Journal of Experimental Social Psychology **37,** 443–451 (2001)
doi:10.1006/jesp.2000.1464, available online at http://www.idealibrary.com on IDEAL®

A Structural and Process Analysis of the Implicit Association Test

Jan De Houwer

University of Southampton, Southampton, United Kingdom

Received May 24, 2000; revised November 6, 2000; accepted November 6, 2000; published online April 3, 2001

The Implicit Association Test (IAT) is based on the observation that participants find it easier to respond in the same way to exemplars of two concepts when these concepts are similar (e.g., "positive" and "flower") compared to when the concepts are dissimilar (e.g., "positive" and "insect"). In the first part of this article, I argue that the IAT is structurally similar to stimulus–response compatibility tasks. On the basis of this analogy, I then present two response conflict accounts of IAT effects. The data of an experiment that was designed to test these accounts showed that IAT effects reflect attitudes toward the target concepts rather than attitudes toward the individual exemplars of those concepts. The results shed light on the processes that underlie IAT effects, suggest that automatic attitude activation may depend on the construal of the object that is fostered by the context, and clarify the relation between different indirect measures of attitudes.

Greenwald, McGhee, and Schwartz (1998) recently introduced the Implicit Association Test (IAT). The simple but ingenious idea behind the IAT is that it should be easier to map two concepts onto a single response when those concepts are somehow similar or associated in memory than when the concepts are unrelated or dissimilar. To test this idea, Greenwald et al. (1998, Experiment 1) presented names of flowers (e.g., TULIP), names of insects (e.g., SPIDER), positive words (e.g, LOVE), and negative words (e.g., UGLY) on a computer screen. Participants were asked to categorize these words by pressing one of two keys. It can be assumed on a priori grounds that the concept "flower" and the concept "positive" are associated in memory, whereas the concept "insect" is associated with the concept "negative." Therefore, when "flower" and "positive" are both assigned to one key and "insect" and "negative" are both assigned to a second key, responses should be fast because the response assignments are compatible with existing associations in memory. When response assignments are incompatible with existing associations (e.g., press left for "insect" and "positive"; press right for "flower" and "negative"), responses should be slower. The results clearly confirmed that reaction times were faster with compatible than with incompatible response assignments.

I first presented the structural analysis of the IAT put forward in this article at a workshop on indirect measures of attitudes that was organized by Tony Greenwald and Marzu Banaji, Chicago, May, 1999. I thank Dirk Wentura, Constantine Sedikides, and Aiden Gregg for their comments on an earlier draft of this article.

Address correspondence and reprint requests to Jan De Houwer, Department of Psychology, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom. E-mail: JanDH@soton.ac.uk.

A STRUCTURAL ANALYSIS OF THE IAT

In this section, I argue that the IAT is structurally similar to stimulus–response compatibility tasks. It has long been know that responses can be emitted more quickly and accurately if the responses are somehow similar to the stimuli to which these responses need to be made than when the responses and stimuli are dissimilar (e.g., Fitts & Seeger, 1953; Kornblum & Lee, 1995). For instance, if participants are asked to press a left key in response to a stimulus presented on the left side of a screen and to press a right key in response to a stimulus presented on the right side of the screen, responses are faster and more accurate than when the response assignments are reversed. In traditional stimulus–response compatibility tasks, the match between the responses and a *relevant* feature of the stimuli varies over trials. The relevant stimulus feature is the feature that participants need to process in order to select the correct response. In the example given above, the spatial position of the stimulus is the relevant feature. With compatible response assignments, the spatial position of a stimulus and the response always match (e.g., press a left key for left stimuli), whereas with incompatible response assignments, the spatial position of a stimulus and the correct response always differ (e.g., press a right key for left stimuli).

0022-1031/01 $35.00

Research on a phenomenon known as the Simon effect has demonstrated that the match between the responses and an *irrelevant* feature of the stimuli can also influence task performance. A feature is said to be irrelevant when participants are not asked and do not need to process this feature in order to respond. For instance, Craft and Simon (1970) presented a red or green stimulus on either the left or right side of the screen (irrelevant feature) and asked participants to press a left or right key on the basis of the color of the stimulus (relevant feature). Results showed that responses were faster and more accurate when the spatial position of the correct response matched the spatial position of the stimulus than when the spatial position of the response and stimulus differed.

Stimulus–response compatibility can involve evaluative aspects of the stimuli and responses. For example, one can present stimuli with a positive or negative valence and ask participants to say the word "GOOD" in response to positive stimuli and "BAD" in response to negative stimuli (compatible combination) or to say "GOOD" to negative words and "BAD" to positive words (incompatible combination). In such a task, stimulus valence is relevant and responses should be faster and more accurate when the valence of the stimulus and the correct response match than when it differs.[1]

Recently, De Houwer and Eelen (1998; De Houwer et al., 2001) demonstrated that the similarity between the valence of the stimulus and the valence of the response also influences reaction time and accuracy when stimulus valence is irrelevant. For instance, De Houwer and Eelen (1998, Experiment 2) presented nouns and adjectives that had a positive (e.g., FLOWER and HAPPY) or negative (e.g., CANCER and UGLY) valence. Half of the participants were asked to say the word "POSITIVE" out loud as quickly as possible when a noun was presented and to respond with "NEGATIVE" when an adjective was presented. For the other participants, the response assignments were reversed (i.e., say "POSITIVE" to adjectives and "NEGATIVE" to nouns). Despite the fact that the valence of the words was irrelevant for the task and had to be ignored, reaction times were shorter when the valence of the presented stimulus and the correct response matched (e.g., say "POSITIVE" to FLOWER) compared to when the valence of the stimulus and the response differed (e.g., say "NEGATIVE" to FLOWER).

In traditional stimulus–response compatibility tasks, similarity depends on long-term associations that have developed as the result of past experiences (Proctor & Lu, in press). For instance, the word FLOWER is similar to the response "POSITIVE" because both are associated in memory with the representation of positive valence. Although the responses that are used in IAT studies are typically unrelated to valence (e.g., press a left or right key), they are mapped onto positive or negative concepts through task instructions. As a result, temporary or short-term associations (Proctor & Lu, in press; Zorzi & Umilta, 1995) are created between the representations of the responses on the one hand and the representations of positive and negative valence on the other hand. Assuming that one defines similarity not only in terms of long-term associations but also in terms of short-term associations, one can argue that the similarity between stimuli and responses in an IAT task varies as the result of task instructions.

Consider the IAT as implemented by Greenwald et al. (1998, Experiment 1). When participants are asked to press a left key for positive words and flower names (compatible response assignments), a short-term association will be established between the representation of the response "press the left key" and the representation of positive valence because the concepts that are mapped onto the left response both have a positive valence. Therefore, with compatible response assignments, flower names will be similar to the correct response. When participants are instructed to press the right key for negative words and flower names (incompatible response assignments), the right response is mapped onto one concept with a negative valence (i.e., the concept "negative") and one concept with a positive valence (i.e., the concept "flower"). Therefore, the representation of the response "press the right key" will become associated both with negative valence and with positive valence. As a result, flower names will be less similar to the correct response when the response assignments are compatible than when they are incompatible.

According to this analysis, the IAT and stimulus–response compatibility tasks have in common that the similarity between the presented stimulus and the to-be-emitted response varies between different trials. The main difference between the two types of tasks is that in stimulus–response compatibility tasks, the match between the responses and stimuli varies as a function of their long-term associations, whereas in the IAT it depends on the short-term associations of the responses and the long-term associations of the stimuli. Rather than having responses that have long-term associations with positive or negative valence (such as saying the word "GOOD" or "BAD"), in the IAT, responses have short-term associations with the representations of positive or negative valence as the result of task instructions.

A PROCESS ANALYSIS OF THE IAT

The fact that the IAT is structurally similar to stimulus–response compatibility tasks has implications for theories about the processes that underlie the IAT. Current models of stimulus–response compatibility effects postulate that these

[1] In fact, this prediction has been confirmed (Anthony Greenwald, May, 1999, personal communication).

effects occur at the stage of response selection (e.g., Hommel, 1997; Kornblum & Lee, 1995; Zhang, Zhang, & Kornblum, 1999). Without going into too much detail, these accounts postulate that the representations of responses that are similar to the relevant or irrelevant feature of the presented stimulus become activated automatically upon presentation of the stimulus. For instance, when a stimulus is presented on the left side, this will automatically activate the representations of responses that are associated with a left spatial position (e.g., pressing a left key). When the automatically activated response representation differs from the representation of the correct response (e.g., press a right key in response to a stimulus presented on the left side of a screen), response selection will be delayed relative to a situation in which the automatically activated response representation is the representation of the correct response. On the basis of the structural similarity between the IAT and stimulus–response compatibility tasks, one could speculate that IAT effects are also due to processes that occur at the response selection stage. I now discuss two possible response selection accounts of IAT effects.

Relevant Feature Account

When participants classify flower names, insect names, positive words, and negative words by pressing a left or right key (Greenwald et al., 1998, Experiment 1), the semantic category of the words is the relevant stimulus feature. These categories differ with regard to their valence: The concepts "flower" and "positive" have a positive valence; the concepts "insect" and "negative" have a negative valence. As was explained above, one can argue that the representations of the responses "press the left key" and "press the right key" become associated with the representations of positive and negative valence because they are mapped onto concepts that have a positive or negative valence. According to the *relevant feature account,* the extent to which a response representation will be activated automatically depends on the overlap between the long-term associations of the specific instantiation of the relevant feature and the short-term associations of the response.

For instance, because TULIP is the name of a flower and the concept "flower" has a positive valence, TULIP will automatically activate response representations that are associated (either in long-term or short-term) with positive valence. When flower names and positive words are assigned to the left key and insect names and negative words are assigned to the right key (i.e., compatible response assignments), the response "press the left key" will be associated with positive valence and the response "press the right key" will be associated with negative valence. In this case, the word TULIP will activate the representation of the left response automatically because the relevant feature of that stimulus (i.e., it belongs to the category "flower") has a positive valence and the left response is associated with positive valence. Because the automatically activated response representation is the representation of the correct response, response selection is facilitated and responses will be fast and accurate. Likewise, a word such as CANCER belongs to the category "negative" that has a negative valence. It will therefore automatically activate the right response because this response is associated with negative valence as the result of task instructions. This will facilitate response selection and performance.

When participants are asked to press the left key for positive words and insects and the right key for negative words and flowers (i.e., incompatible response assignments), each response representation will be associated both with positive and with negative valence. As a result, stimuli will (a) automatically active the representation of the incorrect response and/or (b) automatically activate the representation of the correct response to a lesser extent than with compatible response assignments. Both factors will delay response selection with incompatible relative to compatible response assignments, thus resulting in poorer performance with incompatible than with compatible response assignments.

Irrelevant Feature Account

At this point, I need to make a distinction between target concepts and attribute concepts. The IAT examines whether target concepts possess (i.e., are associated with) a certain attribute. When the IAT is used to measure attitudes, the attribute concepts always correspond to the concepts "positive" and "negative" and the target concepts are those concepts whose valence is being measured (Greenwald et al., 1998, p. 1465). In the example I have used (Greenwald et al., 1998, Experiment 1), the concepts "flower" and "insect" are the target concepts. Target concept trials are trials on which an exemplar of one of the target concepts is presented (e.g., a flower or insect name); attribute concept trials are trials on which an exemplar of one of the attribute concepts is presented (e.g., a positive or negative word).

On target concept trials, one can make a distinction between a relevant and irrelevant feature of the stimuli. For instance, flower names and insect names do not only differ with regard to their semantic category but also with regard to their individual valence. Because participants do not need to process the valence of the individual flower and insect names and because they are instructed to respond to these names on the basis of the semantic category to which the names belong, one can argue that stimulus valence is an irrelevant feature of the target concept stimuli. According to the irrelevant feature account, the extent to which a response representation is automatically activated depends on the overlap between the long-term associations of the presented stimulus and the short-term associations of the response representation. For instance, when participants are asked to press the left key for flower names and positive words and

the right key for insect names and negative words, TULIP will activate the representation of the left response because TULIP has a positive valence and the left key is associated with positive concepts. In other words, automatic response activation will not depend on the valence of the category to which the stimulus belongs (e.g., "flower") but on the valence of the stimulus itself.

It is important to note that in a typical IAT, there is a perfect confounding between the relevant and irrelevant feature of target concept stimuli. That is, all exemplars of one target concept are positive (e.g., flower names), whereas all exemplars of the other target concept are negative (e.g., insect names). As a result, it is unclear whether poorer performance with incompatible than with compatible response assignments reflects response conflicts as induced by the valence of the individual exemplars of the target concepts or the valence of the target concepts as such. According to the relevant feature account, the IAT effect reflects how positive or negative the different target concepts are. According to the irrelevant feature account, however, IAT effects reflect the difference between the mean valence of the exemplars of the first target concept and the mean valence of the exemplars of the second target concept.

However, one should note that the distinction between the irrelevant and relevant feature account only holds for target concept trials. On attribute concept trials, participants do need to process the valence of the individual stimuli in order to select the correct response. Therefore, an attribute concept stimulus can only activate response representations because it is positive or negative, that is, because it is an exemplar of the concept "positive" or "negative" (relevant feature account). A target concept stimulus, on the other hand, can activate response representations either because the stimulus itself is positive or negative (irrelevant feature account) or because it is an exemplar of a positive or negative target concept (relevant feature account). It is interesting to note that until now, there are no published data about the effect of compatibility on target and attribute concept trials separately. It is thus unclear whether the overall compatibility effects that have been reported until now occurred mainly on the target concept trials, the attribute concept trials, or to the same extent on both types of trials.

AN EMPIRICAL TEST OF THE TWO ACCOUNTS

In order to test these accounts of the IAT, I conducted an experiment that was modeled after the experiments reported by Greenwald et al. (1998). The main difference was that the valence of the individual target concept stimuli varied within each target category. As target concepts, I used "British" and "Foreign." Assuming that people generally have a favorable attitude toward concepts that apply to themselves (e.g., Farnham, Greenwald, & Banaji, 1999; Nuttin, 1985), one can infer that the British students who participated in this experiment would have a more positive attitude toward the concept "British" than toward the concept "foreign." However, even though such an in-group bias might be quite strong, most people do not have a negative attitude toward all foreign individuals or a positive attitude toward all individuals of their own nationality. This allowed me to select both positive and negative exemplars of the two target concepts.

On the basis of the irrelevant feature account, one can predict that performance on target concept trials will be superior when both the presented stimulus and the correct response are associated with the same valence. With compatible response assignments (press left for British names and positive words; press right for foreign names and negative words), the left response is associated only with positive valence, whereas the right response is associated only with negative valence. According to the irrelevant feature account, positive British names and positive foreign names will automatically activate the representation of the left response, whereas negative British and negative foreign names will activate the representation of the right response. As such, positive British names and negative foreign names automatically activate the representation of the correct response, whereas negative British and positive foreign names activate the representation of the incorrect response. Therefore, responses to the positive British and negative foreign names should be faster and more accurate than responses to negative British and positive foreign names. Such an impact of the valence of the individual target stimuli should not occur when the response assignments are incompatible (press left for foreign names and positive words; press right for British names and negative words). With incompatible assignments, each response representation is associated with both positive and negative valence and will thus be activated by both positive and negative exemplars of the categories British and Foreign. This will lead to a response conflict on each trial and thus to overall inferior performance regardless of stimulus valence.

According to the relevant feature account, automatic activation of response representations is determined by the valence of the target concepts rather than by the valence of the individual exemplars. Therefore, the valence of the target category exemplars should have no effect on performance even when the response assignments are compatible. The relevant feature account only predicts better performance with compatible than with incompatible response assignments. Because the target concept "British" has a positive valence, whereas the target concept "foreign" has a negative valence, both positive and negative British names will induce a tendency to give the response that is associated with positive valence, whereas both positive and negative foreign names will induce a tendency to give the response that is associated with negative valence. With compatible

response assignments, the left response is associated only with positive valence and the right response only with negative valence. Therefore, the automatically activated response representation will correspond to the representation of the correct response on every single trial. This will lead to fast and accurate responses. With incompatible response assignments, however, each response is associated with both positive and negative valence and all stimuli will thus activate both response representations. This will lead to a response conflict on each trial and thus to slow and error-prone performance.

Method

Participants. Twenty-eight female sixth-form students who visited the University of Southampton during an information day volunteered to take part. All were British subjects. They were tested in one group of 17 and one group of 11 students.

Materials. Target concept stimuli were the names of three liked British persons, three disliked British persons, three liked foreign persons, and three disliked foreign persons (see the Appendix). When presenting these names, the first word of each name was presented as an initial. The names were selected after consulting the results of recent popularity polls and after informal discussions with British members of the Department of Psychology at the University of Southampton. As attribute concept stimuli, I used six positive and six negative adjectives (see the Appendix). Before each phase and during each phase, the name of the target and/or attribute concept that was assigned to the left key was printed in the top left corner of the screen, whereas the name of the target and/or attribute concept that was assigned to the right key was written in the top right corner of the screen. All words were written in white letters presented on a black background. A letter was 7 mm high and 5 mm wide. Presentations were controlled by a Turbo Pascal 5.0 program that operated in graphics mode. The program was implemented on IBM-compatible computers that were situated in one room that contained 21 of such computers. Each participant was seated in front of one of the computers at a distance of approximately 40 cm from the computer screen. Participants could respond by pressing the key "q" or the key "p" of the (QWERTY) keyboard. The time between the presentation of a word and the first key press was measured using a highly accurate (beyond 1 ms) Turbo Pascal Timer (Bovens & Brysbaert, 1990).

Procedure. After filling in an informed consent form, participants were first given written instructions on the computer screen. These instructions informed participants that names of British and foreign persons would be presented on the computer screen together with positive and negative words. Their task was to classify these stimuli by pressing one of two keys. The assignment of responses to categories was said to vary from phase to phase. The categories assigned to the left key would be shown in the top left corner of the screen, whereas the categories assigned to the right key would be shown in the top right corner of the screen. Participants were also told that, if they made a correct response, the next word would appear on the screen immediately. If, however, the response was incorrect, they would hear a beep and the word would stay on the screen until they made the correct response. Finally, participants were informed that the experiment would last about 20 min and were asked to respond as quickly but also as accurately as possible. After allowing participants a few minutes to read these instructions, the experimenter held up a paper on which all the names were printed in a random order. The experimenter read these names out loud and briefly reminded the participants of who these persons were by giving a short description (see the Appendix). The experimenter also reminded the participants about which persons were British and which were foreign.

The task itself consisted of five phases. During the first phase, all 6 British names and 6 foreign names were presented 4 times, twice during a first block of 24 trials and twice during a second block of 24 trials. During the 24 trials of the second phase, each of the 6 positive and 6 negative words was presented twice. In the third phase, all 24 stimuli (12 names and 12 adjectives) were presented twice during a first block of 48 trials and twice during a second block of 48 trials. The fourth and fifth phase were identical to the first and third phase respectively, except with regard to the response assignments for British and foreign names. All blocks were separated by a self-terminated pause during which the labels for the next block and their allocation to the responses were presented on the screen. Information about whether the next block would be a practice or test block also appeared on the screen. All blocks except the second block of Phases 3 and 5 were described as practice blocks. The order in which the different stimuli were presented was randomized for each phase, block, and participant separately with the following restrictions. First, the same stimulus could not be presented on two or more consecutive trials. Second, the correct response could not be the same on more than four consecutive trials.

Regardless of the phase, all participants were asked to press the left key for positive words and the right key for negative words. Half of the participants were asked to press the left key for British names and the right key for foreign names during Phases 1 and 3, but to press the left key for foreign names and the right key for British names during Phases 4 and 5 (Ordercondition 1). The other participants pressed left for foreign and right for British names during Phases 1 and 3 and left for British and right for foreign names during Phases 4 and 5 (Ordercondition 2).

On each trial, a word was presented until the participant gave the correct response. If the participant made an incorrect response, a tone of 200 Hz was presented for 250 ms

TABLE 1

Mean Untransformed Reaction Times (in Milliseconds) and Percentage of Errors (*SD* in Parentheses) on Target Concept Trials as a Function of Combination, Target Concept, and Stimulus Valence

Stimulus valence	Target category: British	Target category: Foreign
Compatible combination		
Positive		
Reaction time	741 (216)	713 (204)
Percentage of errors	.07 (.08)	.09 (.12)
Negative		
Reaction time	730 (203)	776 (253)
Percentage of errors	.08 (.12)	.08 (.13)
Incompatible combination		
Positive		
Reaction time	856 (199)	857 (251)
Percentage of errors	.12 (.15)	.05 (.10)
Negative		
Reaction time	903 (281)	833 (220)
Percentage of errors	.07 (.11)	.04 (.09)

while the word remained on the screen. The next trial was initiated 400 ms after the participant entered the correct response. At the end of the experiment, participants were asked to rate their liking of (-100 = *dislike very much* and $+100$ = *like very much*) and familiarity with (0 = *totally unfamiliar* and $+100$ = *very familiar*) each of the British and foreign persons whose name was presented during the experiment.

Results

I only took into account the time and accuracy of the first response on the test trials, that is, the trials in the second block of Phases 3 and 5. In accordance with Greenwald et al. (1998), the first trial of each block was discarded, as were reaction times on trials where the response was incorrect (7.75% of all trials). Reaction times below 300 ms or above 3000 ms were recoded to 300 ms and 3000 ms respectively (0.4% of all correct responses). Finally, all latencies were log-transformed.

Liking and familiarity ratings. The Target Concept (British or Foreign) × Stimulus Valence (positive or negative) ANOVA of the liking ratings showed that positive target concept names indeed received a higher liking rating than negative target concept names [$F(1, 27) = 412.29$, $MS_e = 912.62$, $p < .001$.] *t*-Tests confirmed that the effect of stimulus valence was present for both British names [$Mpositive = 57.32$, $SD = 31.58$, $Mnegative = -53.75$, $SD = 17.07$, $t(27) = 15.34$, $p < .001$] and foreign names [$Mpositive = 45.12$, $SD = 19.84$, $Mnegative = -75.65$, $SD = 22.14$, $t(27) = 20.03$, $p < .001$]. The ANOVA also revealed a main effect of target concept [$F(1, 27) = 15.43$, $MSe = 527.85$, $p = .001$] showing that overall, British names were liked more than foreign names.

The Target Concept × Stimulus Valence ANOVA that was performed on the familiarity ratings revealed a main effect of target concept [$F(1, 27) = 4.41$, $MSe = 195.93$, $p = .045$], a main effect of stimulus valence [$F(1, 27) = 9.38$, $MSe = 121.81$, $p = .005$], and an interaction between target concept and stimulus valence [$F(1, 27) = 30.99$, $MSe = 158.74$, $p < .001$]. *t*-Tests showed that positive British names ($M = 76.55$, $SD = 22.71$) were more familiar than negative British names ($M = 56.90$, $SD = 23.58$), $t(27) = 6.42$, $p < .001$, whereas the positive foreign names ($M = 57.74$, $SD = 21.50$) were somewhat less familiar than negative foreign names ($M = 64.61$, $SD = 22.12$), $t(27) = -2.10$, $p = .045$.

Target concept trials. The Ordercondition (compatible or incompatible combination first) × Combination (compatible or incompatible) × Target Concept (British or foreign) × Stimulus Valence (positive or negative) ANOVA revealed that the three-way interaction between the last three variables was not significant, $F < 1$. Table 1 shows the means that are involved in this interaction. Contrary to what was predicted by the irrelevant feature account, responses to positive British and negative Foreign names were not faster than responses to negative British and positive Foreign names, neither when the response assignments were compatible, $t < 1$, nor when the response assignments were incompatible, $t < 1$.

There was, however, a clear main effect of combination, [$F(1, 26) = 25.80$, $MS_e = 0.047$, $p < .001$]. Reaction times were significantly shorter with compatible than with incompatible combinations (see Table 2). The only other effect that approached significance was the interaction between ordercondition and combination, [$F(1, 26) = 3.64$, $MS_e = 0.047$, $p = .068$], which indicated that the effect of combination tended to be stronger when the compatible

TABLE 2

Mean Untransformed Reaction Times (in Milliseconds) and Percentage of Errors (*SD* in Parentheses) on Target Concept and Attribute Concept Trials as a Function Combination of Response Assignments

Trial type	Combination: Compatible	Combination: Incompatible
Target concept trials		
Reaction time	740 (143)	862 (152)
Percentage of errors	.08 (.07)	.07 (.07)
Attribute concept trials		
Reaction time	724 (140)	882 (204)
Percentage of errors	.05 (.04)	.10 (.07)

combination came first (Ordercondition 1) compared to when it came second (Ordercondition 2) (also see Greenwald et al., 1998, Fig. 2). It is likely that this result occurred because participants had difficulties with switching from one combination to another. As a result, there was a general advantage for the combination that came first (i.e., the combination presented during Phase 3) which added to the main effect of combination when the compatible combination came first but counteracted the effect of combination when the incompatible combination came first.

An ANOVA performed on the percentage of errors showed that the crucial three-way interaction between combination, target concept, and stimulus valence was not significant,[2] [$F(1, 26) = 2.17$, $MS_e = 0.180$, $p = .15$]. Regardless of compatibility, accuracy was the same on trials with positive British or negative Foreign names than on trials with negative British or positive Foreign names, ts < 1. The main effect of combination was not significant, $F < 1$, suggesting that accuracy was the same with compatible than with incompatible response assignments (see Table 2). The analysis of the error data did reveal a number of less interesting interactions [combination $\times$ target concept, $F(1, 26) = 5.32$, $MS_e = 0.010$, $p = .029$; ordercondition $\times$ stimulus valence, $F(1, 26) = 6.81$, $MS_e = 0.0066$, $p = .015$; four-way interaction, $F(1, 26) = 6.48$, $MS_e = 0.180$, $p = .017$] that were unrelated to the main hypothesis and which I therefore do not discuss further.

Attribute concept trials. An Ordercondition $\times$ Combination $\times$ Attribute Concept ANOVA of the reaction time data revealed a significant main effect of combination [$F(1, 26) = 35.50$, $MS_e = 0.027$, $p < .001$], resulting from the fact that reaction times on attribute concept trials were significantly shorter with compatible than with incompatible response assignments (Table 2). The only other significant effect was the interaction between ordercondition and attribute concept [$F(1, 26) = 8.34$, $MS_e = 0.013$, $p = .008$]. This effect indicated that in Ordercondition 1, responses to positive words ($M = 797$, $SD = 184$) were faster than responses to negative words ($M = 872$, $SD = 216$), whereas in Ordercondition 2, responses to positive words ($M = 806$, $SD = 138$) were slower than responses to negative words ($M = 735$, $SD = 97$).

A similar analysis of the error data revealed a significant main effect of combination [$F(1, 26) = 15.40$, $MS_e = 0.0052$, $p = .001$], showing that participants made less errors with compatible than with incompatible response assignments (Table 2). There were also three marginally significant effects of lesser importance that I do not discuss further [ordercondition $\times$ combination, $F(1, 26) = 3.85$, $MS_e = 0.0052$, $p = .061$; ordercondition $\times$ attribute concept, $F(1, 26) = 3.25$, $MS_e = 0.0074$, $p = .083$; main effect of ordercondition, $F(1, 26) = 3.37$, $MS_e = 0.0074$, $p = .078$].

DISCUSSION

In this article, I presented a structural analysis of the IAT that led to the formulation of two new accounts of IAT effects. An experiment that was designed to test these accounts showed that the valence of the individual target concept stimuli had little or no impact on performance. Rather, only the valence of the target concepts (i.e., "British" or "Foreign") mattered. The results thus suggest that IAT effects reflect differences between the valence of the target concepts rather than differences between the valence of the exemplars of both concepts. The reported data are also the first to demonstrate that the compatibility of the response assignments has an effect on both target concept trials and attribute concept trials. As explained above, these results question the validity of the irrelevant feature account but support the relevant feature account: Once a stimulus is categorized, response representations that are associated with the same valence as the target concept will be activated automatically. When response assignments are compatible, only the to-be-emitted response is activated automatically in this way. When response assignments are incompatible, however, the incorrect response will also be activated, which will interfere with the selection of the correct response.

However, the conclusion that the relevant feature account provides the best explanation of IAT effects needs to be qualified. Mierke and Klauer (in press) recently pointed out that IAT effects could also be due to the fact that participants use different strategies with compatible than with incompatible response assignments. With compatible response assignments, it is irrelevant whether target concept stimuli are treated as target concept stimuli or as attribute concept stimuli. For instance, classifying a flower name as a flower will lead to the same response as classifying it as a positive word when participants need to press the same key for flower names and positive words. Therefore, participants could perform the task without having to switch between the

[2] The test blocks in Phases 3 and 5 were both preceded by one practice block. The data from the practice blocks were not included in the main analysis because they were meant to offer participants with an opportunity to practice the combined classification task and because participants were informed that the practice blocks were intended for practice only (also see Greenwald et al., 1998; Rudman, Greenwald, Mellott, & Schwartz, 1999). Including the data of the practice blocks in the main analysis lead to only one important difference. The ANOVA of the error data of both blocks of Phases 3 and 5 did reveal a significant three-way interaction between combination, target concept, and stimulus valence [$F(1, 26) = 6.63$, $MS_e = 0.0028$, $p = .016$]. As predicted by the irrelevant feature account, less errors were made in response to positive British and negative Foreign names than to negative British and positive Foreign names when the response assignments were compatible [5 and 8% of errors respectively, [$t(27) = -2.64$, $p = .016$] but not when response assignments were incompatible (7% in both cases, $t < 1$). Adding the reaction time data of the practice trials to the main analysis of the reaction time data, however, had no effect on the crucial three-way interaction, $F < 1$.

task of classifying names and the task of classifying words. This would lead to faster performance with compatible than with incompatible response assignments. In the present experiments, participants could not adopt a more simple strategy with compatible than with incompatible response assignments because different exemplars of the same target concept had a different valence. It is possible, however, that when stimulus and target concept valence are confounded (as was the case in previous IAT tasks), (some) participants do adopt different strategies with compatible than with incompatible response assignments. Therefore, the present results do not exclude the possibility that strategic factors underlie IAT effects when stimulus and target concept valence are confounded. One can only conclude that IAT effects can occur in the absence of such strategic factors and the relevant feature account provides the best explanation for the effects that occur under these conditions.

The present results do not only provide an insight into the processes that produce IAT effects, they also have important implications for our understanding automatic attitude activation. Stimuli in our environment always consist of several features or elements. Sometimes we have conflicting attitudes toward the features of a single stimulus. Imagine seeing a good friend who displays a negative facial expression. Our affective reaction toward this stimulus could reflect our positive attitude toward the person we are seeing, the negative attitude toward the facial expression that the person is displaying, or both. The present results suggest that affective reactions will be mainly guided by the attitude toward the feature that is most salient or relevant to us within the context where we encounter the attitude object (see also Macrae, Bodenhausen, & Milne, 1995). The data showed that the name of a British person automatically induced a tendency to give a response that was associated with positive valence, regardless of whether the person was a popular comedian or a convicted mass murderer. Likewise, foreign names induced a tendency to give a response that was associated with negative valence, regardless of whether it was the name was of a liked person (such as Einstein) or a disliked person (such as Hitler).[3] In other words, the attitude toward the nationality of the person dominated the attitude toward all other characteristics of the person (e.g., is the person a comedian or a mass murderer, someone how advanced science or caused the holocaust). The fact that patriotic attitudes dominated responses makes sense only if one considers that the nationality of the presented name was relevant for the task in which participants were engaged in.[4]

This discussion also has implications for the relation between different indirect measures of attitudes. Recent studies showed that when the IAT is used to measure attitudes, the results do not appear to converge with the results of other indirect measures of the same attitudes (Cameron, Alvarez, & Bargh, 2000). The present data suggest that IAT effects are induced by and thus reflect the valence of the relevant but not irrelevant features of the presented stimuli. There are good reasons to assume that other indirect measures, such as the affective priming task (Fazio et al., 1986), might be more sensitive to the global attitude toward a stimulus rather than the attitude toward one (relevant) feature of that stimulus (De Houwer, in press). Because of this difference, it is possible that IAT and other indirect measures sometimes diverge.

For instance, a prime stimulus such as the name "Gandhi" would most likely facilitate responses to positive compared to negative targets in a priming task. In the IAT as implemented in the present experiment, however, the same stimulus activated negative rather than positive responses. The data suggest that the British–Foreign IAT measured the difference between the valence of the target concept "British" and the target concept "foreign" rather than the difference between the mean valence of the individual British names and the mean valence of the presented foreign names. It is likely that if one would measure the valence of the concepts "British" and "foreign" using an affective priming task, the difference between the mean affective priming scores for the British names and the mean affective priming score for the foreign names would reflect the difference between the mean valence of the individual British and foreign names rather than the difference between the va-

[3] An anonymous reviewer suggested that affective reactions toward British and foreign names might have differed because the foreign names were orthographically less familiar than the British names for our British participants. However, this hypothesis is at odds with the observation that the effect of compatibility on reaction times had the same direction and magnitude for the foreign name "B. Pitt" (135 ms) as for the other foreign names (M = 104 ms) despite the fact that "Pitt" is a common British name.

[4] One could argue that the present task was for some reason insensitive to the effects of the valence of individual stimuli. However, I also conducted a second experiment in which the same task was used but the target concepts were neutral (i.e., "person" and "animal"; De Houwer, 2000a). In this experiment, responses were on average 81 ms faster and 11% more accurate when the stimulus and response were associated with the same valence than when they were associated with a different valence. For instance, responses to the word FRIEND were faster and more accurate than responses to the word LIAR when person names and positive words were assigned to the same key but the reverse was true when person names and negative words were assigned to the same key. These effects of stimulus valence occurred even when target concept and attribute concept stimuli were written in different colors so that they could easily be discriminated. It thus seems to be the case that the valence of the target concepts only dominates the valence of the individual stimuli when the valence of the target concepts clearly differs (as is the case in most IATs; e.g., "British" versus "foreign") but not when they have a more or less similar valence (e.g., "person" versus "animal"). More generally, this suggests that IAT-like effects can be due to several processes and that the nature of the task (e.g., differential valence of target concepts and consistency of the valence of different exemplars of the same category) can determine which processes actually produce the effects.

lence of the concepts "British" and "foreign." This example illustrates that the results of the IAT and the affective priming task need not necessarily converge.

APPENDIX

Target Concept and Attribute Concept Stimuli

Positive British names
Princess Diana (recently deceased Princess of Wales), Lenny Henry (popular British comedian), Queen Mother (mother of the British Queen)
Negative British names
Margaret Thatcher (former British prime minister), Rosemary West (convicted mass murderer), Donald Shipman (mass murderer convicted just before the experiment took place)
Positive foreign names
Albert Einstein (well know scientist of German descent), Mahatma Ghandi (former Indian leader), Brad Pitt (popular American actor)
Negative foreign names
Adolf Hitler (fascist leader of Nazi Germany), President Pinochet (former Chilean dictator who was placed under house arrest in Britain at the time that the experiment was conducted), Hoessein Saddam (Iraqi leader)
Positive adjectives
pure, sincere, funny, polite, good, happy
Negative adjectives
hideous, aggressive, mean, brutal, bad, ugly, angry

REFERENCES

Bovens, N., & Brysbaert, M. (1990). IBM PC/XT/AT and PS/2 Turbo Pascal timing with extended resolution. *Behavior Research Methods, Instruments, and Computers,* **22,** 332–334.

Cameron, J. A., Alvarez, J. M., & Bargh, J. A. (2000). *Examining the validity of implicit measures of prejudice.* Poster presented at the first annual meeting for the Society of Personality and Social Psychology, Nashville, TN.

Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display: Interference from an irrelevant directional cue. *Journal of Experimental Psychology,* **83,** 415–420.

De Houwer, J. (2000a). Stimulus–response compatibility effects without dimensional overlap. Manuscript in preparation.

De Houwer, J. (in press). A structural analysis of indirect measures of attitudes. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion.* Mahwah, NJ: Lawrence Erlbaum.

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

De Houwer, J., & Eelen, P. (1998). An affective variant of the Simon paradigm. *Cognition and Emotion,* **12,** 45–61.

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

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

Fitts, P. M., & Seeger, C. M. (1953). SR compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology,* **46,** 199–210.

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.

Hommel, B. (1997). Toward an action-concept model of stimulus–response compatibility. In B. Hommel & W. Prinz (Eds.), *Theoretical issues in stimulus–response compatibility* (pp. 281–320). Amsterdam: North-Holland.

Kornblum, S., & Lee, J.-W. (1995). Stimulus–Response compatibility with relevant and irrelevant stimulus dimensions that do and do not overlap with the response. *Journal of Experimental Psychology: Human Perception and Performance,* **21,** 855–875.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology,* **69,** 397–407.

Mierke, J., & Klauer, K. C. (in press). Implicit association measurement with the IAT: Evidence for an effect of supervisory processes. *Zeitschrift fur Experimentelle Psychologie.*

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

Proctor, R. W., & Lu, K.-P. L. (in press). Eliminating, magnifying, and reversing spatial compatibility effects with mixed location-relevant and irrelevant trials. In W. Prinz & B. Hommel (Eds), *Attention and Performance XIX.*

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.

Zhang, H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus–stimulus and stimulus–response compatibility. *Cognitive Psychology,* **38,** 386–432.

Zorzi, M., & Umilta, C. (1995). A computational model of the Simon effect. *Psychological Research,* **58,** 193–205.