology, 52,* 245–259.

Pelham, B. W. (1991). On confidence and consequence: The certainty and importance of self-knowledge. *Journal of Personality and Social Psychology, 60,* 518–530.

Pelham, B. W., & Hetts, J. J. (1999). *Implicit self-evaluation.* Unpublished manuscript.

Pelham, B. W., Mirenberg, M. C., & Jones, J. K. (2000). *Why Susie sells seashells by the seashore: Implications of implicit egotism for major life decisions.* Unpublished manuscript.

Pelham, B. W., & Swann, W. B. (1989). From self-conceptions to self-worth: On the sources and structure of self-worth. *Journal of Personality and Social Psychology, 57,* 672–680.

Pyszczynski, T., Greenberg, J., & Solomon, S. (1999). A dual-process model of defense against conscious and unconscious death-related thoughts: An extension of terror management theory. *Psychological Review, 106,* 835–845.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review, 96,* 341–357.

Sedikides, C. (1993). Assessment, enhancement, and verification determinants of the self-evaluation process. *Journal of Personality and Social Psychology, 65,* 317–338.

Sedikides, C., Campbell, W. K., Reeder, G. D., & Elliott, A. J. (1998). The self-serving bias in relational context. *Journal of Personality and Social Psychology, 74,* 378–386.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, to thine own self be better. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 209–269). New York: Academic Press.

Shapiro, B., Eppler, M., Haith, M., & Reis, H. (1987). *An event analysis of facial attractiveness and expressiveness.* Paper presented at the meeting of the Society of Research in Child Development, Baltimore.

Simon, L., Greenberg, J., Harmon-Jones, E., Solomon, S., Pyszczynski, T., Arndt, J., & Abend, T. (1997). Terror management and cognitive-experiential self-theory: Evidence that terror management occurs in the experiential system. *Journal of Personality and Social Psychology, 72,* 1132–1146.

Smith, E. R., & Lerner, M. (1986). Development of automatism of social judgments. *Journal of Personality and Social Psychology, 37,* 2240–2252.

Solomon, S., Greenberg, J., & Pyszczynski, T. (1991). A terror management theory of social behavior: The psychological functions of self-esteem and cultural worldviews. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 24, pp. 93–159). San Diego: Academic.

Stapel, D., Koomen, W., & Zeelenberg, M. (1998). The impact of accuracy motivation on interpretation, comparison, and correction processes: Accuracy × Knowledge accessibility effects. *Journal of Personality and Social Psychology, 74,* 878–893.

Stipek, D. J., & Daniels, D. H. (1988). Declining perceptions of competence: A consequence of changes in the child or in the environment? *Journal of Educational Psychology, 80,* 352–356.

Strack, F., & Hannover, B. (1996). Awareness of influence as a precondition for implementing correctional goals. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 579–596). New York: Guilford Press.

Swann, W. B., Jr., Hixon, G., Stein-Seroussi, A., & Gilbert, D. T. (1990). The fleeting gleam of praise: Cognitive processes underlying behavioral reactions to self-relevant feedback. *Journal of Personality and Social Psychology, 59,* 17–26.

Swann, W. B., Jr., & Schroeder, D. G. (1995). The search for beauty and

truth: A framework for understanding reactions to evaluations. *Personality and Social Psychology Bulletin, 21,* 1307–1318.

Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization–minimization hypothesis. *Psychological Bulletin, 110,* 67–85.

Tesser, A., Martin, L. L., & Cornell, D. P. (1996). On the substitutibility of self-protective mechanisms. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 48–68). New York: Guilford Press.

Tice, D. M., Butler, J. L., Muraven, M. B., & Stillwell, A. M. (1995). When modesty prevails: Differential favorability of self-presentation to friends and strangers. *Journal of Personality and Social Psychology, 69,* 1120–1138.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98,* 219–235.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 29, pp. 141–208). London: Academic.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influence on judgments and evaluations. *Psychological Bulletin, 116,* 117–142.

Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B., & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude–behavior consistency. *Journal of Personality and Social Psychology, 47,* 5–16.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 22, pp. 287–343). Orlando, FL: Academic Press.

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

Woike, B. A. (1995). Most-memorable experiences: Evidence for a link between implicit and explicit motives and social–cognitive processes in everyday life. *Journal of Personality and Social Psychology, 68,* 1081–1091.

Received May 12, 2000
Revision received October 10, 2000
Accepted October 10, 2000 ■