Journal of Experimental Psychology: General
2004, Vol. 133, No. 2, 139–165

0096-3445/04/$12.00 DOI: 10.1037/0096-3445.133.2.139

Underlying Processes in the Implicit Association Test: Dissociating Salience From Associations

Klaus Rothermund
University of Trier

Dirk Wentura
University of Jena

The authors investigated whether effects of the Implicit Association Test (IAT) are influenced by salience asymmetries, independent of associations. Two series of experiments analyzed unique effects of salience by using nonassociated, neutral categories that differed in salience. In a 3rd series, salience asymmetries were manipulated experimentally while holding associations between categories constant. In a 4th series, valent associations of the target categories were manipulated experimentally while holding salience asymmetries constant. Throughout, IAT effects were found to depend on salience asymmetries. Additionally, salience asymmetries between categories were assessed directly with a visual search task to provide an independent criterion of salience asymmetries. Salience asymmetries corresponded to IAT effects and also accounted for common variance in IAT effects and explicit measures of attitudes or the self-concept.

Cognitive psychology lives on its basic experimental paradigms—be it the Stroop task (e.g., MacLeod, 1991), the Simon task (e.g., Simon, 1990), the semantic priming paradigm (e.g., Neely, 1991), or the negative priming paradigm (e.g., Tipper, 2001), to name just a few. These paradigms are associated with replicable, robust, and nontrivial effects. Researchers share the strong belief that these paradigms are suited to reveal essential truths about the functional architecture of the human cognitive system. Thus, in some sense, they are considered as windows on the mind.

Recently, the family of these paradigms has gotten a newborn child: the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT is an experimental paradigm that reveals strong and robust response interferences that were not known before. In essence, the IAT is a double discrimination task that maps four categories onto two responses. The four stimulus categories represent two classes of stimuli—the target class and the attribute class—consisting of two categories each. For example, flowers and insects are target categories, pleasant and unpleasant words are attribute categories (Greenwald et al., 1998, Experiment 1). The two categories of each class are assigned to two responses, yielding a complex task that is made up of four different category–response assignments (e.g., for flowers and pleasant words, press the right key; for insects and unpleasant words, press the left key). In successive blocks of the IAT, the two binary classification tasks for the targets and attributes are combined in two different ways: In Block A, flowers and pleasant words are assigned one response, whereas insects and unpleasant words are assigned the other response. In Block B, however, response assignments are switched for one class of stimuli (e.g., for the flower–insect dichotomy), so that flowers and unpleasant words (as well as insects and pleasant words) are now assigned the same response. An IAT effect is calculated as the difference between the average response times in the two blocks (i.e., B − A). What is found is a difference that is clearly positive for most participants. In other words, if those categories that intuitively go together are assigned the same response, the task is much easier than in the other condition.

Since the seminal article of Greenwald et al. (1998), a wide range of different materials have been used for the IAT. Basically, results can be categorized into two sets. First, whenever the two blocks can be classified as compatible versus incompatible on the basis of a priori reasoning about what generally seems to go together, the response-time difference incompatible minus compatible is positive. Second, whenever interindividual differences can be assumed with regard to which categories seem to go together for a given participant, the IAT difference reflects those differences. For example, Greenwald et al. (Experiment 2) found interindividual differences in an IAT with Japanese versus Korean names as the target dimension. If assignment of the same response to Japanese names and pleasant words is (arbitrarily) considered as the compatible task, Japanese participants had a positive IAT effect, whereas Korean participants had a negative one. Given these findings, it is not surprising that this paradigm has attracted much attention beyond cognitive psychology as a potential assessment tool for attitudes and other kinds of implicit associations in social psychology (e.g., Greenwald et al., 1998, 2002), personality psychology (e.g., Asendorpf, Banse, & Mücke, 2002), abnormal psychology (e.g., Teachman, Gregg, & Woody, 2001), developmental psychology (e.g., Hummert, Garstka, O'Brien, Greenwald, & Mellott, 2002), and market research (e.g., Maison, Greenwald, & Bruin, 2001).

Despite an ever-growing number of studies that reveal IAT effects with high face validity (taking into account what intuitively

Klaus Rothermund, Department of Psychology, University of Trier, Trier, Germany; Dirk Wentura, Department of Psychology, University of Jena, Jena, Germany.

Correspondence concerning this article should be addressed to Klaus Rothermund, Department of Psychology, University of Trier, D-54286 Trier, Germany. E-mail: rothermu@uni-trier.de

goes together) for either the average participant (e.g., the flower–insects example, above) or the average member of a specified group (e.g., the Japanese–Korean example, above), the question of which processes contribute to these effects is still unsettled (see, e.g., Greenwald & Nosek, 2001). As suggested by the name of the task, an account in terms of associations is most prominent. According to Greenwald et al. (1998), IAT effects reflect a cognitive association between the concepts that make up the target and attribute categories (e.g., an association between the concepts flower and pleasant and/or an association between the concepts insect and unpleasant). Although the assumption that the IAT measures associations between target and attribute categories seems straightforward, it is not an easy task to specify the basic processes that mediate between associations and IAT effects (see, e.g., Brendl, Markman, & Messner, 2001; De Houwer, 2001; Mierke & Klauer, 2001). Different accounts have been proposed, but up to now, "research on alternative theoretical interpretations has not yet progressed enough to establish any theoretical interpretation of the IAT effect" (Greenwald & Nosek, 2001, p. 90).

Besides, there are some results that are difficult to reconcile with an association account of IAT effects. For example, in a modified version of the flower–insect IAT, nonwords were used instead of flowers as one of the target categories (Brendl et al., 2001). In this arrangement, a reversed IAT effect emerged: Responses were faster when insects and pleasant words were assigned one response and (affectively neutral) nonwords and unpleasant words were assigned the other response. This result is incompatible with the assumption that the flower–insect IAT result was (at least partly) caused by an insect-negative association because this association should also have dominated the insect–nonword IAT. One might argue that the flower–insect IAT is exclusively based on a flower-positive association. Then, however, an insect–nonword IAT should yield no effect for the average participant, rather than a reversed effect, because in this case, no clear associations can be assumed to dominate the IAT. To explain this finding, it might thus be fruitful to consider alternative accounts of IAT effects. In this article, we propose an account that is based on the assumption that IAT effects can also reflect figure–ground asymmetries between categories. According to this account, associations between categories are not necessary to get IAT effects.

The Figure–Ground Model of IAT Effects

The main thesis of the figure–ground model is that differences in salience between the categories of the task lead to IAT effects (Rothermund & Wentura, 2001). Assume for a moment that the two categories of both the target and the attribute dimension indeed differ in salience (e.g., assume that insects are more salient than flowers and that unpleasant words are more salient than pleasant words). In this case, participants would find it easier to respond if the salient categories of both dimensions (the figures) were mapped onto one response and the nonsalient categories (the background) were mapped onto the other response. With such a consistent assignment of salient and nonsalient categories to responses, category salience helps to discriminate between responses. The categorization task now resembles a simple visual search task with a display set size of one stimulus per trial (see, e.g., Wolfe, 1998): If a stimulus belongs to a salient category, a "*yes*" response is executed (i.e., the response assigned to the figure categories). If the stimulus does not belong to a salient category, a "*no*" response is executed (i.e., the response assigned to the background categories). In incompatible blocks, by contrast, the salience and response dimensions are orthogonal, that is, salient and nonsalient categories are not mapped consistently onto responses. In this case, there is no facilitative influence of salience on responding. The figure–ground model thus assumes that IAT effects reflect independent salience asymmetries within the target and attribute dimensions.

Starting from the assumption that the categories within the target and attribute dimensions differ in salience, we have argued that responses are faster for a compatible assignment of salient and nonsalient categories to responses. In the following, we argue that figure–ground asymmetries are actually a common feature of categorization tasks in the IAT and can be used to explain a wide range of existing IAT effects. Figure–ground asymmetries between categories can have many origins (e.g., linguistic properties of the category labels such as marked vs. unmarked language codes—see Greenberg, 1966; perceptual or gestalt qualities relating to differences in brightness, color, or form; etc.). Most important for an explanation of typical IAT effects, however, is the fact that differences in the salience of categories are also closely related to valence and familiarity. Attentional asymmetries between negative and positive stimuli have been repeatedly demonstrated (e.g., Fox et al., 2000; Öhman, Flykt, & Esteves, 2001; Pratto & John, 1991; Wentura & Rothermund, 2003). Negative categories automatically attract or hold attention, presumably because negative information is typically more relevant for the regulation of behavior than positive information (Peeters, 1983). A negative category thus constitutes the figure of the respective dimension, whereas positive categories form the background (Kanouse & Hanson, 1972). Attentional asymmetries have also been found between familiar and unfamiliar stimuli. Wang, Cavanagh, and Green (1994) found that unfamiliar targets pop out among familiar distractors but not vice versa (see also Malinowski & Hübner, 2001; Shen & Reingold, 2001; Wolfe, 2001). In a similar vein, Johnston and Hawley (1994) reported attentional effects of a "novel popout" (p. 56) and a "familiar sink-in" (p. 56) for new and recently presented stimuli, respectively. Lubow and Kaplan (1997) replicated this finding and offered an explanation in terms of conditioned inattention to frequently encountered stimuli. The less familiar of two categories is thus typically more salient and constitutes a figure against the background of the familiar category.[1]

The previous arguments indicate that valence, familiarity, and salience are typically confounded. Explaining standard IAT effects in terms of figure–ground asymmetries thus does not rule out an association account of the IAT that is based on valence (e.g., De

[1] Salience asymmetries between categories can be driven by differences in attentional engagement as well as by differences in attentional disengagement. The exact nature of the attentional processes underlying salience asymmetries based on familiarity or valence is still an object of empirical debate (e.g., Fox, Russo, Bowles, & Dutton, 2001; Tipples, Young, Quinlan, Broks, & Ellis, 2002). We assume that salience asymmetries in the IAT manifest themselves by processes regarding attentional disengagement (*attentional dwell time*) because all stimuli appear in the same location, which makes it unnecessary to shift attention between different spatial locations.

Houwer, 2001). The figure–ground model, however, can also explain recent findings with nonstandard versions of the IAT that are difficult to accommodate on the basis of an association account. The result found by Brendl et al. (2001; see above) with the insect–nonword IAT seems to suggest that insects are associated with a positive valence, which is contrary to what Greenwald et al. (1998) found in their original flower–insect IAT. The figure–ground model offers a simple explanation for the seemingly contradictory findings. Although the category of nonwords has no valence, nonwords should nevertheless tend to pop out compared with the category of insects because nonwords are highly unfamiliar to participants (Johnston & Hawley, 1994; Lubow & Kaplan, 1997; Wang et al., 1994). The figure–ground asymmetry in the target dimension is thus reversed, and nonwords and negative words form the figures against a background of positive words and insects. Therefore, faster responses should emerge for a condition that assigns insects and positive words the same response.

Separating Effects of Salience and Associations

In spite of their structural dissimilarity, the association model and the figure–ground model often yield identical predictions because of the confounding of the valence and salience dimensions. Post hoc explanations of previous experimental findings also seem somewhat limited in scope because of other potentially confounding factors. To demonstrate that figure–ground asymmetries have unique effects on the IAT that are not mediated by associations, it is necessary to experimentally separate effects of salience and valence. Four series of such *dissociation experiments* are reported here.

The first series of experiments eliminated associations between the target and attribute categories but retained the salience asymmetries. In this case, the association account would not predict IAT effects because associations between the target and attribute categories no longer exist. The figure–ground model, by contrast, would predict faster responses for a consistent mapping of salient–nonsalient categories onto responses. The first series of experiments also introduced a visual search task to measure salience asymmetries independent of IAT effects. We investigated whether these salience asymmetries corresponded to the observed IAT effects.

A second series of experiments was concerned with group differences in IAT effects. We investigated whether interindividual differences in salience asymmetries could account for interindividual differences in IAT effects and also for the relation between IAT effects and self-report measures of self-concept variables or subjective attitudes.

In the third series of experiments, salience asymmetries within the target and attribute dimensions were manipulated while holding the stimuli—and evaluative associations between them—constant. An association account would not predict an influence of salience on IAT effects that is independent of a change in associations. The figure–ground model, however, predicts a shift of IAT effects corresponding to the salience manipulations.

Finally, a fourth series of experiments manipulated the valence of the target categories while holding figure–ground asymmetries constant. An association account would predict that IAT effects should vary according to the valence of the stimuli of the target categories. The figure–ground model, by contrast, predicts similar IAT effects as long as salience asymmetries remain unchanged.

Experiment 1A

The first series of experiments aimed at a positive demonstration of IAT effects based on figure–ground asymmetries that could not be explained on the basis of associations. Because most of these experiments involved modified versions of an old–young IAT, we first present an experiment that documents the standard version of the old–young IAT (see also Nosek, Banaji, & Greenwald, 2002; Rothermund & Wentura, 2001; Rudman, Greenwald, Mellott, & Schwartz, 1999).

Method

Participants. Sixteen University of Trier psychology undergraduates (15 women, 1 man) volunteered for partial course credit.

Materials. Ten stimuli were selected for each category (young and old names, pleasant and unpleasant words). Explicit evaluations for the stimuli of Experiments 1A and 1B were gathered in a pilot study. All stimuli were presented in a random order on a computer screen, and evaluative ratings were given on a 7-point scale (-3 = *negative*, 3 = *positive*). Mean evaluative ratings were clearly positive for the pleasant words ($M = 2.04$) and clearly negative for the unpleasant words ($M = -2.05$). Evaluative ratings were significantly positive (larger than zero) for the young names, $M = 0.43$, $t(9) = 3.86$, $p < .01$, $d = 1.22$, but did not differ significantly from zero for the old names, $M = 0.02$, $t(9) < 1$. All stimuli used in the experiment are listed in Appendix A.

Design. We manipulated compatible versus incompatible response assignments within participants. Response assignments were classified as compatible when the categories old and bad were assigned the same response key. Half of the participants received a compatible response assignment first. All possible initial assignments of categories to responses for the target and attribute categories were realized equally often. In the second part of the experiment, response assignments were inverted for the target categories for half of the participants, whereas response assignments were inverted for the attribute categories for the other half of the participants.

Procedure. Participants first received two practice blocks, one with the name stimuli and the old–young categorization task and one with the remaining stimuli and the good–bad categorization task. During the practice blocks, each stimulus appeared once. In a third block, stimuli of the target and attribute categories had to be categorized simultaneously according to the response schedule of the practice blocks. During this block, each stimulus was presented twice, yielding a total of 80 trials that were presented in an individually randomized sequence. The first 20 trials of this sequence were presented as practice trials. In a fourth block, a practice block for one of the simple categorization tasks was presented again, inverting the response assignments of either the target or the attribute categories. In a fifth block, participants again received 80 trials (of which the first 20 trials served as practice trials) of a combined classification task in which responding was inverted for one dimension, as in the fourth block.

Presentation of stimuli and registration of responses were controlled by a TurboPascal program operating in graphics mode. All stimuli were presented in white uppercase letters in the middle of a black computer screen. Category labels (old, young, good, bad) were constantly shown at the top right and top left corners of the display, indicating the assignment of categories to responses. Two keys on the computer keyboard were marked as response keys ($D \rightarrow$ left, $L \rightarrow$ right). In each trial, the stimulus remained on the screen until a response was registered. In case of an incorrect response, the stimulus remained on the screen, and an error message was displayed in red beneath the stimulus ("ERROR—press correct key and continue"). The intertrial interval was 150 ms.

Results and Discussion

Erroneous responses (9.64%) and outlier values[2] (2.86%) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments (see Table 1). Response latencies were 153 ms shorter for the compatible assignments, that is, if the categories old and bad were assigned the same response, $t(15) = 7.06$, $p < .001$, $d = 1.77$. Converging with previous research (Rothermund & Wentura, 2001; Rudman et al., 1999), an IAT using old and young names as target stimuli and pleasant and unpleasant words as attribute stimuli yielded a strong compatibility effect.

Experiment 1B

To demonstrate an IAT effect based on figure–ground asymmetries that could not be explained on the basis of associations, we conducted a modified version of the previous old–young IAT that used neutral words and nonwords as attribute categories instead of pleasant and unpleasant words. This modification eliminated the valence component from the attribute dimension (nonwords are not associated with either a negative or a positive valence; this has recently been demonstrated experimentally with an alternative measure of implicit associations; De Houwer, 2002; Rothermund & Kaul, 2002). Accordingly, evaluative associations between the concepts old and nonword (or between young and word) could be ruled out. The association account thus no longer predicted an IAT effect. The dichotomy of words and nonwords nevertheless contained a figure–ground asymmetry (see, e.g., Flowers & Lohr, 1985; Wentura, 2000). Nonwords are highly unfamiliar and should pop out against a background of common neutral words. The figure–ground model therefore predicted faster responses when the unfamiliar figure categories (old and nonword) were assigned the same response.

Method

Participants. Sixteen University of Trier psychology undergraduates (13 women, 3 men) volunteered for partial course credit.

Table 1
Mean Reaction Times (in Milliseconds; Standard Deviations in Parentheses) in Compatible and Incompatible Blocks of Standard and Modified Versions of the Old–Young IAT (Experiments 1A, 1B, 1C, and 1E)

Labels of attribute categories	Response assignment[a]: Compatible	Response assignment[a]: Incompatible	IAT effect
Experiment 1A			
Good–bad	661 (57)	815 (107)	153 (87)
Experiment 1B			
Word–no word	712 (88)	803 (115)	91 (69)
Experiment 1C			
Siwob–no siwob	735 (84)	838 (128)	103 (108)
Experiment 1E			
Multicolored–single-colored	621 (76)	657 (91)	37 (76)

Note. IAT = Implicit Association Test; Siwob = wordlike letter sequence having no meaning.
[a] Categories old and no word or old and siwob were assigned the same response in compatible blocks.

Materials. Old and young names were identical to Experiment 1A. Ten stimuli were selected for each of the categories of words and nonwords. Valence ratings for the word stimuli did not differ significantly from zero, $M = 0.04$, $t(9) < 1$. Nonwords were created out of neutral words by changing letters; evaluative ratings for the nonword strings also did not deviate significantly from zero, $M = -0.12$, $t(9) = 1.57$, *ns*, nor did they deviate significantly from the evaluations of the word stimuli, $t(18) = 1.62$, *ns*. The stimuli are listed in Appendix A.

Design and procedure. Response assignments were classified as compatible when the categories old and nonword were assigned the same response key. In all other respects, the design and procedure were identical to those of Experiment 1A.

Results

Erroneous responses (7.45%) and outlier values (2.45%; see footnote 2) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments (see Table 1). Response latencies were 91 ms shorter for the compatible assignments, that is, if the categories old and nonword were assigned the same response, $t(15) = 5.29$, $p < .001$, $d = 1.32$.

Discussion

As predicted by the figure–ground model, a compatibility effect was observed for the modified version of the old–young IAT in which words and nonwords were used as attribute categories. A consistent mapping of salient and nonsalient categories onto responses led to faster responses compared with an inconsistent mapping, despite the fact that the attribute categories did not differ in valence. The finding replicates the results of a previous study that also used words and nonwords as attribute categories (Rothermund & Wentura, 2001). Experiment 1B thus supports the assumption that figure–ground asymmetries within the target and attribute categories do in fact produce compatibility effects in the IAT.

These findings are difficult to explain on the basis of associations between target and attribute categories because the attribute categories were specifically chosen to rule out explanations in terms of valence or associations: Words and nonwords had neutral valence and were clearly not associated with old and young names. It might be objected, however, that although nonwords themselves are neutral (De Houwer, 2002; Rothermund & Kaul, 2002), the category label *no word* (German: *Kein Wort*) might be perceived as more negative due to the prefixed *no.* This grammatical negation could be responsible for a perceived negativity of the category—and, in turn, for the observed IAT effect. To rule out this possibility, we conducted a conceptual replication of Experiment 1B in which the labels for the attribute categories were changed.

Experiment 1C

Method

Participants. Sixteen University of Trier psychology undergraduates (12 women, 4 men) volunteered for partial course credit.

[2] Values that were below 250 ms or that were more than three interquartile ranges above the median of the overall distribution of response latencies were treated as outliers (Tukey, 1977).

Materials, design, and procedure. Design, stimuli, and procedure were identical to those of Experiment 1B. The category of nonwords was now introduced as a distinct, positively defined category of *siwob,* which is a German acronym for *sinnlose, wortähnliche Buchstabenfolge* (wordlike letter sequence having no meaning). The words were referred to as *kein siwob* (no siwob), that is, with a label that contained a grammatical negation (see also Wentura, 2000).

Results

Erroneous responses (9.17%) and outlier values (3.13%; see footnote 2) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments (see Table 1). Response latencies were 103 ms shorter for the compatible assignments, that is, if the categories old and siwob were assigned the same response, $t(15) = 3.81$, $p < .001$, $d = 0.95$. Adding a negation prefix to one or the other of the attribute categories had no effect on the sign of the IAT effect, and the means did not differ significantly between experiments, $F < 1$. This finding rules out an alternative explanation of the observed effect in terms of a grammatical negativity of the category labels.

Discussion

Experiments 1B and 1C demonstrated compatibility effects in the IAT although associations between the target and attribute categories were eliminated. In these cases, associations between categories were nonexistent and thus can be ruled out as an explanation for the compatibility effects. We argue instead that these findings corroborate the figure–ground model. This interpretation rests on the assumption that one of the categories of the target and attribute dimensions is more salient than its counterpart. Although it is plausible to assume the existence of salience asymmetries for each of the dimensions of the previous experiments, we have not yet provided an independent measure of salience to test this assumption. This was done in Experiment 1D.

Visual search tasks provide a direct operational criterion of salience (Wolfe, 1998; Yantis, 1996): The salience concept implies that it should be easier to allocate attention to a stimulus of a salient category and more difficult to disengage attention from a stimulus of a salient category. This further implies that it should be easier to detect a salient stimulus among stimuli of the less salient category than the other way around; perhaps most important, it should take longer to process a display with distractors of the salient category compared with a display containing distractors of the less salient category (cf. Wolfe, 1998). Experiment 1D used visual search tasks to investigate the relative salience of two categories.

Experiment 1D

For each of the target and attribute dimensions of the previous IAT experiments, a different visual search task that contained only the stimuli of the two categories of the respective dimension was conducted. In each trial of a search task, four stimuli were presented simultaneously. Participants had to decide whether all stimuli belonged to the same category (e.g., all old names or all young names) or whether one of the stimuli belonged to a different category (e.g., one old name among three young names or one young name among three old names).[3] If stimuli of the salient category automatically draw and/or hold attention, it should take longer to process displays that contain a majority of these stimuli. We therefore predicted that response latencies would be longer if the majority of the stimuli (the distractors) belonged to the figure category. This prediction should have held for each of the dimensions of the previous experiments (old names vs. young names, nonwords vs. neutral words, unpleasant words vs. pleasant words).

Experiment 1D also included a search task for another dimension that had not been contained in the previous experiments. In one of the search tasks, multicolored vs. single-colored strings of zeros were presented as stimuli. Multicolored strings are more complex and do not form a simple perceptual gestalt. Processing of these stimuli should thus require more attentional resources, and multicolored strings should be perceptually more salient than single-colored strings. Accordingly, detecting a homogeneous single-colored string among multicolored distractor strings should take longer than detecting a multicolored string among single-colored distractor strings. This additional dimension was included in Experiment 1D for two reasons: First, we wanted to show that if a salience asymmetry could be demonstrated for two categories in the search task, then these categories should also produce an IAT effect when used as attribute categories in an IAT. Second, we wanted to use these purely perceptual categories to demonstrate a compatibility effect in an old–young IAT using attribute categories that were definitely neutral with respect to valence and that did not have any associations whatsoever with the categories old versus young. These predictions were tested in a later experiment of this series (Experiment 1E).

Method

Participants. Ninety-nine University of Trier psychology undergraduates (69 women, 30 men) volunteered for partial course credit. Each participant received either one or two out of the four search tasks specified below. A subsample of participants that completed only one search task of the present experiment also received one of two other search tasks that related to a later experiment (Experiment 4).

Materials. For the search tasks relating to the target and attribute dimensions of the previous IAT experiments, the same stimuli were used as before (old vs. young names, nonwords vs. neutral words, bad vs. good words). For the search task with colored strings, each stimulus consisted of a string of five zeros (*00000*). The zeros of the single-colored strings all had the same color, which was either blue, red, green, brown, or cyan. Different single-colored strings, however, also differed in color. For the multicolored strings, one of the five colors was selected at random for each of the five zeros of the string, with the restriction that at least two different colors had to be selected for the zeros of a string.

Design. Selection of search tasks was determined randomly for each participant. The category of the distractor stimuli (figure vs. ground) was manipulated across trials within each search task.

[3] This is a category version of visual search tasks that were previously used by Johnston and Hawley (1994) and Lubow and Kaplan (1997). We did not specify in advance which of the two categories would be the target category and which would be the distractor category in the upcoming trial. Specifying a search category before a trial might eliminate existing attentional asymmetries (see Experiment 3A). In a similar task with neutral words and nonwords as stimuli, Krueger, Stadtlander, and Blum (1992) reported longer latencies and more errors for lists containing nonwords as the distractor category.

Procedure. Before each search task, participants received 40 trials of a simple categorization task to practice the assignment of stimuli to their respective categories. The stimuli of one dimension had to be assigned to one of the two categories. In these practice blocks, each stimulus of the two categories of the respective dimension was presented twice. The corresponding search task was conducted immediately after the practice block. Each search task consisted of 12 practice trials and 64 experimental trials. The experimental trials were created according to the following rules: In each trial, four stimuli were presented. In half of the trials, the majority of the stimuli (three or four out of four) belonged to the figure category of the respective search task; in the other half of the trials, the majority of the stimuli belonged to the ground category. Furthermore, all stimuli belonged to the same category in half of the trials (*same* trials), whereas in the other half of the trials, three stimuli belonged to one category, and the fourth stimulus belonged to the other category (*different* trials). In half of the trials, the four stimuli were presented in the corners of a virtual square, whereas in the other half of the trials, the stimuli were presented in the corners of a virtual diamond. All of these factors were varied orthogonally, yielding a total of 8 trials for each possible combination. In the *different* trials within each type of trials, the target stimulus appeared twice at each of the four possible locations. Stimuli were selected randomly for each trial, with the restriction that each stimulus appeared only once within a display. Order of presentation of the experimental trials was randomized for each participant.

Presentation of stimuli and registration of responses were controlled by a TurboPascal program operating in text mode. Except for the colored strings, all stimuli were presented in white letters around the center of a black computer screen. Each trial consisted of the following sequence of events: A ready signal (#) was displayed in the middle of the screen. When participants pressed the space bar, the ready signal was replaced with a cue (*). After an interval of 750 ms, the four stimuli were presented around the cue, either in the shape of a square (two stimuli above the cue and two stimuli below the cue) or in the shape of a diamond (one stimulus above the cue, one stimulus each to the right and to the left of the cue, and one stimulus below the cue). The stimuli remained on the screen until one of the response keys was pressed. Two keys on the computer keyboard were marked as response keys ($D \rightarrow$ left, $L \rightarrow$ right). During the search task, response labels (same, different) were constantly shown at the top right and top left corners of the display. In case of an incorrect response, the stimuli remained on the screen and an error message was displayed beneath the stimulus ("ERROR—press correct key and continue"). The intertrial interval was 500 ms.

Results

Erroneous responses (old–young: 6.73%, nonword–word: 6.13%, bad–good: 6.18%, multicolored–single-colored: 9.52%) and outlier values (old–young: 0.84%, word–nonword: 0.66%, good–bad: 1.56%, multicolored–single-colored: 1.03%; see footnote 2; outlier criteria were determined separately for each type of search task) were excluded from the analyses. For each participant and search task, we calculated mean response latencies for trials in which the distractor stimuli belonged to the figure category (old, nonword, bad, multicolored) or to the ground category (young, word, good, single-colored; see Table 2). Separate *t* tests comparing the response latencies for trials with figure and ground distractors were conducted for each type of visual search task. In each case, response latencies were significantly longer for trials with figure distractors, old–young: $t(25) = 2.54$, $p < .01$, $d = .50$; nonword–word: $t(25) = 4.88$, $p < .001$, $d = .96$; bad–good: $t(21) = 3.60$, $p < .01$, $d = .77$; multicolored–single-colored: $t(31) = 10.13$, $p < .001$, $d = 1.79$.

Table 2
Mean Response Latencies (in Milliseconds; Standard Deviations in Parentheses) in Visual Search Tasks Depending on the Salience of the Distractor Stimuli (Experiments 1D and 4)

Categories of the search task (figure–ground)	Distractor type: Figure	Distractor type: Ground	Search asymmetry
Experiment 1D			
Old–young	2,249 (551)	2,147 (499)	102 (205)
Nonword–word	1,820 (301)	1,699 (290)	121 (126)
Bad–good	2,185 (371)	2,001 (334)	184 (240)
Multicolored–single-colored	1,145 (219)	923 (182)	222 (124)
Experiment 4			
Unknown–known-positive	2,447 (706)	2,258 (561)	189 (205)
Unknown–known-negative	2,785 (603)	2,570 (542)	214 (216)

Discussion

The results gathered with the different search tasks confirm the salience asymmetries that were presumed in the previous IAT experiments. It took longer to process displays containing a majority of old names, nonwords, or bad words. Supposedly, processing stimuli of these categories should require more attentional resources than processing stimuli of the ground categories, which is functionally equivalent to being more salient.

A clear-cut salience asymmetry also emerged for the colored strings. Multicolored strings were more easily identified in a display of single-colored strings than the other way around. Multicolored strings thus mark the more salient category. Hence, the salience account would predict that a compatibility effect should also emerge in an IAT in which multicolored and single-colored strings were combined with old and young names. Experiment 1E was conducted to test this prediction.

Experiment 1E

Method

Participants. Thirty-two University of Trier psychology undergraduates (25 women, 7 men) volunteered for partial course credit.

Materials. Old and young names were identical to the stimuli used in the previous experiments. Multicolored and single-colored strings were created as described in Experiment 1D.

Design and procedure. Response assignments were classified as compatible when the categories old and multicolored were assigned the same response key. In all other respects, the design and procedure were identical to those of Experiment 1A.

Results

Erroneous responses (8.93%) and outlier values (3.57%; see footnote 2) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments (see Table 1). Response latencies were 37 ms shorter for the compatible assignments, that is, if the categories old and multicolored were assigned the same response, $t(31) = 2.73$, $p < .01$, $d = 0.48$.

Discussion

Converging with the predictions of the salience account, a compatibility effect emerged in a modified version of an old–

young IAT even when purely perceptual categories (multicolored–single-colored) that differed in salience were introduced as an attribute dimension. As in the previous experiments, the nature of the attribute categories ruled out alternative explanations of this effect in terms of valence or other associations between the colored strings and old and young names. In this regard, Experiment 1E is perhaps even more convincing than Experiments 1B and 1C because this time, all stimuli of both attribute categories were completely devoid of meaning.

One might wonder why the size of the compatibility effect in Experiment 1E was somewhat smaller than in the previous experiments although the single- and multicolored strings yielded a large figure–ground asymmetry in the visual search task. The comparatively small IAT effect with the colored strings may have been due to the low difficulty of the combined classification task: Single- versus multicolored strings were very easy to discriminate (as is evident from the average response times in the search task; see Table 2). Furthermore, target and attribute stimuli were perceptually more dissimilar than in the previous IAT experiments. It was thus much easier to retrieve the correct response rule in each trial of the IAT, which allowed a fast responding (mean response times were approximately 100 ms shorter than in the other IAT experiments) and reduced the potential for a facilitating effect of salience on responding in the compatible block. Such an attenuating influence of category discriminability on IAT effects has been demonstrated empirically by Mierke and Klauer (2001): Increasing the discriminability of the target and attribute categories by advance cues strongly reduced the size of IAT effects. Mierke and Klauer argued that this effect of category discriminability is mediated by reduced task switching costs. Task shift costs in the incompatible block are reduced if target and attribute categories can be easily distinguished by an obvious perceptual cue; facilitating operations in the incompatible block in turn reduces the potential benefit that can be gained from a compatible assignment of figure categories to responses, resulting in a kind of floor effect (Mierke & Klauer, 2001; Rothermund & Wentura, 2001).

The previous experiments demonstrated the influence of figure–ground asymmetries in the old–young IAT. Effects of figure–ground asymmetries are not restricted to this particular variant of the IAT. Similar IAT effects without associations were also found in a recent set of studies by Kinoshita and Peek-O'Leary (2002): Significant compatibility effects emerged in a modified flower–insect IAT using neutral words versus nonwords as attribute categories and in IATs using odd versus even numbers as target categories. All of these studies exemplify general IAT effects that emerge in the same direction for all participants. In a large number of studies, however, the IAT has also been used to investigate interindividual differences (e.g., shy–nonshy IAT: Asendorpf et al., 2002; Japanese–Korean IAT: Greenwald et al., 1998; male–female IAT: Greenwald & Farnham, 2000). On the basis of an association account of the IAT, interindividual or group differences in the magnitude or direction of IAT effects have been taken to reflect differences in implicit associations underlying attitudes or the self-concept. A similar logic applies to the interpretation of correlations of IAT effects with explicit measures of attitudes, personally held stereotypes, or the self-concept: Positive correlations between IAT effects and respective self-report measures might be taken to support the validity of the IAT as a measure of implicit associations (e.g., zero-order correlations with explicit measures have been used as a criterion for evaluating the quality of different IAT methodologies; Greenwald, Nosek, & Banaji, 2003; Nosek, Greenwald, & Banaji, 2003).

According to the figure–ground account, interindividual differences in IAT effects might also reflect differences in salience asymmetries. The following series of experiments attempted to investigate whether differences in figure–ground asymmetries could account for interindividual differences in IAT effects and for associations between IAT effects and explicit measures.

Experiment 2A

Greenwald and Farnham (2000) used an IAT with the target categories self versus other and the attribute categories male versus female. For women, responses were faster when the categories self and female were assigned the same response, whereas for men, responses were faster with the opposite assignment (Greenwald & Farnham, 2000; see also Rudman, Greenwald, & McGhee, 2001, Experiment 3). On the basis of the association account, this group difference in the sign of the IAT effect would reflect differences in the gender self-concept of women and men: It is assumed that the concept self is linked to the concept female for women (and/or other is linked to male), whereas for men, self is associated with male (and/or other is related to female).

An explanation of this gender difference in terms of the figure–ground model can be given if it is assumed that for men, the self-descriptive category masculine is a background category, against which the self-aschematic category feminine pops out, whereas the reverse is true for women. The category other should constitute a figure against the background category self for all participants, regardless of gender.[4] In this case, the response assignment other–feminine versus self–masculine is compatible for men because it maps salient and nonsalient categories in a consistent manner onto responses, whereas for women, other–masculine versus self–feminine would be the compatible response assignment.

To differentiate between these explanations of gender differences in the male–female IAT in terms of self-concept associations and figure–ground asymmetries, we conducted an additional modified version of the male–female IAT in which the categories self and other were replaced with the neutral categories word and nonword. As demonstrated in Experiment 1D, the unfamiliar category nonword constitutes a figure against the familiar background category word. As with the categories self and other, this asymmetry should hold regardless of gender. The figure–ground model thus would predict a pattern of group differences in IAT effects for the modified male–female IAT similar to that of the original

[4] Prima facie, the assumption that self-related information is more accessible and that processing this information requires fewer attentional resources than a processing of other-related information may appear implausible: As illustrated by the famous cocktail-party phenomenon, self-related information (i.e., the participant's own name) automatically captures attention when it is presented to the unattended channel in a dichotic listening task (Moray, 1959). In a study by Bargh (1982), however, it was shown that the reverse is true—that is, fewer attentional resources are required to process self-related information—when this information is presented to the attended channel. This situation closely resembles the IAT situation in which all stimuli are presented in focal awareness.

variant: For women, faster responses should emerge if the figure categories nonword and male are assigned the same response, whereas for men, responses should be faster for the opposite assignment. The association account, by contrast, does not predict IAT effects for this modified variant of the task at all, nor does it predict group differences in compatibility effects for the modified task between women and men because the categories word and nonword are not associated with the categories male or female.

We also assessed search asymmetries between the categories male–female, self–other, and word–nonword directly with visual search tasks that were analogous to those in Experiment 1D. These search tasks provided an independent test of the proposed salience asymmetries between categories. We predicted that male would be a figure category for women, whereas female would be a figure category for men. The categories other and nonword would constitute figure categories against the background categories self and word regardless of gender. The Bem Sex Role Inventory (BSRI; Bem, 1974) was used as an explicit measure of participants' gender self-concept. Greenwald and Farnham (2000) found a positive correlation between BSRI scores and IAT effects. We predicted that an association between the explicit gender self-concept and IAT effects would be reduced or eliminated when differences in salience asymmetries were statistically controlled.

Method

Participants. Sixty-four University of Trier psychology undergraduates (32 women, 32 men) volunteered for partial course credit.

Materials. Words and nonwords were identical to those in the previous experiments. Eight masculine adjectives (e.g., *brave, self-confident*) and eight feminine adjectives (e.g., *tender, gentle*) were selected for the categories male and female, respectively. Four generic pronouns were selected for each of the categories self (e.g., *I, me*) and other (e.g., *they, them*). All stimuli are listed in Appendix A.

Design and procedure. Each participant conducted the original and the modified versions of the male–female IAT. Response assignments in the IATs were (arbitrarily) classified as compatible when the categories male and self (original version) or male and word (modified version) were assigned the same response key. The order of presentation of the original IAT and the modified IAT was counterbalanced across participants. Procedural details of the IATs were identical to those of Experiment 1A. Following the IATs, each participant conducted three visual search tasks for the categories male–female (first), self–other (second), and word–nonword (third). The sequence of the search tasks was identical for all participants. Procedural details of the search tasks were identical to those of Experiment 1D. Finally, participants completed a German translation of the BSRI (Schneider-Düker & Kohler, 1988). Twenty-six items that differentiated between women and men assessed either femininity (13 items) or masculinity (13 items).

Results

IAT analyses. Erroneous responses (original version: 8.02%; modified version: 8.36%) and outlier values (original version: 3.09%; modified version: 2.98%; see footnote 2, above) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments in the original and modified IATs (see Table 3; assignments of the categories male and self or male and word to the same response key were coded as compatible). A 2 (compatibility) × 2 (IAT type: original vs. modified variant) × 2 (gender of participants) analysis of variance (ANOVA) revealed a main effect of IAT type, $F(1, 62) = 25.83$, $MSE = 1{,}338.25$, $p < .001$, $f = .65$, indicating that on average, response times were 23 ms faster in the original than in the modified version of the IAT. Furthermore, the Compatibility × Gender interaction was also significant, $F(1, 62) = 55.14$, $MSE = 1{,}723.79$, $p < .001$, $f = .94$, indicating that for men, response times were 29 ms shorter when the category male was assigned the same response as the categories self or word, whereas for women, responding was facilitated by 48 ms when the categories female and self or female and word were assigned the same response. It is important to note that the Compatibility × Gender interaction was not further qualified by the type of IAT, $F(1, 62) = 1.06$, *ns*, that is, it emerged in the same fashion for the original IAT using self versus other as target categories, $F(1, 62) = 34.95$, $MSE = 1{,}768.81$, $p < .001$, $f = .75$, as for the modified IAT in which word versus nonword were used as target categories, $F(1, 62) = 20.42$, $MSE = 1{,}719.25$, $p < .001$, $f = .57$. Compatibility effects in the original and modified variants of the male–female IAT were significantly intercorrelated, $r = .29$, $p < .05$.

Search task analyses. Erroneous responses (male–female: 10.47%; self–other: 9.40%; word–nonword: 8.45%) and outlier values (male–female: 1.34%; self–other: 0.93%; word–nonword: 1.00%; see footnote 2; outlier criteria were determined separately for each type of search task) were excluded from the analyses. For each participant and search task, we calculated mean response latencies for trials in which the distractor stimuli belonged to the figure category (male, other, nonword) or to the ground category (female, self, word; see Table 4). Separate 2 (distractor type) × 2 (gender of participants) ANOVAs were conducted on the mean response latencies for each type of visual search task. In the male–female search task, a main effect of distractor type was found, $F(1, 62) = 24.61$, $MSE = 15{,}367.55$, $p < .001$, $f = .63$, that was qualified by an interaction with gender, $F(1, 62) = 26.56$, $MSE = 15{,}367.55$, $p < .001$, $f = .65$. Analyses of simple main effects revealed that for women, response latencies were 222 ms longer for trials with male distractors, $t(31) = -7.77$, $p < .001$, $d = 1.37$, whereas for men, response latencies were 4 ms longer for trials with female distractors, $t < 1$. In the other–self search task, only the main effect of distractor type was significant, $F(1,$

Table 3
Mean Reaction Times (in Milliseconds; Standard Deviations in Parentheses) in Compatible and Incompatible Blocks of Standard and Modified Versions of the Male–Female IAT for Men and Women (Experiment 2A)

	Response assignment[a]		
Labels of target categories	Compatible	Incompatible	IAT effect
Self–other			
Men	658 (75)	672 (88)	14 (62)
Women	696 (88)	623 (74)	−74 (57)
Word–no word			
Men	672 (79)	716 (83)	43 (56)
Women	689 (97)	666 (83)	−23 (61)

Note. IAT = Implicit Association Test.
[a] Categories male and self or male and word were assigned the same response in compatible blocks.

Table 4
Mean Response Latencies (in Milliseconds; Standard Deviations in Parentheses) in Visual Search Tasks Depending on the Salience of the Distractor Stimuli (Experiment 2A)

Categories of the search task (figure–ground)	Distractor type		Search asymmetry
	Figure	Ground	
Female–male[a]			
Men	2,001 (376)	1,996 (370)	4 (188)
Women	1,867 (356)	2,089 (429)	−222 (161)
Other–self			
Men	1,796 (308)	1,565 (282)	231 (173)
Women	1,709 (281)	1,543 (234)	166 (167)
Nonword–word			
Men	1,562 (242)	1,420 (262)	142 (123)
Women	1,507 (257)	1,429 (275)	78 (110)

[a] The category female was arbitrarily assigned the status as figure category for both male and female participants.

62) = 87.42, MSE = 14,431.80, $p < .001$, f = 1.19, indicating that regardless of gender, response latencies were 199 ms longer for trials with distractors of the figure category other. In the nonword–word search task, again, a strong main effect of distractor type emerged, $F(1, 62) = 56.92$, MSE = 6,793.29, $p < .001$, f = .96, indicating that response latencies were 110 ms longer for trials with distractors of the figure category nonword. For the latter task, the Distractor Type × Gender interaction was also significant, $F(1, 62) = 4.83$, MSE = 6,793.29, $p < .05$, f = .28. This interaction, however, was ordinal because longer response latencies for trials with nonword distractors were observed for men, M = 142 ms, $t(31) = 6.52$, $p < .001$, d = 1.15, and women, M = 78 ms, $t(31) =$ 4.02, $p < .001$, d = .71, alike.

Self-reported masculinity–femininity. For each participant, we subtracted the average self-rating for the femininity items from the average self-rating for the masculinity items to yield an index of relative masculinity–femininity. Women and men differed significantly on the resulting masculinity–femininity score, $t(62) = 4.38$, $p < .001$, d = .55. On average, women rated themselves higher on the femininity items than on the masculine attributes, resulting in a negative difference score, M = −1.05, SD = .90, whereas for men, average ratings for the masculine and feminine attributes did not differ, M = −.06, SD = .92.

Interrelations between IAT effects, salience asymmetries, and self-reported masculinity–femininity. Additional analyses were conducted to estimate interrelations between measures. All variables were represented by multiple indicators in structural equation models, which allowed an estimation of path coefficients that corrected for the unreliability of the measures (e.g., Baron & Kenny, 1986). To obtain parallel indicators for the RT data, we computed effect scores (IAT effects, salience asymmetries in the search task) for the odd- and even-numbered trials of the respective tasks for each participant. The femininity items and the masculinity items of the BSRI were also divided into two sets of items, respectively, to obtain two parallel indicators of masculinity–femininity. All indicators were standardized before they were entered into the structural equation models to obtain latent variables with comparable variances.

A first structural equation model revealed that BSRI masculinity–femininity predicted IAT effects in the original male–female IAT and also in the modified version of the task (see Figure 1A). In a second step, search asymmetries for the categories female and male were entered as an additional predictor into the equation (see Figure 1B). Three findings are noteworthy in this second set of equations. First, salience asymmetries for the categories female and male correlated positively with the BSRI masculinity–femininity score, indicating that self-reported masculinity–femininity was associated with an increased salience of the nonschematic female–male attributes in the search task, respectively. Second, salience asymmetries significantly predicted IAT effects for both versions of the IAT task over and above the effects of the BSRI masculinity–femininity score. Third, introducing the salience asymmetries into the equations rendered effects of BSRI masculinity–femininity on the IAT effects nonsignificant.

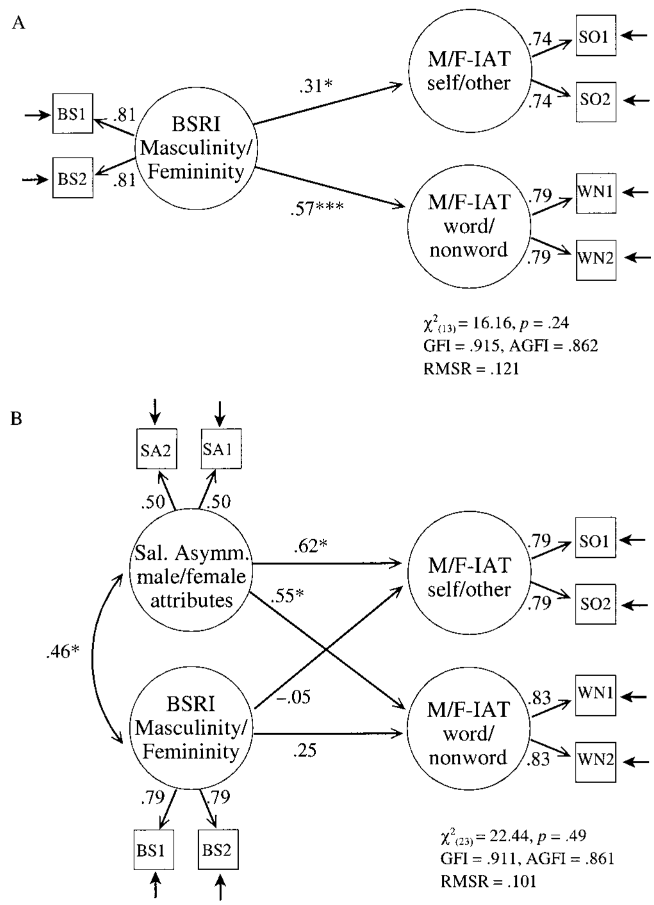

Figure 1. Structural equation models estimating the interrelations between self-reported masculinity–femininity, salience asymmetries regarding male and female attributes, and Implicit Association Test (IAT) effects in different versions of a male–female IAT (Experiment 2A). A: A first structural equation model revealing that Bem Sex Role Inventory masculinity–femininity predicted IAT effects in the original male–female IAT and in the modified version of the task. B: A second step in which search asymmetries for the categories female and male were entered as an additional predictor into the equation. AGFI = adjusted goodness of fit index; BS = Bem Sex Role Inventory; BSRI = Bem Sex Role Inventory; GFI = goodness of fit index; M/F = male/female; RMSR = root-mean-square residual; SA = salience asymmetries; Sal. Asymm. = salience asymmetries; SO = self/other; WN = word/nonword. *$p < .05$. ***$p < .001$.

Discussion

As predicted by the figure–ground model, an identical pattern of gender-dependent compatibility effects emerged in the original and in the modified versions of the male–female IAT. The IAT effects for women and men closely replicated the pattern of findings reported by Greenwald and Farnham (2000). Gender differences in the original version of this task are compatible with explanations drawing on differences in the gender self-concepts of men and women (association account) as well as with differences in salience asymmetries between schematic and aschematic categories (figure–ground account). In the modified version of the IAT, however, associations between the target and attribute categories were eliminated. The pattern of findings that emerged for the word–nonword variant of the male–female IAT exactly matches the predictions of the figure–ground model.

An account of the IAT effects in terms of salience asymmetries is further corroborated by the results found with the visual search tasks: Search asymmetries for the categories male and female were dependent on the gender of participants, indicating that the nonschematic categories constituted figure categories against the self-schematic background categories. As predicted by the model, the categories other and nonword constituted figure categories against the background categories self and word for men and women alike.

Interrelations between IAT effects, salience asymmetries, and self-reported femininity–masculinity were assessed in a series of structural equation models. These analyses revealed that salience asymmetries for the female and male categories predicted IAT effects. Salience asymmetries also accounted for the relation between IAT effects and self-reported gender self-concept. Again, the structural pattern of effects was highly similar for the original male–female IAT and for the modified version of the task. These findings support the assumption that salience asymmetries have a unique impact on IAT effects. The finding also suggests that the common variance underlying zero-order correlations between compatibility effects in male–female IATs and the explicit measure of gender self-concept is related to salience asymmetries.

Experiment 2B

Experiment 2B was a conceptual replication of the previous study and was conducted to investigate the generality of the findings in a completely different domain of content. This time, we analyzed group differences between East and West German participants. In spite of the fact that Germany was reunited more than 10 years ago, stereotypes of the typical East and West German person are prevalent throughout the country, and many East and West German citizens identify themselves with their own group and origin (see, e.g., Kessler & Mummendey, 2001; Schmitt & Maes, 2002).

Participants received two different versions of an East–West IAT (cf. Kühnen et al., 2001), crossing the target categories East and West with either self–other or with good–bad, and a third version of the task that used the categories word and nonword as attribute categories to assess East–West IAT effects with a task in which associations were eliminated. On the basis of the figure–ground model, we predicted similar group differences between East and West German participants for all three types of IAT effects. In a second part of the experiment, we assessed salience asymmetries for the pairs of categories with visual search tasks. Here, we predicted a group difference for the salience asymmetry regarding the categories East and West: For West Germans, East should be the figure category against the familiar background category West, whereas the opposite asymmetry should obtain for East Germans. Identical salience asymmetries, however, were predicted for the East and West German subsamples regarding the categories *other*–self, *bad*–good, and *nonword*–word (figure categories italicized).

As in Experiment 2A, we were also interested in the interrelations between IAT effects, salience asymmetries, and an explicit measure of attitude or self-concept. As an explicit measure of group and/or self-evaluation, each participant rated a set of positive and negative personality attributes on a scale ranging from "typically West" to "typically East." The difference in East–West typicality ratings for positive and negative attributes is an index of in-group favoritism and is indicative of personally identifying oneself as an East or West German (see, e.g., Otten & Wentura, 2001). Again, we predicted that salience asymmetries regarding the categories East and West would be associated with IAT effects and would account for potential relations between IAT effects and the explicit group evaluations.

Method

Participants. Ninety-six psychology undergraduates of the University of Jena (East German subsample, $n = 48$; 38 women, 10 men) and of the University of Trier (West German subsample, $n = 48$; 40 women, 8 men) volunteered for partial course credit. All participants from Jena and Trier confirmed that they had been born in East Germany or West Germany, respectively, and had lived there for the largest part of their life.

Materials. Stimuli of the categories self, other, good, bad, word, and nonword were identical to those of the previous experiments. Eight names of East German cities (e.g., *Cottbus*, *Stralsund*) and eight names of West German cities (e.g., *Bochum*, *Mannheim*) were selected for the categories East and West. Stimuli are listed in Appendix A.

Design and procedure. Each participant took three versions of the East German–West German IAT (East–West × other–self, East–West × bad–good, East–West × word–nonword). Response assignments in the IATs were (arbitrarily) classified as compatible when the categories East and self, East and good, or East and word were assigned the same response key. The order of presentation of the three IATs was counterbalanced across participants. Procedural details of the IATs were identical to those of Experiment 1A. Following the IATs, each participant undertook four visual search tasks for the categories East–West (first), self–other (second), good–bad (third), and word–nonword (fourth). The sequence of the search tasks was identical for all participants. Procedural details of the search tasks were identical to those of Experiment 1D. Finally, participants rated a set of personality attributes on a scale ranging from −4 (*the typical West German*) to 4 (*the typical East German*) on a computer screen (the scale had no midpoint to enforce a classification of attributes as either East or West). Five of these attributes were positive and stereotypical for an East German person (*cheerful, friendly, honest, open, reliable*), five attributes were negative and stereotypical for an East German person (*dependent, lazy, passive, pessimistic, sad*), five attributes were positive and stereotypical for a West German person (*active, contented, energetic, self-confident, strong*), and another five attributes were negative and stereotypical for a West German (*greedy, malicious, mean, miserly, repelling*).

Results

IAT analyses. Erroneous responses (East–West × self–other: 8.44%; East–West × good–bad: 8.89%; East–West × word–

nonword: 9.01%) and outlier values (East–West × self–other: 2.71%; East–West × good–bad: 3.88%; East–West × word–nonword: 3.17%; see footnote 2) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments in the different versions of the East German–West German IAT (see Table 5; assignments of the categories East and self or good or word to the same response were coded as compatible). A 2 (compatibility) × 3 (IAT type: self–other, good–bad, word–nonword) × 2 (origin of participants: East vs. West) ANOVA revealed a main effect of IAT type, $F(2, 188) = 10.85$, $MSE = 2{,}621.64$, $p < .001$, $f = .34$, indicating that on average, response times were faster for the self–other IAT ($M = 699$ ms) than for the word–nonword IAT ($M = 715$ ms) and the good–bad IAT ($M = 723$ ms). As predicted, the Compatibility × Origin of Participants interaction was also significant, $F(1, 94) = 380.03$, $MSE = 1{,}906.04$, $p < .001$, $f = 2.01$, indicating that for the East German subsample, response times were 65 ms shorter when the category East was assigned the same response as the categories self, good, or word, whereas for the West German subsample, responding was facilitated by 76 ms when the categories West and self, good, or word were assigned the same response. Although the three-way interaction Compatibility × Origin of Participants × IAT Type was significant, $F(2, 188) = 4.14$, $MSE = 1{,}481.98$, $p < .05$, $f = .21$, a similar pattern of group differences emerged for all types of IAT effects (see Table 5); the interaction Compatibility × Origin was highly significant for the self–other version of the East German–West German IAT, $F(1, 94) = 188.27$, $MSE = 1{,}178.90$, $p < .001$, $f = 1.42$; for the good–bad version, $F(1, 94) = 161.65$, $MSE = 2{,}065.74$, $p < .001$, $f = 1.31$; and for the word–nonword version of the task, $F(1, 94) = 111.21$, $MSE = 1{,}625.36$, $p < .001$, $f = 1.09$. The compatibility effects for the three versions of the East German–West German IAT were highly intercorrelated (self–other—good–bad: $r = .64$, $p < .001$; self–other—word–nonword: $r = .69$, $p < .001$; good–bad—word–nonword: $r = .62$, $p < .001$).

Search task analyses. Erroneous responses (East German cities–West German cities: 11.56%; self–other: 12.65%; good–bad: 10.53%; word–nonword: 9.42%) and outlier values (East German cities–West German cities: 1.48%; self–other: 1.20%; good–bad: 1.46%; word–nonword: 0.73%; see footnote 2; outlier criteria were determined separately for each type of search task) were excluded from the analyses. For each participant and search task, we calculated average response latencies for trials in which the distractor stimuli belonged to the figure category (East, other, bad, nonword) or to the ground category (West, self, good, word; see Table 6). Separate 2 (distractor type) × 2 (origin of participants) ANOVAs were conducted on the mean response latencies for each type of visual search task. In the East–West search task, a main effect of distractor type was found, $F(1, 94) = 6.05$, $MSE = 13{,}372.44$, $p < .05$, $f = .25$, that was qualified by an interaction with origin of participants, $F(1, 94) = 7.89$, $MSE = 13{,}372.44$, $p < .01$, $f = .29$. Analyses of simple main effects revealed that for the West German subsample, response latencies were 88 ms longer for trials containing East German city names as distractors, $t(47) = -3.43$, $p < .001$, $d = .50$, whereas for the East German subsample, response latencies were 6 ms longer for trials that contained West German city names as distractors, $t < 1$. In all other search tasks, only the main effects of distractor type were significant—other–self: $F(1, 94) = 152.97$, $MSE = 17{,}325.30$, $p < .001$, $f = 1.28$; bad–good: $F(1, 94) = 46.08$, $MSE = 13{,}620.56$, $p < .001$, $f = .70$; nonword–word: $F(1, 94) = 71.99$, $MSE = 5{,}137.93$, $p < .001$, $f = .88$; on average, response latencies were longer for trials with distractors of the figure categories other, bad, and nonword. Interactions with origin were nonsignificant in these analyses, all Fs < 1.64, indicating that salience asymmetries for these categories were similar for the East and West German subsamples.

Explicit group evaluations. A global score of East–West favoritism was computed for each participant by subtracting the average East–West typicality rating for the negative attributes from the average East–West typicality rating for the positive attributes (high and low values on the resulting difference variable represent positive evaluations of the typical East German and West German person, respectively). The East German and West German subsamples differed significantly on the resulting group evaluation score, $t(94) = 3.70$, $p < .001$, $d = .38$. On average, participants in the East German subsample rated positive attributes as being more typical of the East German person than negative attributes ($M = .62$, $SD = 1.29$), whereas West German participants showed a somewhat weaker but opposite bias ($M = -.20$, $SD = .82$).

Interrelations between IAT effects, salience asymmetries, and explicit group evaluations. Structural equation analyses were conducted to estimate interrelations between the different measures of the study. As in Experiment 2A, two effect scores were computed for each version of the IAT and for the East–West search task, based on the odd- and even-numbered trials of the tasks, respectively.[5] The stereotypical attributes were also divided

Table 5
Mean Reaction Times (in Milliseconds; Standard Deviations in Parentheses) in Compatible and Incompatible Blocks of Different Versions of an East German–West German IAT for East German and West German Participants (Experiment 2B)

Labels of target categories	Response assignment[a] Compatible	Incompatible	IAT effect
Self–other			
East Germans	678 (75)	761 (88)	83 (52)
West Germans	705 (83)	652 (79)	−53 (45)
Good–bad			
East Germans	694 (70)	768 (85)	74 (58)
West Germans	761 (95)	668 (78)	−93 (70)
Word–no word			
East Germans	701 (88)	740 (83)	39 (58)
West Germans	751 (88)	667 (72)	−84 (56)

Note. IAT = Implicit Association Test.
[a] Categories East and self, East and good, or East and word were assigned the same response in compatible blocks.

[5] To increase the reliability of the salience asymmetry measure for East German–West German cities, we computed effect variables only on the basis of those trials in which no target was presented (*same* trials). Reliability of the salience asymmetry effects in the target-present trials is markedly weaker because detecting two neighboring stimuli of different categories allows a fast different decision irrespective of which of the two categories is the distractor category in the current trial.

Table 6
Mean Response Latencies (in Milliseconds; Standard Deviations in Parentheses) in Visual Search Tasks Depending on the Salience of the Distractor Stimuli (Experiment 2B)

Categories of the search task (figure–ground)	Distractor type		Search asymmetry
	Figure	Ground	
West–East[a]			
East Germans	1,761 (335)	1,755 (370)	6 (148)
West Germans	1,708 (301)	1,796 (329)	−88 (178)
Other–self			
East Germans	1,712 (342)	1,494 (313)	218 (193)
West Germans	1,780 (372)	1,528 (313)	252 (179)
Bad–good			
East Germans	1,690 (292)	1,562 (295)	128 (158)
West Germans	1,771 (324)	1,670 (342)	101 (172)
Nonword–word			
East Germans	1,463 (228)	1,362 (230)	101 (111)
West Germans	1,503 (234)	1,429 (237)	75 (90)

[a] The category West was arbitrarily assigned the status as figure category for both East German and West German participants.

into two sets of items, respectively, to obtain two parallel indicators of East–West favoritism. All indicators were standardized before they were entered into the structural equation models to obtain latent variables with comparable variances.

A first structural equation model revealed that East–West favoritism predicted IAT effects in the East–West IAT tasks (see Figure 2A). In a second step, we entered search asymmetries for the categories East and West into the equation as an additional predictor (see Figure 2B). To a large degree, the findings parallel the results of Experiment 2A. First, salience asymmetries for the categories East and West were positively correlated with the group evaluation score, indicating that explicit East–West favoritism was associated with corresponding salience asymmetries in the West German–East German city names in the search task. Second, salience asymmetries predicted compatibility effects in the East–West IATs over and above effects of the East–West favoritism score. Third, introducing the salience asymmetries into the equations rendered effects of East–West favoritism on the effect variables in the self–other and word–nonword IATs nonsignificant and reduced the strength of the effect of the favoritism score on the good–bad IAT.

Discussion

Experiment 2B yielded a conceptual replication of the findings of Experiment 2A. A highly similar pattern of group differences between East German and West German participants was found regarding the compatibility effects in the different variants of an East German–West German IAT, including the word–nonword variant of this task. The latter finding is difficult to reconcile with an association account of the IAT effect but was predicted on the basis of the figure–ground model.

The pattern of salience asymmetries that is assumed to underlie the group differences in the East–West IATs was investigated in a series of visual search tasks: As predicted, salience asymmetries for the categories East and West differed between the East and West German subsamples, indicating that the nonfamiliar categories constituted figure categories against the familiar background categories. For the category pairs *other*–self, *bad*–good, and *nonword*–word (figure categories in italics), we found homogeneous salience asymmetries for both subsamples.

Interrelations between IAT effects, salience asymmetries, and explicit East–West favoritism or group identification were assessed in a series of structural equation models. These analyses revealed that salience asymmetries for the East and West categories were associated with IAT effects. Salience asymmetries also accounted for a large part of the common variance between the explicit attitude measure and IAT effects. The structural pattern of effects was similar for the different versions of the East German–West German IAT.

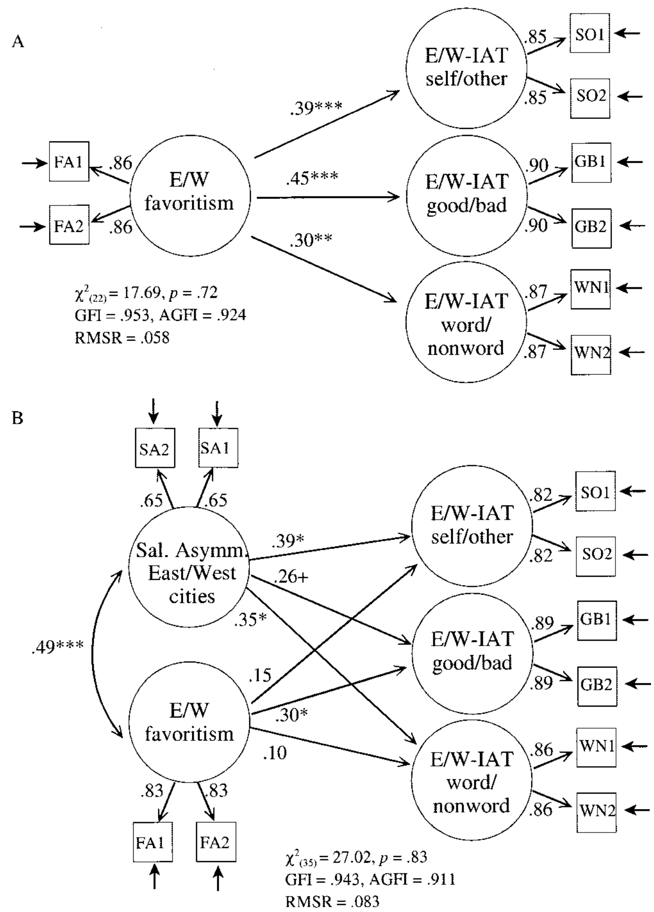

Figure 2. Structural equation models estimating the interrelations between East German–West German favoritism, salience asymmetries regarding East German and West German cities, and Implicit Association Test (IAT) effects in different versions of an East German–West German IAT. A: A first structural equation model revealing that East–West favoritism predicted IAT effects in the East–West IAT tasks. B: A second step in which search asymmetries were entered into the equation for the categories East and West as an additional predictor. AGFI = adjusted goodness of fit index; E/W = East German/West German; FA = favoritism; GB = good/bad; GFI = goodness of fit index; RMSR = root-mean-square residual; SA = salience asymmetries; Sal. Asymm. = salience asymmetries; SO = self/other; WN = word/nonword. $+p < .05$, one-tailed. $*p < .05$. $**p < .01$. $***p < .001$.

Taken together, the results of Experiments 2A and 2B provide evidence that figure–ground asymmetries are an important factor in explaining interindividual differences in IAT effects. The analyses revealed that interindividual or group differences in salience asymmetries are at least sometimes responsible for interindividual differences in IAT effects and might also account for associations of IAT effects and self-report measures. We want to stress, however, that the reported studies are only exemplary in character and do not warrant general conclusions regarding the relation of self-report measures, IAT effects, and salience asymmetries. Different patterns of association between compatibility effects, search asymmetries, and explicit measures of self-concept or evaluative associations might be found for other IAT tasks, search tasks, and self-report measures. However, the results of the Experiments 2A and 2B suggest that salience asymmetries can be a proximal cause of IAT effects and that these asymmetries can account for some of the common variance in IAT effects and self-report measures of attitudes or the self-concept.

The previous studies do not shed any light on the causal origin of salience asymmetries. If self-report measures of evaluative (or other) associations and salience asymmetries are related, this could have various reasons: For instance, salience asymmetries might have their origins in evaluative (or other) associations. Alternatively, salience asymmetries might influence IAT effects and measures of associations in a similar way. A third possibility is that salience asymmetries and associations might mutually influence each other. A clarification of causal sequences is beyond the scope of the previous studies.

In sum, the first two series of experiments eliminated the influence of associations on IAT effects by replacing associated categories with neutral categories that differed in salience. The attribute categories of nonwords versus words and multicolored versus single-colored strings were chosen to reflect clear figure–ground asymmetries, but they possibly also differed on other dimensions that might have contributed to the observed effect. To rule out such possible confounds, we conducted a third series of studies, in which we manipulated figure–ground asymmetries experimentally within a fixed set of categories and stimuli. Varying salience asymmetries while holding associations and other factors constant was a further step toward establishing that salience asymmetries have a unique impact on IAT effects, an impact that does not have its origins in associations.

Experiment 3A

A standard old–young IAT with the categories old names (old), young names (young), pleasant words (good), and unpleasant words (bad) was used for the present experiment. Before conducting the IAT, we manipulated figure–ground asymmetries within the target and attribute categories by a go/no-go search task for the target and attribute dimensions, respectively. In these search tasks, participants had to search for only one of the categories within each dimension (see Figure 3). Searching for exemplars of only one of two categories should establish an attentional focus toward the particular category (this kind of top-down effect on selective attention has repeatedly been demonstrated with the visual search paradigm; for reviews, see, e.g., Pashler, 1998; Yantis, 1998). The figure–ground model thus predicts that the sign of the IAT effects should depend on the figure–ground asymmetries established by the previous go/no-go tasks. Responses in the IAT should be faster when the figure categories of the previous go/no-go tasks are assigned the same response in the subsequent IAT task, regardless of which categories are chosen as figure categories. The association model does not predict an influence of previous figure–ground manipulations on the direction or size of the IAT effect.

Block	Category Label		Stimuli	Response Rules
1		young	old/young names	young → right old → no response
2	bad		pleasant/unpleasant words	pleasant → no response unpleasant → left
3	not young bad	young not bad	old/young names + pleasant/unpleasant words	young → right old → left pleasant → right unpleasant → left
4	young		old/young names	young → left old → no response
5	bad		pleasant/unpleasant words	pleasant → no response unpleasant → left
6	young bad	not young not bad	old/young names + pleasant/unpleasant words	young → left old → right pleasant → right unpleasant → left

Figure 3. Schematic overview of the sequence of go/no-go tasks and combined categorization tasks for a condition in which young and bad are established as figure categories (Experiment 3A).

Method

Participants. Thirty-two participants (23 women, 9 men) were recruited through an advertisement in a local newspaper. Most of them were students of different faculties of the University of Münster, Germany. They were paid DM 20 (about $10) for their participation in this experiment and a second, unrelated experiment.

Materials and design. Stimuli and design were identical to those of Experiment 1A. In addition, the sample was randomly divided into four groups, each of which received one of the four possible combinations of go/no-go tasks before conducting the combined tasks of the IAT. Two of these conditions created figure–ground asymmetries that were consistent with the default figure–ground asymmetries (old names–unpleasant words, young names–pleasant words), whereas the other conditions created figure–ground asymmetries that were inconsistent with the default figure–ground asymmetries (young names–unpleasant words, old names–pleasant words).

Procedure. Figure 3 gives an overview of the experiment. The practice blocks introducing the target and attribute stimuli were now presented as go/no-go detection tasks. In the first block, the label of the search category for the target dimension was shown in one of the upper corners of the screen. Old and young names were presented in random order, and participants had to press the corner-corresponding key whenever the name of a search category was presented. No response should have been executed for names of the other category. In a second block, the same type of go/no-go task was conducted with the stimuli of the attribute dimension. During the go/no-go blocks, each stimulus was presented for 1,000 ms or until a response key was pressed. An error message was displayed on the screen whenever a response key was pressed for a stimulus of the nonsearched category. Each stimulus was presented four times in the corresponding search task, yielding a total of 80 trials per block. The third block was analogous to the combined classification task in a standard old–young IAT. Two category labels were shown on the upper left side of the screen; the other two category labels were shown on the upper right side of the screen. The position (and assigned key) of the two search categories corresponded to the practice blocks. Stimuli of all four categories were presented at random, and each stimulus required a response according to the position of the corresponding category label on the screen. In the third block, each stimulus was presented twice, yielding a total of 80 trials; 10 additional trials were presented as practice trials. In the fourth and fifth blocks, the go/no-go tasks of the first and second blocks were presented again, with a reversed response assignment for either the target or the attribute dimension. The search categories always remained unchanged throughout the experiment for each participant. In the sixth block, a combined classification task was conducted as in the third block but with the reversed response assignment. Presentation parameters during the combined classification tasks were identical to those of Experiment 1A. Two provisions were taken to maintain the figure–ground asymmetries that were established in the go/no-go tasks during the combined classification tasks as well. Participants could gain points by correctly identifying stimuli of the go categories. Correct classifications for the stimuli of the background categories did not yield points. In addition, negated labels (e.g., “not old” instead of “young”) were used for the background categories in a somewhat smaller type font. These provisions should have stressed the importance of attending primarily to the categories that represented the figures.

Results

Erroneous responses (4.55%) and outlier values (3.22%; see footnote 2) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments. Compatibility of response assignments was coded according to the default, that is, assigning old names and unpleasant words the same response key was always coded as a compatible response assignment. Mean reaction times for compatible and incompatible response assignments for the different figure–ground asymmetries established by the previous go/no-go tasks are shown in Table 7. A 2 (compatibility) × 2 (figure–ground asymmetries: consistent vs. inconsistent with default) ANOVA revealed a main effect of compatibility, $F(1, 30) = 16.52$, $MSE = 12{,}288.29$, $p < .001$, $f = .74$, that was qualified by an interaction with the type of figure–ground asymmetries, $F(1, 30) = 76.90$, $MSE = 12{,}288.29$, $p < .001$, $f = 1.60$. An analysis of simple main effects revealed a positive effect of compatibility if the established figure–ground asymmetries were consistent with the default asymmetries, $t(15) = 10.13$, $p < .001$, $d = 2.53$. When figure–ground asymmetries that were inconsistent with the default were created, the sign of the compatibility effect was reversed, $t(15) = -3.04$, $p < .01$, $d = .76$.

Table 7
Mean Reaction Times (in Milliseconds; Standard Deviations in Parentheses) in Compatible and Incompatible Blocks of Standard Old–Young IATs, Depending on Previously Established Figure–Ground Asymmetries (Experiment 3A)

	Response assignment[a]		
Figure–ground asymmetries	Compatible	Incompatible	IAT effect
Consistent with default			
Old–bad	649 (81)	1,012 (128)	363 (99)
Young–good	708 (98)	1,056 (187)	349 (180)
Inconsistent with default			
Old–good	880 (137)	689 (75)	−190 (139)
Young–bad	860 (122)	790 (145)	−70 (189)

Note. IAT = Implicit Association Test.
[a] Categories old and bad were assigned the same response in compatible blocks.

Discussion

The results provide direct evidence in support of the figure–ground model. The sign of IAT effects was found to depend on the type of figure–ground asymmetries that had been previously created by the go/no-go tasks. Standard IAT effects were observed if the figure–ground asymmetries were consistent with the default, but IAT effects were reversed if, for either the target or the attribute dimension, a figure–ground asymmetry had been established that was opposite to the default. This rules out alternative explanations of the observed reversal of IAT effects in terms of background variables related to the materials used (e.g., associations, valence, familiarity, etc.). The present results replicate and extend findings from a previous pilot study in which a reversed compatibility effect was observed after inverting the figure–ground asymmetry for the old–young dimension (Rothermund & Wentura, 2001). Meanwhile, a conceptual replication of these findings has also been reported by Kinoshita and Peek-O’Leary (2002), who found reversed IAT effects for the insect–flower IAT and for an odd–even IAT after experimentally reversing the figure–ground asymmetries by the means of a go/no-go task.

A main effect of compatibility was nevertheless observed in the present experiment, indicating that compatibility effects are stronger when the established figure–ground asymmetries correspond to the default. Presumably, it is difficult for participants not to attend to categories that are salient by default (Theeuwes, 1996; Yantis, 1996). In the case of inconsistent figure–ground asymmetries, categories that are figure categories by default might tend to interfere with the attentional control settings required by the task. This kind of interference undermines the potential benefit of a consistent mapping of experimentally created figures onto responses. The main effect of compatibility thus might reflect automatic attentional capture by salient categories that have not been used as figure categories. Of course, valence might also contribute to the observed effect, which would explain why IAT effects are stronger when figure–ground asymmetries correspond to the default.

To rule out these problems, we conducted another experiment using simple colors as attribute categories. These categories are not characterized by a strong default asymmetry regarding salience, which also allows a more subtle manipulation of salience that does not involve a change of category labels or other changes in the procedural parameters of the IAT. Furthermore, compatibility effects with neutral materials would add to the evidence that IAT effects do not require associations between the target and attribute categories.

Experiment 3B

A modified version of the old–young IAT was used for Experiment 3B. Neutral words written in either yellow or green were presented as attribute stimuli instead of pleasant and unpleasant words. Salience asymmetries between the color categories were created by the help of a subtle context manipulation involving only an irrelevant feature of the target stimuli: In one variant of the IAT, all stimuli of the target categories (i.e., old and young names) were written in yellow. This should have established yellow as the default or background color against which the color green should have popped out as the more salient attribute category. In another variant of the IAT, all stimuli of the target categories were written in green, which should have established the category yellow as the figure category. Each participant received both variants of the IAT. The figure–ground model predicts a reversal of the sign of the compatibility effects from the first to the second variant of the IAT because the context manipulation should invert the salience asymmetry of the color categories. The association model does not predict compatibility effects in either the first or the second IAT because the target categories (old vs. young) are unrelated to the attribute categories (yellow vs. green). Furthermore, it does not predict a difference in compatibility effects because stimuli, labels, and procedural parameters do not differ between the two variants of the task.

Method

Participants. Forty-eight University of Trier psychology undergraduates (40 women, 8 men) volunteered for partial course credit.

Materials. Old and young names were identical to the stimuli used in the previous experiments. Depending on condition, the names were presented either in yellow or in green. In addition, a set of 10 neutral words was selected for both categories of the color categorization task (see Appendix A, Colored Words). The same neutral words were written in yellow for the color category yellow and in green for the color category green.

Design and procedure. Each participant received two variants of a modified old–young IAT. In one version, old and young names were presented in yellow; in the other version, all target stimuli were presented in green. Order of presentation of the two versions of the IAT was counterbalanced across participants. Responses were coded as compatible (incompatible) if the categories old and the salient (nonsalient) color category of the first IAT were assigned the same response key. During the second IAT, compatibility was coded according to the same rules as in the first IAT. The salience asymmetry of the color categories, however, was now reversed. All stimuli were presented with initial letters in uppercase. In all other respects, the procedural details of both variants of the IAT were identical to those of Experiment 1A.

Results

Erroneous responses (8.17%) and outlier values (2.38%; see footnote 2) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments separately for the first and second IATs (see Table 8). No significant effects were observed for the order of presentation of the two versions of the IAT, all *F*s < 2.86, *ns.* To simplify the presentation of the results, we report results that are aggregated across this factor. A 2 (compatibility) $\times$ 2 (block: initial salience asymmetries vs. reversal of salience asymmetries) ANOVA yielded only a significant interaction of Compatibility $\times$ Block, $F(1, 47) = 13.08$, $MSE = 2{,}305.10$, $p < .001$, $f = .53$. An analysis of simple main effects revealed a significant compatibility effect during the first IAT, $t(47) = 2.77$, $p < .01$, $d = .40$, and a reversed compatibility effect in the second IAT, $t(47) = -2.12$, $p < .05$, $d = .31$.

Table 8
Mean Reaction Times (in Milliseconds; Standard Deviations in Parentheses) in Compatible and Incompatible Blocks of Modified Old–Young IATs, for the First (Initial Color Asymmetries) and Second IAT (Reversed Color Asymmetries; Experiment 3B)

Context-dependent figure–ground asymmetries between the color categories	Response assignment[a]		
	Compatible	Incompatible	IAT effect
First IAT (initial color asymmetries)	719 (96)	748 (95)	29 (72)
Second IAT (reversed color asymmetries)	739 (100)	718 (89)	−21 (69)

Note. IAT = Implicit Association Test.
[a] The category old and the salient color category of the first IAT were assigned the same response key in compatible blocks.

Discussion

Manipulating the salience asymmetries of neutral perceptual attribute categories in a modified old–young IAT produced compatibility effects with opposite signs. Two aspects of this finding are noteworthy. First, significant compatibility effects were observed in a modified old–young IAT although there was no association between the target and attribute categories. This finding converges with the assumption that compatibility effects in the IAT do not require the existence of these associations. Second, the sign of the compatibility effects depended on subtle context manipulations that were intended to create specific figure–ground asymmetries between the two color categories. Establishing one of the colors as the figure category by presenting the old and young names in the other color led to faster response latencies when the figure category old and the salient color category were assigned the same response. This finding strongly supports the figure–ground model. It is difficult to provide an alternative explanation for this finding because the reversal of the IAT effects depended only on the manipulation of the color context; all other parameters of the task remained unchanged.

In sum, the first three series of experiments have demonstrated an influence of existing and experimentally created figure–ground asymmetries on IAT effects while simultaneously excluding effects of valence or associations. These experiments have revealed a unique contribution of salience on IAT effects. In the following studies, we manipulated the valence of the target categories while holding figure–ground asymmetries constant. This type of dissociation allowed another test of whether effects of valence in the IAT were mediated by salience. If effects of valence were in fact mediated by figure–ground asymmetries, controlling for salience asymmetries between categories should weaken or eliminate the influence of valent associations on IAT effects.

This kind of dissociation effect has already been demonstrated in an experiment by De Houwer (2001), who varied the valence of target stimuli within target categories: In his experiment, half of the stimuli of each target category (British vs. foreign) referred to negative and positive stimuli, respectively. An overall IAT effect emerged, indicating a positive evaluation of the category British that was independent of the valence of the stimuli. Specifically, the same pattern of compatibility effects was found for the negative and the positive stimuli of the category British. From the perspective of the figure–ground model, the results of De Houwer reflect a salience asymmetry between the target categories. For British participants, names of British persons are familiar and form a background category, whereas the category foreign forms the figure—irrespective of the valence of the stimuli. A possible objection to this interpretation of the findings, however, stems from a recent account of IAT effects in terms of differences in decision thresholds that might vary on a global level between compatible and incompatible blocks of the IAT (Brendl et al., 2001). On the basis of this account, IAT effects for a specific subset of the presented stimuli are difficult to interpret. In the following experiment, we tried to avoid this problem.

Experiment 4

A standard IAT was conducted with names referring to publicly known persons and names of unknown persons as target categories; pleasant and unpleasant words were used as attribute categories. The evaluation of the category of well-known names was manipulated experimentally on a between-subjects basis: For one group of participants, all well-known names referred to persons with a clear positive evaluation, whereas for another group of participants, all well-known names referred to persons with a clear negative evaluation. Varying the evaluations on a between-participants basis allowed an analysis and comparison of global IAT effects and did not require an analysis of compatibility effects for subsets of stimuli. Manipulating the valence of the well-known names, however, should not have altered salience asymmetries between the categories of known and unknown names: Processing names that do not refer to a known person should bind attentional resources because no perceptual or semantic schema was available for these stimuli. Unknown names should thus have formed a salient category. Processing the names of well-known persons, by contrast, should not have required attentional resources. Accordingly, these stimuli should form the background—regardless of their valence.

An account of the processes underlying IAT effects based on automatic evaluations predicts opposite IAT effects for target categories consisting only of positively evaluated stimuli and only of negatively evaluated stimuli. However, if effects of valence on the IAT are mediated by figure–ground asymmetries, changing the valence of a category without changing its salience should not have an influence on the resulting IAT effect: Responses should be faster if the figure categories of unfamiliar names and unpleasant words are assigned the same response, even when the category of familiar names consists only of negative stimuli.

An additional experiment was conducted to confirm the presumed figure–ground asymmetries between the known and unknown names. As in Experiment 1D, the names of known and unknown persons were presented as stimuli in a visual search task. Participants had to decide whether all names of a display belonged to the same category or not. For one group of participants, only the positively valent familiar names were presented together with the unknown names; the other group of participants received only the negatively valent names and the neutral names. It was predicted that it would be easier to detect an unknown name among familiar distractors than the other way around, regardless of the valence of the names of the well-known persons.

Method

Participants. Thirty-two University of Trier psychology undergraduates (26 women, 6 men) participated in the IAT experiment for partial course credit. Another 43 University of Trier psychology undergraduates (27 women, 16 men) participated in the search task experiment. These participants were a subsample of the participants in Experiment 1D. Each participant also received another visual search task during this experiment. The search tasks with the known and unknown names, however, were never presented after having received a search task with unpleasant and pleasant words.

Materials. Names of well-known persons were selected on the basis of a pilot study in which evaluative ratings (7-point scale: −3 = *unequivocally negative,* 3 = *unequivocally positive*) were gathered for a large number of names of well-known persons from different domains of public life, for example, politics, sports, literature, music, movies, and television. Ten clearly positive names ($M = 1.63$; range: 0.94 to 2.38) and 10 clearly negative names ($M = -1.49$; range: −0.64 to −2.33) were selected as target stimuli for the IAT experiment (all names are listed in Appendix B). In addition, 10 unknown names were generated. Although the unknown names did not refer to publicly known people, the names themselves were not uncommon and consisted of combinations of familiar German first names and surnames (see Appendix B). No participant of the pilot study identified one of these names as referring to a known person. Ten pleasant and 10 unpleasant words were selected for the attribute categories.

Design and procedure. A standard IAT was conducted in which *bekannt* (known) and *unbekannt* (unknown) served as target categories, and good and bad were used as attribute categories. The design was identical to that of the previous simple IAT experiments. In addition, the sample was randomly divided into two groups. For one group, only the names of positively evaluated persons were used as stimuli for the category known, whereas for the other group, only the names of negatively evaluated persons were presented. Presentation parameters and procedure were identical to those of Experiment 1A. For the search task experiment, design and procedure were identical to those of Experiment 1D.

Results

IAT analyses. Erroneous responses (8.70%) and outlier values (2.59%; see footnote 2) were excluded from the analysis. For each participant, we calculated mean response latencies for compatible and incompatible response assignments. Compatibility of response assignments was coded according to the figure–ground asymmetries, that is, assigning the same response key to unknown names and unpleasant words was always coded as a compatible response assignment, irrespective of the valence of the known names. Mean reaction times for compatible and incompatible response assignments for the two groups that received positively and negatively evaluated names of well-known people are shown in Table 9. A 2 (compatibility) × 2 (valence of the known names) ANOVA revealed a main effect of compatibility, $F(1, 30) = 57.49$, $MSE = 12{,}870.13$, $p < .001$, $f = 1.38$. Response latencies were 152 ms ($d = 1.35$) faster for compatible response assignments. The effect

Table 9
Mean Reaction Times (in Milliseconds; Standard Deviations in Parentheses) in Compatible and Incompatible Blocks of IATs With Names Referring to Well-Known Persons Versus Unknown Persons, Depending on the Valence of the Familiar Names (Experiment 4)

Evaluation of known names	Response assignment[a] Compatible	Incompatible	IAT effect
Positive	693 (122)	828 (131)	135 (99)
Negative	736 (99)	906 (184)	170 (127)

Note. IAT = Implicit Association Test.
[a] Categories unknown and bad were assigned the same response in compatible blocks.

of compatibility was not qualified by the valence of the category of known names, $F < 1$. An IAT effect of 135 ms ($d = 1.37$) was observed when the category known consisted entirely of names with a clearly positive evaluation, and an IAT effect of 170 ms ($d = 1.34$) in the same direction was found when the category known consisted entirely of names with a clearly negative evaluation.

Search task analyses. Erroneous responses (8.79%) and outlier values (0.87%; see footnote 2) were excluded from the analyses. For each participant, we calculated mean response latencies for trials in which the distractor stimuli belonged to the figure category (names of unknown persons) or to the ground category (positive or negative familiar names; see Table 2). Separate *t* tests comparing the response latencies for trials with figure distractors and ground distractors were conducted for each type of the visual search task. In both cases, response latencies were significantly longer for trials with figure distractors, unknown versus familiar–positive: $t(21) = 4.33$, $p < .001$, $d = .92$; unknown versus familiar–negative: $t(20) = 4.54$, $p < .001$, $d = .99$.

Discussion

The main point of Experiment 4 was to investigate whether evaluative associations influenced IAT effects independently of salience asymmetries. Although previous studies established that differences in the strength of evaluative associations between the target and attribute stimuli can have an influence on the magnitude of the resulting IAT effects (Steffens & Plewe, 2001), it is unclear whether this relation is a genuine effect of evaluative associations or whether it reflects related differences in salience asymmetries. The results of Experiment 4 suggest that positive or negative evaluations of the target stimuli do not modulate IAT effects independently of figure–ground asymmetries. Varying the evaluations of the stimuli of the target category known had no influence on the sign or the magnitude of the observed IAT effect, even though we used a very powerful manipulation of evaluative associations by presenting two sets of target stimuli of completely opposite valence. The results of the additional search task experiment confirm that salience asymmetries between the categories unknown and known did not depend on the valence of the well-known persons: It took longer to process the names of unknown persons than the names of well-known persons regardless of the valence of the known persons.

The pattern of findings is difficult to reconcile with the association model of IAT effects. Proponents of this account might perhaps assume that explicit negative evaluations of familiar names reported in the pilot study do not reflect true attitudes because, in fact, names of well-known persons are always associated with a positive implicit attitude. In our opinion, however, such an assumption would stretch the concept of implicit attitudes beyond its limits. One might argue, however, that IAT effects are mainly driven by the category labels rather than by the exemplar stimuli that are used to represent the categories (De Houwer, 2001; Mitchell, Nosek, & Banaji, 2003; Neumann, 1999; but see Steffens & Plewe, 2001). Because identical labels were used in both conditions of Experiment 4, this perspective offers an explanation why IAT effects were similar regardless of the valence of the target stimuli. Whether or to what extent IAT effects are driven by exemplar stimuli or by category labels is a difficult question and is clearly beyond the scope of the present article. We do not want to rule out that the category labels might have influenced the IAT effect in Experiment 4. It seems unlikely, however, that fairly neutral labels like known versus unknown could have produced such large IAT effects because of a difference in their valence. That is, even if the category labels contributed to the IAT effects, this probably reflects a figure–ground asymmetry between the categories rather than a difference in valence.[6]

The present results yield additional support for the figure–ground model. The direction of the IAT effects corresponds to the prediction of the figure–ground model: Responses were faster when the figure categories (unknown and bad) were assigned the same response. Because this effect did not depend on the valence of the target stimuli, this finding is a conceptual replication of the results of Experiments 3A and 3B. We want to stress, however, that the results of Experiment 4 do not allow any general conclusions regarding the relation between evaluative associations, salience asymmetries, and IAT effects. Experiment 4 illustrates, for a specific type of IAT, that varying evaluative associations has no influence on the resulting IAT effects when salience asymmetries are held constant. Whether this relation also holds for other types of IATs is a question for future research.

General Discussion

Four series of experiments that aimed at dissociating the influence of associations and figure–ground asymmetries on IAT effects have been reported. In Experiments 1B, 1C, and 1E, an influence of associations was eliminated by replacing the standard attribute categories of pleasant–unpleasant words with neutral dichotomies (nonwords vs. neutral words, multicolored vs. single-colored strings). In Experiments 2A and 2B, associations were eliminated by replacing the categories self versus other and good versus bad of standard male–female and East German–West German IATs with the dichotomy of word versus nonword. In spite of

[6] Effects of category labels on IAT effects can easily be incorporated into the figure–ground model by assuming that the labels can determine the way in which the items are processed and which features of the items attract attention. Accordingly, particular labels might conceal salience asymmetries that might emerge with other labels.

this removal of any plausible associations between the target and attribute categories, IAT effects that were in line with predictions of the figure–ground model emerged: Faster responses were observed for a consistent mapping of salient (and nonsalient) categories onto responses.

Experiments 2A and 2B also provided evidence that salience asymmetries can sometimes account for the relation between IAT effects and explicit measures. We were able to show that IAT effects, salience asymmetries (measured with the visual search task), and self-report measures of associations (gender self-concept, in-group favoritism) were intercorrelated. Structural equation models revealed that the salience measures were better predictors of IAT effects than the self-report measures and eliminated or reduced effects of the latter variables on IAT effects in four of five cases.

Experiments 3A and 3B attempted a direct manipulation of salience asymmetries while holding categories and stimuli constant, thereby controlling for effects of evaluative associations. Again, the sign of IAT effects varied as a function of experimentally created figure–ground asymmetries. The standard old–young IAT effect was reversed after nondefault salience asymmetries had been established for either the old and young names or for the pleasant and unpleasant words (Experiment 3A). A reversal of compatibility effects was also observed in a modified old–young IAT after inverting the salience asymmetries of simple color categories (Experiment 3B).

Finally, Experiment 4 revealed that varying the valence of a target category while holding figure–ground asymmetries constant had no impact on the sign or magnitude of IAT effects: Similar compatibility effects emerged in a standard IAT contrasting unknown neutral names with positively or negatively valenced names of well-known persons.

Additional experiments using visual search tasks that allowed an independent estimation of figure–ground asymmetries between the categories of the IAT experiments were conducted. In each case, the presumed salience asymmetry between categories was confirmed: Detecting stimuli of the salient categories among nonsalient distractors was always faster than detecting stimuli of the nonsalient categories among salient distractors (Experiments 1D, 2A, 2B, and 4).

Taken together, the results support the hypothesis that figure–ground asymmetries have a unique influence on IAT effects. Salience asymmetries between the categories of the target and the attribute dimensions seem to play an important role in the emergence of compatibility effects in the IAT. In our studies, with one exception (Experiment 2B), associations did not have an impact on IAT effects independently of differences in category salience, which casts some doubt on the hypothesis that the processes underlying IAT effects always depend on the valence of the categories. In Experiment 2B, a residue of a valence effect remained after controlling for effects of salience asymmetries in a structural equation model. We return to this issue below.

The remainder of the General Discussion is structured as follows: First, since any alternative explanation of a paradigm's basic results gets its worth from its scope of explanation, we want to discuss previous IAT findings in terms of figure–ground asymmetries. In our view, many of the most interesting findings can be explained by the figure–ground account. In the next step, the association account and the figure–ground account are pitted against each other. We want to define the scopes of applicability of both accounts. Third, we point out that figure–ground asymmetries reveal basic features of human mental representations to prevent the impression that we want to trivialize findings gotten with the IAT.

Scope of the Figure–Ground Model: Interpreting Previous Research Findings With the IAT

A multitude of studies using the IAT have been published recently that address a large number of interesting research questions. In the following, we discuss some selected research findings and try to give an account of the results in terms of the figure–ground model.

Associations. The present experiments were based on only a small set of IATs (old vs. young names, male vs. female attributes, East vs. West German cities, well-known vs. unknown names). The scope of the figure–ground model, however, is not limited to the specific target and attribute categories of the present experiments. In principle, an account of compatibility effects in an IAT can be construed in most cases, provided that salience asymmetries exist between the categories of the target and attribute dimensions. In IATs that investigate evaluative associations or automatic prejudice (regarding social groups, the self, certain objects, political leaders, etc.), such an asymmetry can be presumed to exist at least for the attribute dimension: As was shown in Experiment 1D, stimuli of the category good are less salient than stimuli of the category bad. As we argue below, salience asymmetries might also exist between many of the target categories that are typically investigated in this kind of research. We therefore expect that future research could uncover salience asymmetries for many of the target dimensions of previous IATs. Similar arguments also apply to uses of the IAT as a measure of nonevaluative associations between categories. Again, explanations of these effects in terms of the figure–ground model presuppose the existence of salience asymmetries within the target and attribute categories. For example, Experiments 2A and 2B revealed salience asymmetries between the categories other and self, which play an important role in many IATs related to self-concept research (e.g., Asendorpf et al., 2002; Bosson, Swann, & Pennebaker, 2000; Greenwald et al., 2002; Greenwald & Farnham, 2000; Nosek et al., 2002; Rudman, Greenwald, & McGhee, 2001). Salience asymmetries are thus a promising alternative that might explain self-concept-related IAT effects. Exemplary demonstrations of such an influence of salience asymmetries in nonevaluative IATs were made for the gender self-concept IAT in Experiment 2A and for the East German–West German identity in Experiment 2B.

Context effects. Recent findings also reveal that compatibility effects in the IAT are highly susceptible to context manipulations. For example, IAT effects can be reduced, neutralized, or reversed by a confrontation with salient exemplars of the categories (Dasgupta & Greenwald, 2001; Lowery, Hardin, & Sinclair, 2001), by specific semantic or pictorial backgrounds (Wittenbrink, Judd, & Park, 2001), by focused mental imagery concerning the target categories (Blair, Ma, & Lenton, 2001; Kühnen et al., 2001), or after being massively exposed to persuasive messages regarding the target categories (Rudman, Ashmore, & Gary, 2001). Context effects on the size or direction of IAT effects were also observed in the present research (Experiments 3A and 3B). Seen from the perspective of the figure–ground model, moderating effects of the context on IAT effects are mediated by changes in salience asymmetries within the target or attribute categories. For example, the presence of a Black experimenter was found to reduce IAT effects in a Black–White IAT (Lowery et al., 2001). Part of this effect might stem from the fact that a Black experimenter increased the accessibility of Black names, thus decreasing the salience of the category of Black names.[7] Similar influences on category salience might be produced by the context manipulations of other studies.

Familiarity. As we have argued in the introductory section, above, salience asymmetries might often be related to differences in familiarity. For example, the results of the Japanese–Korean IAT (Greenwald et al., 1998) can be explained by assuming that for people who are strongly immersed in the Japanese culture, Korean names are unfamiliar and form a figure against the background of familiar Japanese names, whereas for people with a Korean background, the reverse is true. Similarly, the finding that African Americans show no or even an inverted Black–White IAT effect (Nosek et al., 2002) may reflect group differences in the familiarity of same-race and cross-race faces that give rise to corresponding salience asymmetries (e.g., Levin, 2000, found that in a visual search task, Black faces were more easily detected by White participants, whereas the opposite search asymmetry emerged for native African participants).

However, the relation between familiarity and IAT effects has been investigated in a number of recent studies that seem to undermine the presumed relation between familiarity and IAT effects. Ottaway, Hayden, and Oakes (2001) found reliable IAT effects although the stimuli of the two target categories were similar with regard to familiarity. Rudman et al. (1999) found no correlation between self-reported familiarity of the target concepts and IAT effects and also found IAT effects of comparable magnitude when the familiarity of the stimuli of the two target categories was varied orthogonally. Dasgupta, McGhee, Greenwald, and Banaji (2000) regressed IAT effects onto a measure of differential familiarity for the stimuli of the two target categories. The regression equation yielded a positive intercept, indicating a significant IAT effect in case of equal familiarity, and a positive slope, indicating an increase in the magnitude of the IAT effect with increasing differential familiarity. The findings of these studies seem to suggest that IAT effects cannot be attributed to asymmetries in familiarity alone. A number of caveats, however, should be taken into account when evaluating the results of these studies: First, the studies investigated only the familiarity of the stimuli. Findings by De Houwer (2001), Mitchell et al. (2003), and Neumann (1999), however, suggest that the categories are also important for IAT effects. Second, differences in familiarity between target categories were not manipulated directly in the studies by Dasgupta et al. and Ottaway et al., which precludes strong causal interpretations of the findings. Third, the differential familiarity measures that were used by Dasgupta et al. and Rudman et al. are only rough indicators that do not yield error-free estimates of differences in familiarity. Error variance in a predictor variable, however, leads to an underestimation of the slope parameter and a corresponding overestimation of the intercept (Klauer, Draine, & Greenwald, 1998). The regression analyses reported by Dasgupta et al. and Rudman et al. might thus be biased in favor of a familiarity-independent IAT effect and against finding an effect of familiarity. Therefore, although these studies suggest that IAT effects cannot be explained by familiarity alone, they do not show that familiarity is unimportant for IAT effects (for related arguments and findings, see Kinoshita & Peak-O'Leary, 2002).

Correlations with explicit attitude measures and behavior. Another important issue relates to the predictive validity of the IAT. Various studies have investigated correlations of IAT effects with explicit ratings (relating to, e.g., attitudes, self-esteem, or other attributes of the self), group membership, and behavioral indicators. The evidence on this issue to date is mixed. Whereas some studies have reported fairly strong correlations between IAT effects and reference variables (e.g., Banse, Seise, & Zerbes, 2001; McConnell & Leibold, 2001; Neumann et al., 1998; Rudman et al., 1999; for a review, see Greenwald & Nosek, 2001), other studies have failed to find these associations (e.g., Bosson et al., 2000; Karpinski & Hilton, 2001), have found only moderate associations (Nosek et al., 2003), or have identified moderating conditions of the relation between IAT effects and explicit ratings or behavior (e.g., Asendorpf et al., 2002; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Florack, Scarabis, & Bless, 2001; Marsh, Johnson, & Scott-Sheldon, 2001; Neumann & Seibt, 2001).

How can the figure–ground model account for these findings? Prima facie, the figure–ground model does not predict correlations between explicit measures of, for example, positive or negative attitudes and corresponding IAT effects because IAT effects are assumed to be based on salience asymmetries and not on evaluative differences. Therefore, studies showing no or only weak covariations between explicit measures and IAT effects easily fit with our approach. Finding correlations, however, is not a strong case against the model. Assuming that IAT effects are based on figure–ground asymmetries, correlations between IAT effects and reference measures of associations should be spurious. People living in different environments typically differ with regard to self-views, with regard to attitudes, and—because of selective exposure—also with regard to the salience and familiarity of categories. Because associations between category dimensions are typically confounded with asymmetries in salience, positive correlations between the IAT and explicit measures are also expected on the basis of the figure–ground model. In line with the spuriousness assumption, we found a similar pattern of correlations with explicit measures of associations for standard and modified, nonevaluative variants of the IAT (Experiments 2A and 2B). Additionally, structural equation analyses suggested that salience asymmetries are a mediator of these explicit–implicit correlations.[8]

The previous arguments have revealed that the figure–ground model provides an interesting perspective on many research findings involving the IAT. Of course, extrapolating the results of the

[7] The presence of a Black experimenter might also have reminded participants of social norms of not expressing social stereotypes against Black people (Lowery et al., 2001). Participants might therefore have been reluctant to make strategic use of a negative stereotype of Black people in giving their responses (see related arguments below).

[8] Whenever IAT effects, salience asymmetries, and associations are confounded empirically, IAT effects can be used to reliably diagnose associations between categories (or interindividual differences in these associations), even though these associations might not produce IAT effects directly. It is important to note, however, that using the IAT as an instrument to detect associations can yield misleading results because salience asymmetries and associations are not necessarily related.

present experiments to other IATs remains highly speculative, and we do not want to create the impression that all previous findings can be easily explained with the figure–ground model. Nevertheless, the figure–ground model provides an interesting theoretical account of the underlying mechanisms of compatibility effects in the IAT and can be used to explain a wide range of IAT applications. In combination with the visual search task that was developed as an operational criterion of salience asymmetries, the figure–ground model can be used to generate empirically testable hypotheses.

Pitting the Two Accounts Against Each Other

As we have argued, the scope of the figure–ground model is potentially wide, including delicate domains of psychology such as the assessment of prejudice and stereotypes, and it is not restricted to cases of using artificial materials such as nonwords or multicolored strings. However, as is the case with other experimental paradigms, a coexistence of different accounts with different scopes of applicability is conceivable. For example, in negative priming research, a fruitful competition of explanations based on forward inhibition (e.g., Tipper, 1985) versus backward retrieval (e.g., Neill, 1997) has led to demarcatable scopes of applicability (e.g., Tipper, 2001). In semantic priming research, automatic activation has been pitted against expectation-based accounts (see Neely, 1977) or postlexical accounts (for a review, see Neely, 1991). In both domains, competition of explanations has led to a vast amount of studies that have contributed to the understanding of basic processes in attention and memory. Given this backdrop, we concede that there presumably are association-based IAT effects. To demarcate the scopes of applicability of the two accounts, we distinguish between two kinds of IATs or between two ways of conducting an IAT.

In a first type of IAT, the target and attribute categories can easily be mapped onto a single dimension. A trivial example of this type of task was described by Greenwald et al. (1998): "you are shown a series of alternating faces and names, and you are to say 'hello' if the face or name is male and 'goodbye' if the face or name is female" (p. 1464). In this case, the response in each trial can be given on the basis of only one dimension (male vs. female), regardless of the nature of the stimulus (face or name). A similar strategy might also be available in some published IATs, in which the stimuli of the two target categories and also of the two attribute categories are easily distinguishable on the basis of just one dimension, for example, valence (see, e.g., Gray, MacCulloch, Smith, Morris, & Snowden, 2003). Take, for instance, the weapon–musical instrument IAT (Greenwald et al., 1998). With a compatible response assignment (musical instrument–good vs. weapon–bad), an almost perfect discrimination between responses can be achieved on the basis of the valence of the stimuli. Participants might then decide to simplify the complex task and respond only on the basis of the valence of the stimuli. Because such a fast and efficient strategy is not available with an incompatible response assignment (musical instrument–bad vs. weapon–good), a large compatibility effect would be found (the existence of such explicit strategies is suggested by a comparison of task-switching costs in compatible and incompatible blocks of the IAT; see Mierke & Klauer, 2001; Rothermund & Wentura, 2001). In the case of such an explicit valence-based responding, compatibility effects in the IAT must be related to the valence of the categories and stimuli, as most of the people taking such an IAT would agree. In this case, participants might draw on common knowledge or stereotypes to determine the valence of the stimuli of a category. The resulting compatibility effects therefore may reflect "environmental associations" (Karpinski & Hilton, 2001, p. 774) or general knowledge concerning a stereotype (Devine, 1989). Of course, valence is not the only stimulus feature that might be used for a strategic recoding of the double discrimination task in the compatible block. In principle, any feature that helps to distinguish between the two groups of stimuli that are assigned to the different responses can be used for a strategic recoding that simplifies the task (e.g., similarity [De Houwer, 2003], color, upper- vs. lowercase, etc.).[9] Given these problems, it is not surprising that Greenwald and Nosek (2001) did not recommend the use of IATs that "make it easy for respondents to generate alternative construals of the concept's identity" (p. 91; e.g., in terms of valence). However, despite this recommendation, there are published IAT studies that are at least suspicious of strategic components. For example, a recent study by Gray et al. (2003) made use of an IAT with the attribute categories pleasant versus unpleasant (with words like *ugly* written in uppercase) and target categories peaceful versus violent (with words like *kill* written in lowercase). Not surprisingly, the average IAT effect amounted to about 500 ms in this case. Given the arguments above, it can be assumed that a strategic recoding (negative vs. positive) suggests itself for the congruent pairing.[10] In sum, then, whenever the complex classification task in the compatible block of an IAT can be strategically recoded in terms of valence, the resulting IAT effect will be determined by the valence of the categories and stimuli, whereas salience asymmetries between categories presumably play no role in the emergence of the IAT effect.

Much of the appeal and popularity of the IAT, however, stem from variants of this task that do not allow for an easy and obvious recoding of target and attribute categories in terms of valence (or any other single feature). In such a case, the valence of the

[9] It is important to note that we do not want to suggest that salience is intentionally used to simplify the categorization task in the compatible block. Informal interrogations of our participants revealed that they did not develop a strategy of recoding the task in terms of salience. Salience effects in the IAT might be compared with effects of the proportion of compatible trials in the Stroop task: Stroop interference on incompatible trials is positively related to the proportion of compatible trials, indicating that more attention is devoted to the irrelevant feature (Logan & Zbrodoff, 1979). It is highly unlikely, however, that the distribution of attention across relevant and irrelevant features is the result of an intentional strategy. Similarly, in the IAT, a consistent mapping of salient and nonsalient categories onto responses in the compatible block should enhance responding in an automatic fashion. Of course, strategic components might also be involved in mediating the effect (e.g., lowering the response criterion and responding in a spontaneous fashion in the compatible block), but this does not imply that associating salient and nonsalient stimuli with different responses is a conscious strategy.

[10] The main result of the study by Gray et al. (2003) was that psychopathic murderers showed a somewhat smaller IAT effect. Although this is of course an interesting finding, its relevance for exposing potential murderers seems to be severely limited if the result can be related to interindividual differences in the use of strategies (compared with differences in automatic processes).

categories or stimuli might be difficult to detect; alternatively, participants might refrain from acknowledging or actively using the valence of the stimuli and categories for their responses because such a use is prohibited by situational, social, or personal norms (Devine, Monteith, Zuwerink, & Elliot, 1991; Lowery et al., 2001). These subtle applications of the IAT are the most interesting ones because compatibility effects in these variants of the IAT occur without the awareness of an influence of stimulus valence on responding and sometimes even against the explicit intentions of the participants (Banse et al., 2001). Because valence is not a dominant feature in the second type of IAT, however, this type of IAT should not evoke the strong conviction that evaluative associations must necessarily be the basis of the observed effects—otherwise, people taking, for example, the Black–White IAT would not be surprised if the test seemed to indicate a racist attitude, which in fact most of the participants have been (Monteith, Voils, & Ashburn-Nardo, 2001).

The findings reported in the present article are primarily concerned with the second type of IAT. The results of our experiments show that compatibility effects in the second type of IAT are—at least partly—a function of figure–ground asymmetries. Furthermore, our findings suggest that effects of associations on IAT effects are sometimes mediated by salience asymmetries. Of course, we do not want to rule out the possibility that evaluative or other associations might have a strong and unique influence on IAT effects in other cases. Clearly, further research is needed to disentangle the relative contributions of salience asymmetries and associations to IAT effects in other cases. Our research, however, leads to a clear and simple recommendation for IAT researchers who are interested in the underlying processes of IAT effects. Standard IATs should be accompanied by a corresponding word–nonword version of the task (or by any other technical version of the task that makes use of an asymmetrical attribute dichotomy that is clearly not associated with the target categories). Finding comparable results in the two IATs would indicate a strong contribution of salience asymmetries (for a related argument, see Kinoshita & Peek-O'Leary, 2002). Alternatively, researchers might make use of the visual search task to assess salience asymmetries directly and to investigate the relative contributions of figure–ground asymmetries and evaluative associations for a particular IAT. In some cases, it might turn out that salience, as well as evaluative or other associations, contributes to an IAT effect. Experiment 2B might be a prototypical case indicating a joint contribution of explicit evaluation and salience asymmetries in predicting an evaluative IAT effect.

Some readers might have gotten the impression that we want to trivialize results gathered with the IAT. That is not our intent. We are impressed by the robust and replicable effects, and we do think that those effects reveal something important about the human cognitive apparatus. In the last part of the General Discussion, we explicitly argue that figure–ground asymmetries are an important feature of that apparatus.

Theoretical Importance of Figure–Ground Asymmetries

In our view, figure–ground asymmetries reflect a fundamental and universal characteristic of psychological functioning, a partitioning of experience into dichotomies of relevant versus irrelevant, novel–strange versus familiar, unexpected versus expected, rare versus frequent, signal versus noise, and so on (see, e.g., Berlyne, 1958; Ketterer, 1985; Öhman, Hamm & Hugdahl, 2000; Taylor & Fiske, 1978). These dichotomies express a fundamental way of structuring a complex environment that is also observed in animals (e.g., Lubow, Rifkin, & Alek, 1976). It is therefore not surprising that figure–ground asymmetries are frequently found between social categories, for example, in the distinction between in-group and out-group. Members of an out-group tend to stand out against the familiar and homogeneous background of the in-group (e.g., Levin, 2000). A similar allocation of attention emerges for members of stigmatized social groups (Langer, Fiske, Taylor, & Chanowitz, 1976), for schema- or stereotype-inconsistent behavior (Hastie & Kumar, 1979; Verplanken, Jetten, & van Knippenberg, 1996; White & Carlston, 1983), and also for trait categories that are related to socially relevant behavioral tendencies (Wentura, Rothermund, & Bak, 2000). Attentional biases related to figure–ground asymmetries also form the basis of actor–observer differences in perceived causality (Jones & Nisbett, 1987; Storms, 1973). Finally, personality psychologists might be reminded of the theory of personal constructs of Kelly (1955). Kelly proposed that the individual belief system consists of asymmetrical bipolar constructs with one pole being the emergent pole (i.e., the construct actively used by a person), the other the implicit pole (i.e., the implicit counterpart).

At this point, it is important to recall that figure–ground asymmetries must not be equated with asymmetries in evaluations. From our point of view, differences in salience reflect an even more fundamental distinction than differences in evaluation. Figure–ground asymmetries are closely related to the regulation of behavior. Confronting a stimulus of a figure category causes an allocation of attention, controlled processing, an interruption of ongoing processing and behavioral routines, and a reorienting of cognition and action. The capacity of salient information to capture processing and action resources, however, is not tied to a specific valence. For example, both positive and negative other-relevant traits (e.g., generous, aggressive; see Peeters, 1983) automatically attract attention and interrupt current behavioral routines irrespective of their valence (Wentura et al., 2000). In a similar vein, Rothermund, Wentura, and Bak (2001) found automatic vigilance effects of equal magnitude for stimuli signalling chances and dangers. Moreover, even evaluatively neutral information can be highly salient, for example, when it is unexpected or when it is useful for the regulation of behavior in a specific action context (e.g., a signpost).

The previous arguments have pointed out that figure–ground asymmetries constitute an important but somewhat neglected dimension of social cognition research that cannot and should not be reduced to evaluative associations. Although salient categories are often associated with either a positive or a negative evaluation, figure–ground asymmetries between social categories might be even more fundamental than asymmetries in evaluation and need not imply any evaluation whatsoever.

Conclusion

The IAT is a new experimental paradigm that reveals strong, replicable, and nontrivial interference effects. The most straightforward and intuitively appealing account of IAT effects is that these results reflect associations between concepts, best illustrated by links between nodes in a semantic–affective network (see, e.g., Greenwald et al., 2002). As with other established experimental paradigms, the most straightforward account can be questioned on the basis of alternative accounts (without trivializing the basic results). Here, we have proposed an explanation of IAT effects in terms of figure–ground asymmetries that does not rely on associations between categories. We hope that a fruitful competition between accounts regarding the scope of explanation will be initiated by this article.

References

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83,* 380–393.

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48,* 145–160.

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology, 43,* 425–436.

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

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42,* 155–162.

Berlyne, D. E. (1958). The influence of complexity and novelty in visual figures on orienting responses. *Journal of Experimental Psychology, 55,* 289–296.

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

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

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

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

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology, 36,* 316–328.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology, 37,* 443–451.

De Houwer, J. (2002). *The extrinsic affective Simon task.* Unpublished manuscript.

De Houwer, J. (2003, January). *Is the IAT a general measure of similarity?* Paper presented at the Third Workshop on “Implicit Diagnostics” in Bonn, Germany.

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

Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60,* 817–830.

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology, 82,* 835–848.

Florack, A., Scarabis, M., & Bless, H. (2001). When do associations matter? The use of implicit association towards ethnic groups in person judgments. *Journal of Experimental Social Psychology, 37,* 518–524.

Flowers, J. H., & Lohr, D. J. (1985). How does familiarity affect visual search for letter strings? *Perception and Psychophysics, 37,* 557–567.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion, 14,* 61–92.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General, 130,* 681–700.

Gray, N. S., MacCulloch, M. J., Smith, J., Morris, M., & Snowden, R. J. (2003, May 29). Violence viewed by psychopathic murderers: Adapting a revealing test may expose those psychopaths who are most likely to kill. *Nature, 423,* 497–498.

Greenberg, J. H. (1966). *Language universals.* The Hague, the Netherlands: Mouton.

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

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

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

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie, 48,* 85–93.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

Hastie, R., & Kumar, P. A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. *Journal of Personality and Social Psychology, 37,* 25–38.

Hummert, M. L., Garstka, T. A., O’Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging, 17,* 482–495.

Johnston, W. A., & Hawley, K. J. (1994). Perceptual inhibition of expected inputs: The key that opens closed minds. *Psychonomic Bulletin & Review, 1,* 56–72.

Jones, E. E., & Nisbett, R. E. (1987). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones & D. E. Kanouse (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Hillsdale, NJ: Erlbaum.

Kanouse, D. E., & Hanson, L. R. (1972). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 47–62). Morristown, NJ: General Learning Press.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–788.

Kelly, G. A. (1955). *The psychology of personal constructs.* New York: Norton.

Kessler, T., & Mummendey, A. (2001). Is there any scapegoat around? Determinants of intergroup conflicts at different categorization levels. *Journal of Personality and Social Psychology, 81,* 1090–1102.

Ketterer, M. W. (1985). Awareness I: The natural ecology of subjective experience and the mind–brain problem revisited. *Journal of Mind and Behavior, 6,* 469–513.

Kinoshita, S., & Peek-O’Leary. (2002). *Implicit Association Test (IAT):*

Support for the figure–ground asymmetry account. Unpublished manuscript.

Klauer, K. C., Draine, S. C., & Greenwald, A. G. (1998). An unbiased errors-in-variables approach to detecting unconscious cognition. *British Journal of Mathematical and Statistical Psychology, 51,* 253–267.

Krueger, L. E., Stadtlander, L. M., & Blum, A. J. (1992). Search for a singular word or nonword. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1331–1341.

Kühnen, U., Schießl, M., Bauer, N., Paulig, N., Pöhlmann, C., & Schmidthals, K. (2001). How robust is the IAT? Measuring and manipulating implicit attitudes. *Zeitschrift für Experimentelle Psychologie, 48,* 135–144.

Langer, E. J., Fiske, S., Taylor, S. E., & Chanowitz, B. (1976). Stigma, staring, and discomfort: A novel-stimulus hypothesis. *Journal of Experimental Social Psychology, 12,* 451–463.

Levin, D. T. (2000). Race as a visual feature: Using visual search and perceptual discrimination tasks to understand face categories and the cross-race recognition deficit. *Journal of Experimental Psychology: General, 129,* 559–574.

Logan, G. D., & Zbrodoff, N. J. (1979). When it helps to be misled: Facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. *Memory & Cognition, 7,* 166–174.

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

Lubow, R. E., & Kaplan, O. (1997). Visual search as a function of type of prior experience with target and distractor. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 14–24.

Lubow, R. E., Rifkin, B., & Alek, M. (1976). The context effect: The relationship between stimulus preexposure and environmental preexposure determines subsequent learning. *Journal of Experimental Psychology: Animal Behavior Processes, 2,* 38–47.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203.

Maison, D., Greenwald, A. G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin, 32,* 61–69.

Malinowski, P., & Hübner, R. (2001). The effect of familiarity on visual-search performance: Evidence for learned basic features. *Perception and Psychophysics, 63,* 458–463.

Marsh, K. L., Johnson, B. T., & Scott-Sheldon, L. A. J. (2001). Heart versus reason in condom use: Implicit versus explicit attitudinal predictors of sexual behavior. *Zeitschrift für Experimentelle Psychologie, 48,* 161–175.

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

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie, 48,* 107–122.

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

Monteith, M. J., Voils, C. I., & Ashburn-Nardo, L. (2001). Taking a look underground: Detecting, interpreting, and reacting to implicit racial biases. *Social Cognition, 19,* 395–417.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11,* 56–60.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General, 106,* 226–254.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. In D. Besner & G. W. Humphreys (Eds.), *Basic processes in reading: Visual word recognition* (pp. 264–336). Hillsdale, NJ: Erlbaum.

Neill, W. T. (1997). Episodic retrieval in negative priming and repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1291–1305.

Neumann, R. (1999, June). *Erleichterung oder Inhibierung: Wie funktioniert der Implizite Assoziationstest (IAT)?* [Facilitation or inhibition: How does the Implicit Association Test (IAT) work?]. Poster presented at the Seventh Meeting of the Division of Social Psychology, Kassel, Germany.

Neumann, R., Ebert, M., Gabel, B., Gülsdorff, J., Krannich, H., Lauterbach, C., & Wiedl, K. (1998). Vorurteile zwischen Bayern und Norddeutschen: Die Anwendung einer neuen Methode zur Erfassung evaluativer Assoziationen in Vorurteilen [Prejudice between Bavarians and North Germans: Applying a new method for assessing evaluative association within prejudice]. *Zeitschrift für Experimentelle Psychologie, 45,* 99–108.

Neumann, R., & Seibt, B. (2001). The structure of prejudice: Associative strength as a determinant of stereotype endorsement. *European Journal of Social Psychology, 31,* 609–620.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and stereotypes from a demonstration website. *Group Dynamics, 6,* 101–115.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2003). *Understanding and using the Implicit Association Test: II. Methodological issues.* Unpublished manuscript.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130,* 466–478.

Öhman, A., Hamm, A., & Hugdahl, K. (2000). Cognition and the autonomic nervous system: Orienting, anticipation, and conditioning. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 533–575). New York: Cambridge University Press.

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the implicit association test. *Social Cognition, 19,* 97–144.

Otten, S., & Wentura, D. (2001). Self-anchoring and in-group favoritism: An individual profiles analysis. *Journal of Experimental Social Psychology, 37,* 525–532.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Peeters, G. (1983). Relational and informational patterns in social cognition. In W. Doise & S. Moscovici (Eds.), *Current issues in European social psychology* (Vol. 1, pp. 201–237). Cambridge, England: Cambridge University Press.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380–391.

Rothermund, K., & Kaul, M. (2002). [The Extrinsic Affective Simon Task (EAST) as a measure of implicit associations]. Unpublished raw data.

Rothermund, K., & Wentura, D. (2001). Figure–ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie, 48,* 94–106.

Rothermund, K., Wentura, D., & Bak, P. (2001). Automatic attention to stimuli signalling chances and dangers: Moderating effects of positive and negative goal and action contexts. *Cognition and Emotion, 15,* 231–248.

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

Rudman, L. A., Greenwald, A. G., & McGhee, D. E. (2001). Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits. *Personality and Social Psychology Bulletin, 27,* 1164–1178.

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

Schmitt, M., & Maes, J. (2002). Stereotypic ingroup bias as self-defense against relative deprivation: Evidence from a longitudinal study of the German unification process. *European Journal of Social Psychology, 32,* 309–326.

Schneider-Düker, M., & Kohler, A. (1988). Die Erfassung von Geschlechtsrollen: Ergebnisse zur deutschen Neukonstruktion des Bem Sex-Role Inventory [Assessment of gender roles: Results for the new German version of the Bem Sex-Role Inventory]. *Diagnostica, 34,* 256–270.

Shen, J., & Reingold, E. M. (2001). Visual search asymmetry: The influence of stimulus familiarity and low-level features. *Perception and Psychophysics, 63,* 464–475.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.), *Stimulus–response compatibility: An integrated perspective* (Vol. 65, pp. 31–86). Amsterdam, the Netherlands: North-Holland.

Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie, 48,* 123–134.

Storms, M. D. (1973). Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology, 27,* 165–175.

Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 249–288). New York: Academic Press.

Teachman, B. A., Gregg, A. P., & Woody, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology, 110,* 226–235.

Theeuwes, J. (1996). Perceptual selectivity for color and form: On the nature of the interference effect. In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 297–314). Washington, DC: American Psychological Association.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *Quarterly Journal of Experimental Psychology, 37A,* 571–590.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A review and integration of conflicting views. *Quarterly Journal of Experimental Psychology, 54A,* 321–343.

Tipples, J., Young, A. W., Quinlan, P., Broks, P., & Ellis, A. W. (2002). Searching for threat. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A,* 1007–1026.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Verplanken, B., Jetten, J., & van Knippenberg, A. (1996). Effects of stereotypicality and perceived group variability on the use of attitudinal information in impression formation. *Personality and Social Psychology Bulletin, 22,* 960–971.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics, 56,* 495–500.

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

Wentura, D., & Rothermund, K. (2003). The "meddling-in" of affective information: A general model of automatic evaluation effects. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 51–86). Mahwah, NJ: Erlbaum.

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

White, J. D., & Carlston, D. E. (1983). Consequences of schemata for attention, impressions, and recall in complex social interactions. *Journal of Personality and Social Psychology, 45,* 538–549.

Wittenbrink, B., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81,* 815–827.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). Hove, England: Erlbaum.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception and Psychophysics, 63,* 381–389.

Yantis, S. (1996). Attentional capture in vision. In A. F. Kramer, M. G. H. Coles, & G. D. Logan (Eds.), *Converging operations in the study of visual selective attention* (pp. 45–76). Washington, DC: American Psychological Association.

Yantis, S. (1998). Control of visual attention. In H. Pashler (Ed.), *Attention* (pp. 223–256). Hove, England: Erlbaum.

Appendix A

Stimuli Used in the Experiments

Old (Experiments 1A, 1B, 1C, 1E, 3A, 3B)
- GERDA
- HILDE
- MAGDA
- HEDWIG
- WALTRAUD
- OTTO
- JOSEF
- ERICH
- HEINZ
- WALTER

Young (Experiments 1A, 1B, 1C, 1E, 3A, 3B)
- JULIA
- SONJA
- JENNY
- KIRSTEN
- JASMIN
- TOBIAS
- MARIO
- SASCHA
- JENS
- PATRICK

Unpleasant (Experiments 1A, 2B, 3A)
- ABGAS [exhaust fumes]
- VERLUST [loss]
- ANGST [anxiety]
- GEWALT [violence]
- KRIEG [war]
- ELEND [misery]
- SCHMERZ [pain]
- EINSAM [lonely]
- GRAUSAM [cruel]
- GEIZIG [miserly]

Pleasant (Experiments 1A, 2B, 3A)
- PARADIES [paradise]
- HUMOR [humor]
- LIEBE [love]
- SOMMER [summer]
- FRIEDEN [peace]
- FREUND [friend]
- URLAUB [vacation]
- SANFT [gentle]
- TREU [faithful]
- GESUND [healthy]

Nonwords (Experiments 1B, 1C)
- TIMPF
- AGDAT
- RESCHLET
- BALORT
- LESURT
- SCHESOL
- MUKET
- FARST
- NIKAM
- SEDLOR

Neutral words (Experiments 1B, 1C)
- TISCH [table]
- PAPIER [paper]
- FINGER [finger]
- GABEL [fork]
- PUMPE [pump]
- BEGRIFF [concept]
- BECHER [cup]
- FLACH [flat]
- KARIERT [checkered]
- ECKIG [angular]

Colored words (Experiment 3B)
- Fenster [window]
- Topf [pot]
- Holz [wood]
- Beutel [bag]
- Glas [glass]
- Flasche [bottle]
- Sand [sand]
- Jacke [jacket]
- Berg [mountain]
- Stift [pen]

Male (Experiment 2A)
- tapfer [brave]
- mutig [courageous]
- stark [strong]
- athletisch [athletic]
- selbstsicher [self-assured]
- unabhängig [independent]
- tatkräftig [vigorous]
- dynamisch [dynamic]

Female (Experiment 2A)
- einfühlsam [sympathetic]
- fürsorglich [provident]
- zärtlich [tender]
- liebevoll [loving]
- warmherzig [warmhearted]
- romantisch [romantic]
- sanft [gentle]
- hilfsbereit [cooperative]

East (Experiment 2B)
- Jena
- Cottbus
- Grimma
- Stralsund
- Dessau
- Schwerin
- Wismar
- Stendal

West (Experiment 2B)
- Trier
- Koblenz
- Augsburg
- Konstanz
- Mainz
- Bochum
- Aachen
- Mannheim

(*Appendixes continue*)

Self (Experiments 2A, 2B)
- ICH [I]
- SELBST [self]
- MICH [me]
- MEIN [mine]

Other (Experiments 2A, 2B)
- ANDERE [others]
- IHR [you]
- EURE [your]
- EUCH [yours]

Nonwords (Experiments 2A, 2B)
- TIMPF
- AGDAT
- BALORT (Experiment 2A)/SCHESOL (Experiment 2B)
- LESURT
- MUKET
- FARST
- NIKAM
- SEDLOR

Neutral words (Experiments 2A, 2B)
- TISCH [table]
- PAPIER [paper]
- FINGER [finger]
- GABEL [fork]
- PUMPE [pump]
- BECHER [cup]
- STRASSE [street]
- BAUM [tree]

Unpleasant (Experiment 2B)
- Abgas [exhaust fumes]
- Verlust [loss]
- Angst [anxiety]
- Gewalt [violence]
- Schmerz [pain]
- einsam [lonely]
- grausam [cruel]
- geizig [miserly]

Pleasant (Experiment 2B)
- Urlaub [vacation]
- Humor [humor]
- Liebe [love]
- Sommer [summer]
- Frieden [peace]
- sanft [gentle]
- treu [faithful]
- gesund [healthy]

Unpleasant (Experiment 4)
- Abgas [exhaust fumes]
- Verlust [loss]
- Angst [anxiety]
- Gewalt [violence]
- Krieg [war]
- Elend [misery]
- Schmerz [pain]
- einsam [lonely]
- grausam [cruel]
- geizig [miserly]

Pleasant (Experiment 4)
- Paradies [paradise]
- Humor [humor]
- Liebe [love]
- Sommer [summer]
- Frieden [peace]
- Freund [friend]
- Urlaub [vacation]
- sanft [gentle]
- treu [faithful]
- gesund [healthy]

Appendix B

Names of Known and Unknown Persons Used in Experiment 4

Name	Characterization	Evaluation
	Known–negative	
Axel Springer	Publisher of the German scandal sheet *BILD*	−1.92
Dieter Bohlen	Pop singer (nonstudent culture)	−2.06
Edmund Stoiber	German right-wing politician	−1.87
Erich Honecker	Communist leader of the former German Democratic Republic	−1.06
Erich Mielke	Leader of the security service in the former German Democratic Republic	−1.55
Gotthilf Fischer	Leader of a folk music choir (nonstudent culture)	−0.64
Helmut Kohl	Former German chancellor, involved in scandal	−1.63
Jörg Haider	Austrian (extremely) right-wing politician	−2.33
Karl Moik	Moderator of a folk music show on TV (nonstudent culture)	−0.78
Uli Hoeneß	Manager of Munich soccer club (perceived as arrogant)	−1.07
	Known–positive	
Albert Einstein	Physicist	2.31
Erich Kästner	Author of children's books	2.38
Günther Jauch	Talk-show host	1.75
Heinz Rühmann	Actor	1.94
Jan Ullrich	Racing cyclist	1.44
Joschka Fischer	German secretary of state	1.63
Manfred Krug	Actor	1.07
Stefan Raab	Talk-show host	0.94
Ulrich Wickert	Anchorman and host of political talk-shows	1.07
Willy Brandt	Former German chancellor	1.81
	Unknown	
Andreas Seiler		
Fabian Brunstein		
Frank Klauer		
Georg Schmitz		
Jörg Buchtal		
Klaus Wild		
Lothar Peters		
Marco Langner		
Michael Klein		
Rainer Müller		

Received January 7, 2003
Revision received November 26, 2003
Accepted December 16, 2003 ■