The Implicit Association Test as a General Measure of Similarity

Jan De Houwer, Tanja Geldof, and Els De Bruycker
Ghent University

Abstract The Implicit Association Test (IAT) is widely used as a measure of semantic similarity (i.e., associations in semantic memory). The results of previous research and of a new study show that IAT effects can, however, also be based on other types of similarity between stimuli. We therefore put forward the hypothesis that the IAT provides a general measure of similarity. Given that similarity is highly dynamic and context-dependent, our view that the IAT measures similarity is compatible with existing evidence showing that IAT effects are highly malleable. We provide further evidence for this in a new study in which the outcome of an IAT depended on whether the perceptual or functional characteristics of the stimuli were made salient.

Within a few years time, the Implicit Association Test (IAT) has become a widely used tool to study attitudes and other associations in memory. This task, which was introduced by Greenwald, McGhee, and Schwartz (1998), is based on the idea that it should be easier to respond to stimuli that represent a particular concept when the same response is also assigned to exemplars of a second associated concept than when the same response is also required for exemplars of a second unassociated concept. For instance, participants might be asked to press one of two keys on the basis of whether the presented word is the name of a flower, the name of an insect, a positive adjective, or a negative adjective. Results show that performance is better when flower names and positive words are assigned to one key and insect names and negative words are assigned to the second key (compatible task) than when pressing the first key is correct for insect names and positive words and pressing the second key is correct for flower names and negative words (incompatible task) (e.g., Greenwald et al., 1998, Experiment 1). This pattern of results is in line with the a priori assumption that, in semantic memory, the target concept "flower" is associated more strongly with the attribute concept "positive" than with the attribute concept "negative" whereas the reverse is true for the target concept "insect."

Results such as these suggest that the IAT can be used as a tool to measure and study associations in semantic memory. As an illustration, we will briefly explain how the IAT has been used to measure self-esteem (defined as the relative strength of the associations between the target concept "self" and the attribute concepts "positive" and "negative," see Greenwald et al., 2002). In a self-esteem IAT, words are presented that represent the concept "self" (e.g, ME), the concept "not-self" or "other" (e.g., THEM), the concept "positive" (e.g., FRIENDLY), and the concept "negative" (e.g., HOSTILE). Results typically show that performance is better in a task where self and positive items are assigned to one key and not-self and negative items are assigned to a second key than in a task where the assignments are reversed (i.e., first key for self and negative items; second key for not-self and negative items; e.g., Greenwald & Farnham, 2000). Such a pattern of results is in line with the idea that people on average possess positive self-esteem. Interindividual differences in the self-esteem IAT effect have also been interpreted as a reliable and valid indicator of actual differences in (implicit) self-esteem (e.g., Greenwald & Farnham, 2000).

Given the potential practical benefits of the IAT, it is important to have a good theoretical understanding of the task. Theoretical research on the IAT has been directed toward two questions: the *what-question* and the *how-question*. The first question is about what the IAT measures. That is, what aspects of the stimuli or knowledge about the stimuli can determine the direction and magnitude of IAT effects? As can be inferred from the above, the most widely accepted view is that IAT effects reflect associations between concepts in semantic memory (e.g., Greenwald & Nosek, 2001, p. 85). Once one has an idea of what information determines IAT effects, there is still the question of how this information produces IAT effects. That is, why is it that certain information results in faster performance in one classification task than in another classification task? This *how-question* is thus about the processes that are responsible for the IAT effect. Research directed at this question tries to unravel the exact mechanisms by which IAT effects are produced (e.g., Brendl, Markman, & Messner, 2001; De Houwer, 2001; Mierke & Klauer, 2003). Of course, both questions are related. Knowledge about the information that is reflected in IAT effects constrains models of the mechanisms that produce those effects. Likewise, knowledge about the mechanisms can provide insight into the type of information that can activate those mechanisms and thus produce IAT effects. Nevertheless, both questions are to a large extent separable. For instance, a particular type of information could lead to IAT effects as the result of a number of different mechanisms.

The present paper focuses mainly on the *what-question*. As mentioned earlier, most researchers assume that IAT effects are based on associations in semantic memory. We will refer to this hypothesis as the semantic association account. Rothermund and Wentura (2001, 2004), however, argued that most IAT effects are not driven by semantic associations but by asymmetries in the salience of stimuli and concepts. This answer to the *what-question* is called the salience account. Although it is doubtful that IAT effects are always due to salience asymmetries, it is clear that they at least sometimes are. In one of their experiments (Rothermund & Wentura, 2004, Experiment 3b), participants saw first names typical of elderly persons, first names typical of young persons, green words that were neutral regarding valence and age, and yellow neutral words. They were asked to press one of two keys on the basis of the age-implication of the names (old or young) and the colour of the neutral words (green or yellow). When all old and young names were written in yellow ink, participants were faster when old names and green words were assigned to the same key than when old names and yellow words were assigned to the same key. However, when all old and young names were written in green ink, the reverse was found. This result can be explained if one assumes that performance on IAT tasks will improve when salient concepts are assigned to the same key. In the condition where all names were yellow, green words were uncommon and thus salient. Old names were also salient because they are less familiar than young names for the (young) participants. Therefore, performance should be better when old names and green words are assigned to the same key. When all names are printed in green, yellow words are salient. Therefore, performance should be better when old names and yellow words are assigned to the same key, as was observed.

Whereas the role of semantic associations and salience has already been acknowledged in the literature on the IAT, little attention has been given to isolated findings that suggest that IAT effects can also be based on perceptual similarity. Mierke and Klauer (2003, Experiment 1a) presented small red objects, big blue objects, small objects that were neither red nor blue, and big objects that were neither red nor blue. Participants were asked to respond on the basis of the colour of the red and blue objects and on the basis of the size of the objects that were not red or blue. Because all red objects were small and blue objects big, red objects were perceptually similar to small objects, and blue objects were perceptually similar to the big objects. Results showed that performance was better when red and small objects were assigned to one key and blue and big objects to the other key than when the response assignments were reversed. That is, performance was better when perceptually similar items were assigned to the same key than when they were assigned to different keys.

We recently came across another paper that was published long before the IAT was introduced, but that also supports the view that IAT effects can be based on perceptual similarity between stimuli. This paper by Lasaga and Garner (1983) contains a number of studies on the oblique effect, that is, the observation that people have greater difficulty processing obliquely aligned (i.e., diagonal) lines than horizontal or vertical lines. Lasaga and Garner argued that the oblique effect was due to the fact that a left and right diagonal line are perceptually similar. To test this idea, they used a classification task that was highly similar to an IAT. In one task, participants were asked to give one response upon presentation of a left or right diagonal line and a second response when a horizontal or vertical line was presented. In other tasks, one of the diagonal lines was assigned to the same response as the horizontal line and the other diagonal line was assigned to the same response as the vertical line. Importantly, responses were much faster in the first task than in the other tasks. Because there are no reasons to assume that simple stimuli such as lines are associated or even represented in semantic memory, it seems unlikely this IAT-like effect was based on associations in semantic memory. A more plausible explanation is that the effect was due to the perceptual similarity between a left and right diagonal line. This is also the account that Lasaga and Garner put forward:

> In classification tasks, the relative magnitudes of overall task RT will reveal possible facilitative stimulus groupings. If the four stimuli are grouped into two groups of two such

> that the groups contain stimuli that are highly similar to each other but different from those in the other group, then classification should be facilitated, since intraclass similarity has been maximized and interclass similarity has been minimized. (p. 217)

The ideas and results of Lasaga and Garner (1983), together with the results of Mierke and Klauer (2003), led us to a new answer to the *what-question*: The IAT measures similarity. Importantly, because similarity can be defined with respect to various intrinsic and structural features of stimuli and concepts (e.g., Kornblum & Lee, 1995; Medin, Goldstone, & Gentner, 1993), this answer encompasses previous answers to the *what-question*. First, associations in semantic memory can be regarded as a particular algorithmic representation of semantic similarity, that is, as one of several possible ways to try to implement similarity with regard to meaning in a theoretical process model (e.g., Medin et al., 1993). For instance, in semantic network models, concepts that are thought to be similar in meaning have a strong direct association between them or a strong indirect connection as the result of many overlapping associations with other concepts (e.g., Collins & Quinlan, 1972; Glaser & Glaser, 1989). Therefore, a similarity view of the IAT encompasses the semantic association account, that is, the traditional idea that IAT effects reflect associations in semantic memory. Second, stimuli and concepts can also be similar or different with regard to their salience rather than their semantic meaning (which Kornblum & Lee, 1995, p. 872, regard as a form of structural similarity). Therefore, our similarity view encompasses the salience account because similarity with regard to salience is seen as a potential source of IAT effects.

The similarity account also provides added value. First, it is the only account from which it follows that IAT effects can be based on perceptual similarity. Although the existing evidence for perceptually based IAT effects is still limited, the initial results that we described above do suggest that such effects can be found (i.e., Lasaga & Garner, 1983; Mierke & Klauer, 2003). Second, the similarity account makes apparent the links between IAT research and cognitive research on similarity, and therefore leads to interesting new ideas and questions. Most important, it is well known that similarity is highly dynamic and context dependent (but still constrained and thus theoretically useful, see Medin et al., 1993). As Medin et al. pointed out in their highly influential paper, "context tends to activate or make salient context-related properties, and, to the extent that examples being judged share values of these activated properties, their similarity is increased" (p. 257). Similarity should therefore not be regarded as being a fixed or given entity but as the outcome of a constructive process that is guided by the context. For instance, the colour black and the colour white are perceptually highly dissimilar but they are similar with respect to the fact that they are both monochrome colours. The extent to which black is seen as similar to white depends on the extent to which the context makes this shared feature salient. For instance, a black and a white object are regarded as more similar in the presence of a red object than in the absence of such an object (Medin et al., 1993, p. 268).

On the basis of the assumption the IAT measures similarity and the assumption that similarity is highly dynamic, one can easily understand the data showing that IAT effects are context-dependent (see Blair, 2002, for a review). For instance, Mitchell, Nosek, and Banaji (2003, Experiment 1) presented names of liked black athletes, disliked white politicians, positive words, and negative words. When participants were asked to respond to the names on the basis of occupation (athlete or politician) and to words on the basis of valence, the IAT revealed a more positive attitude toward black athletes than toward white politicians. When the race of the names was relevant (black or white), however, a more positive attitude toward white politicians was found. Likewise, Dasgupta and Greenwald (2001) found smaller race IAT effects (i.e., less prejudice against black people as measured by the IAT) when participants saw names of admired Black Americans and disliked White Americans before completing the race IAT. Such findings can be understood on the basis of the similarity view because context manipulations such as manipulations of the relevant feature in an IAT (e.g., occupation or race; Mitchell et al., 2003) or of the items to which one is pre-exposed (e.g., Dasgupta & Greenwald, 2001) can make salient certain features of stimuli and concepts. This can influence the extent to which those stimuli and concepts are viewed as being similar to positive or negative concepts. Research has also demonstrated that properties that allow one to group concepts or stimuli become salient and will thus determine similarity (e.g., Tversky, 1977). This could explain why the IAT effect for one particular category of target stimuli (e.g., insects) can depend heavily on the nature of the other target category (e.g., flowers or nonwords; Brendl et al., 2001) if one makes the likely assumption that replacing a category can sometimes alter the properties by which categories are grouped.

Before we describe other implications of our similarity view, we will present the data of two experiments that were designed to further examine a) the conditions under which perceptual similarity can influence IAT performance (Experiment 1, 2) and b) the context-dependency of IAT effects (Experiment 2).

Experiment 1

Gathering evidence for the impact of perceptual similarity on IAT effects is important because it cannot be explained on the basis of previous answers to the *what-question* (i.e., semantic association and salience accounts; Greenwald & Nosek, 2001; Rothermund & Wentura, 2001, 2004) but does fit well with the similarity view. Further research is warranted because the existing evidence is limited in several respects. First, so far there have been only two demonstrations of perceptually based IAT effects (Lasaga & Garner, 1983; Mierke & Klauer, 2003). Second, these studies involved simple, abstract stimuli that differ substantially from the stimuli that are typically used in IAT studies (i.e., words or photographs). Third, they involved very strong manipulations of perceptual similarity given that a) all items of one class of stimuli (e.g., the left diagonal line; the small red objects) were perceptually more similar to all items of one particular other class of stimuli (e.g., the right diagonal line; the small objects) but not to the other stimuli (e.g., the vertical and horizontal line; the large objects) and b) the manipulation of similiarity involved a task-relevant feature (orientation lines; size of objects). It thus remains to be seen whether more subtle forms of perceptual similarity with other types of stimuli can also influence IAT performance.

In our first experiment, participants saw colour photographs of pizzas, coins, snakes, and rivers and were asked to press one of two keys on the basis of the category to which the presented photograph belonged. In daily life, most pizzas and coins are round whereas most snakes and rivers are winding. However, we selected both photographs of round pizzas and coins and photographs of pizzas and coins that were not round (e.g., a triangular slice of pizza and a rectangular coin). Likewise, half of the selected photographs of snakes and rivers depicted winding snakes and rivers whereas on the other photographs, the snakes and rivers were not winding (e.g., a coiled snake and a view across a wide river from a river bank). During the crucial test blocks, participants were asked to press one key for coins and pizzas and the second key for rivers and snakes (coin-pizza task) or to press the first key for coins and snakes and the second key for rivers and pizzas (coin-snake task). Therefore, in the coin-pizza task but not in the coin-snake task, perceptually similar items (round coins and pizzas; winding snakes and rivers) were assigned to the same key. Based on the hypothesis that perceptual similarity between items can influence IAT performance, we expected faster responses to these items in the coin-pizza task than in the coin-snake task. To establish that this effect was indeed due to the perceptual similarity between the items, we also examined the effect of task for the items that were not similar (e.g., slice of pizza, rectangular coin, coiled snake, view of wide river).

Method

Participants. Twenty-eight first-year psychology students at Ghent University took part in exchange for course credit.

Stimuli and materials. We selected four photographs of a coin, four photographs of a pizza, four photographs of a snake, and four photographs of a river. For each category, there were two similarity items and two control items. The similarity items were two photographs of a round coin, two photographs of a round pizza, two photographs of a winding snake, and two photographs of a winding river (taken from an airplane or mountain). The control items were a photograph of a square coin, a photograph of a rectangular coin, two photographs of a triangular slice of pizza, two photographs of nonwinding rivers (taken from a riverbank or a bridge), and two photographs of curled-up snakes. On the 14" screen, the photographs were 9.5 cm wide and 7.2 cm high. Before and during each phase, the name of the categories that were assigned to the left key were printed in the top left corner of the screen and the name of the categories that were assigned to the right key were printed in the top right corner of the screen. As labels, we used the Dutch words MUNTSTUK (coin), RIVIER (river), SLANG (snake), and PIZZA (pizza). The labels were written in black capital letters within a white bar of 1.3 cm high and between 3.7 and 6.0 cm wide. The background colour of the screen was black throughout the experiment. The experiment was programmed in Inquisit and was executed on a Pentium 166 computer with a 14" screen. Participants could respond by pressing the key "q" (on the left side) or the key "m" (on the right side) of the AZERTY keyboard.

Procedure. Participants took part one by one and were seated at approximately 50 cm from the screen. After filling in an informed consent form, they were given written instructions on the computer screen. These informed them that photographs of a coin, a pizza, a snake, or a river would appear one by one on the screen. Their task was to respond to the identity of the depicted object (coin, pizza, snake, or river) by pressing one of two keys. Which types of photographs were assigned to which key would vary from phase to phase. Before and during each phase, the categories assigned to the left response would be presented in the top left corner of the screen and the categories assigned to the right key would appear in the top right corner of the screen. Participants were asked to

TABLE 1
Mean Untransformed Reaction Times, Percentage of Errors, and Within-Subjects Standard Errors as a Function of Item Type and Task in Experiment 1

	Task				IAT effect	
	Coin-Pizza		Coin-River			
Item Type	*M*	*SE*	*M*	*SE*	*M*	SE
Similarity Item						
Reaction Time	613	14	654	18	41*	19
Percentage of Errors	3.77	0.69	4.03	0.77	0.47	0.65
Control Item						
Reaction Time	633	17	644	19	10	19
Percentage of Errors	4.23	0.63	3.06	0.59	-0.96	0.91

**p* < .05.

respond as quickly but also as accurately as possible. Finally, they learned that the task would take about 10 minutes.

The experiment consisted of five phases. During the first phase, all coin and river photographs were presented three times in a block of 24 trials. On the 24 trials of the second phase, all pizza and snake photographs were presented three times each. During the third phase, all 16 photographs were presented three times in a first block of 48 trials and three times in a second block of 48 trials. The fourth and fifth phase were identical to the second and third phase, respectively, except for the response assignments of the pizzas and snakes. Half of the participants were instructed to press the left key for coins and the right key for rivers throughout the experiment. The other participants always pressed the left key for rivers and the right key for coins. Within each group, half of the participants were asked to press the left key for pizzas and the right key for snakes during the second and third phase and the left key for snakes and the right key for pizzas during the fourth and fifth phase. The other participants pressed the left key for snakes and the right key for pizzas during Phase 2 and 3 but pressed the left key for pizzas and the right key for snakes during Phases 4 and 5. As a result, whether the coin-pizza task (press one key for coins and pizzas and the other key for rivers and snakes) was performed before or after the coin-snake task (press one key for coins and snakes and the other key for rivers and pizzas) was also counterbalanced across participants.

Before each block, participants were informed about the categories and response assignments of the upcoming block. Information about whether the upcoming block would be a practice or test block was also provided. All blocks except for the two blocks of the third and fifth phase were said to be practice phases. Photographs were presented in a random order that was determined separately for each block and each participant. On each trial, the photograph was presented at the screen centre and remained there until a response was given. If the response was incorrect, the photograph was replaced by a red X of approximately 1 cm high and 1 cm wide that remained on the screen for 400 ms. The next photograph was presented 400 ms after the removal of the photograph or the red X.

Results

We took into account only the trials of the test blocks of Phases 3 and 5, discarding the first trial of each block as well as reaction times on trials where the response was incorrect. In accordance with Greenwald et al. (1998), reaction times shorter than 300 ms or longer than 3,000 ms were recoded to 300 ms and 3,000 ms, respectively. All latencies were then log-transformed. We conducted an Item Type (similarity or control) × Task (coin-pizza or coin-snake) repeated measures ANOVA on the observed mean log-transformed reaction times. A priori t-tests were used to examine the effect of task for similarity and control items separately. One could argue that analyzing similarity and control trials separately violates the basic principle that one should consider all trials within a task of the IAT together (e.g., Greenwald, Nosek, & Banaji, 2003). Although it might be true that IAT effects for subgroups of items do not provide a valid measure of associations in memory (e.g., because the IAT effect for a particular class of items is known to depend heavily on the nature of the other items, see Brendl et al., 2001), this does not exclude the fact that analyses on subgroups of items can provide a useful tool to increase our theoretical understanding of the IAT. The aim of the present study was not to obtain a valid measure of a particular association in memory. Rather, it was directed at uncovering whether perceptual similarity between items can influence performance in an IAT task. Analyzing the data for similarity items and control items separately is an appropriate manner to achieve this theoretical aim.

All relevant untransformed means can be found in Table 1. The ANOVA did not reveal a main effect of item type, $F < 1$, or task, $F(1, 27) = 2.05$, but the interaction between both variables was significant, $F(1, 27) = 7.66$, $p = .01$. A priori t-tests showed that reaction times were faster in the coin-pizza task than in the coin-snake task on trials where a similarity item was presented, $t(27) = 2.17$, $p = .04$, $d = .41$, but not on trials where a control item was presented, $t < 1$, $d = .11$.

Discussion

When an item was presented that was perceptually similar to items of another category (i.e., similarity items; e.g., a round coin that was similar to round pizzas), participants were faster when that other category was assigned to the same response as the category of the presented item (e.g., press a left key for coins and pizzas) than when that other category was assigned to another response (e.g., press left for coins and right for pizzas). Such a result was not found for category items that were not similar to items of other categories (i.e., control items). These data thus demonstrate that IAT performance can be influenced by the perceptual similarity between items, a result that cannot be explained on the basis of a semantic association or salience account. Unlike what was the case in previous studies (Lasaga & Garner, 1983; Mierke & Klauer, 2003), the crucial perceptual feature was not mentioned in the instructions. Moreover, only half of the items were similar and colour photographs of real objects were used as stimuli. Our results therefore show that perceptual similarity can have an impact on IAT performance even when the similarity is quite subtle.

Experiment 2

As we pointed out above, the same concepts or stimuli can be regarded as similar or dissimilar depending on the features that are made salient. For instance, most pizzas are similar to most coins in that they are round. However, pizzas and coins are dissimilar from a functional point of view in that pizzas but not coins are eatable. If IAT performance is driven by similarity, one can thus predict that it should depend on what features are made salient by the context. It is likely that the perceptual features of the stimuli were salient in Experiment 1 because of the high degree of perceptual similarity of the similarity items. But on the basis of the similarity view, one can predict that performance will be different when functional features are made salient than when perceputal features are salient.

We therefore ran a second experiment similar to Experiment 1 but added a phase during which participants were asked to judge whether pizzas, coins, snakes, and rivers are eatable (functional condition) or whether these objects are round or winding (perceptual condition). Afterwards, participants performed an IAT consisting of one task in which functionally similar items were assigned to the same key (coins and rivers vs. pizzas and snakes; coin-river task) and another task in which perceptually similar items were assigned to the same key (coins and pizzas vs. snakes and rivers; coin-pizza task). As in Experiment 1, only half of the pizzas and coins were round and only half of the snakes and rivers were winding (similarity items). We predicted that performance would be better in the coin-river task than in the coin-pizza task when participants first judged whether the items were eatable. This effect should be the same for similarity items than for control items because the functional similarity applies to both kinds of items. In contrast, when participants first judged the form of the items, we expected results similar to those of Experiment 1: Better performance in the coin-pizza task than in the coin-river task but only for similarity items because the perceptual similarity applies only to those items.

Method

Participants. Fifty-one first year psychology students at Ghent University took part in exchange for course credits. Twenty-four students were assigned to the perceptual condition, the other twenty-seven students were assigned to the functional condition.

Stimuli and materials. The photographs that were presented during the IAT were identical to those used in Experiment 1, except for three snake photographs. In Experiment 1, two snake photographs were of drawings of snakes and a third one was of a sculpture of a snake. Because some participants might argue that drawn or sculptured snakes are not eatable, we replaced these photographs with identically formatted photographs of live snakes. As in Experiment 1, half of the four snake photographs showed a winding snake whereas the other two photographs showed a coiled snake. All other materials were the same as in Experiment 1.

Procedure. The experiment started with a phase during which participants were asked to judge a perceptual or functional feature of pizzas, coins, snakes, and rivers. In the functional condition, they were asked whether pizzas, snakes, coins, and rivers (in that order) are eatable. In the perceptual condition, they judged whether pizzas, coins, snakes, and rivers (in that order) are round or winding. Each question was written in

TABLE 2
Mean Untransformed Reaction Times, Percentage of Errors, and Within-Subjects Standard Errors as a Function of Item Type, Task, and Condition in Experiment 2

	Task				IAT effect	
	Coin-Pizza		Coin-River			
Item Type	*M*	*SE*	*M*	*SE*	*M*	SE
			Perceptual Condition			
Similarity Item						
Reaction Time	557	15	601	18	44*	20
Percentage of Errors	4.60	0.69	5.50	0.76	0.90	0.86
Control Item						
Reaction Time	579	14	588	18	9	15
Percentage of Errors	4.18	0.74	6.03	1.00	1.85	1.09
			Functional Condition			
Similarity Item						
Reaction Time	585	18	553	15	-32#	22
Percentage of Errors	5.26	0.65	5.77	0.85	0.51	0.96
Control Item						
Reaction Time	586	15	551	13	-35#	19
Percentage of Errors	4.90	0.81	4.29	0.60	-0.61	0.80

*$p < .05$; #$p < .10$.

white Arial letters, font size 16, on a black background and appeared in the centre of the screen. Participants responded by pressing the key "1" (eatable, round) or the key "2" (not eatable, winding). The question disappeared after the participants pressed a key. The next question was presented after an interval of 400 ms. After answering all four questions, participants in the functional condition saw the message "Pizzas and snakes are indeed eatable whereas coins and rivers are not" and participants in the perceptual condition saw the message "Pizzas and coins are indeed most often round whereas snakes and rivers are most often winding." This message stayed on the screen during 10 s, after which the instructions for the IAT were presented. The procedure of the IAT was identical to that of Experiment 1 except that the coin and snake (rather than the coin and river) photographs were presented during the first phase and the pizza and river (rather than the pizza and snake) photographs appeared in Phases 2 and 4. This change allowed us to create one task in which functionally similar items were assigned to the same key (coins and rivers vs. pizzas and snakes; coin-river task) and another task in which perceptually similar items were assigned to the same key (coins and pizzas vs. snakes and rivers; coin-pizza task).

Results

Means were calculated in the same way as in Experiment 1 and were analyzed using a Condition (perceptual or functional) × Item Type (similarity or control) × Task (coin-pizza or coin-snake) ANOVA with repeated measures on the last two variables. Untransformed mean reaction times can be found in Table 2. Most importantly, the interaction between condition and task was significant, $F(1, 49) = 6.76$, $p = .01$, showing that the IAT effect depended on whether participants first judged perceptual or functional features of the concepts. The interaction between condition and item type approached significance, $F(1, 49) = 3.49$, $p = .07$, as did the three-way interaction between condition, item type, and task, $F(1, 49) = 2.59$, $p = .11$, all other Fs < 1.18.

We also conducted two Item Type × Task ANOVAs in order to test our predictions regarding the effects in each condition separately. As expected and in line with the results of Experiment 1, the ANOVA of the data of the perceptual condition yielded a significant interaction between item type and task, $F(1, 23) = 4.85$, $p = .04$. A priori t-tests showed that responses to similarity items were faster in the coin-pizza task than in the coin-snake task, $t(23) = 2.26$, $p = .03$, $d = 0.46$, whereas responses to control items were not influenced by task, $t < 1$, $d = 0.16$. A similar analysis of the data of the functional condition revealed only a marginally significant main effect of task, $F(1, 26) = 3.92$, $p = .06$, but no interaction between item type and task, $F < 1$, or main effect of item type, $F < 1$. A priori t-tests revealed that responses tended to be faster in the coin-river task than in the coin-pizza task for both similarity, $t(26) =$

1.76, p = .09, d = 0.34, and control items, t(26) = 2.06, p = .05, d = 0.40.

Discussion

The most important finding was that the direction of the IAT effect depended on the nature of the feature that was made salient. When participants were asked to judge whether pizzas, coins, snakes, and rivers are typically round or winding (perceptual condition), performance was better when perceptually similar items were assigned to the same key (coin-pizza task) than when these items were assigned to different keys (coin-snake task). As in Experiment 1, this effect was significant only for perceptually similar items. When participants first judged whether pizzas, coins, snakes, and rivers are eatable, a clearly different pattern of results emerged: Participants tended to be faster when functionally similar (i.e., eatable) items were assigned to the same key (coin-snake task) than when functionally similar items were assigned to different keys (coin-pizza task). Moreover, this effect tended to be present for both similarity and control items.

Some caution is needed when interpreting the effects in the functional condition because they did not reach conventional levels of significance. However, whether the IAT effect in the functional condition is reliable is not crucial for the aims that we had. What is important is that a) the interaction between condition and task was clearly significant and b) the results in the perceptual condition replicated those of Experiment 1. The interaction is crucial because it demonstrates for the first time that IAT performance can depend on the type of similarity (perceptual or functional) that is made salient by the context. Because the semantic association and salience account focus on only one type of similarity, it is difficult to see how this result can be explained by those accounts. The interaction fits very well, however, with the idea that IAT effects can be based on various types of similarity and that the context can determine the type of similarity that dominates. The significant IAT effect that emerged for the perceptually similar items in the perceptual condition is also important because it is compatible with our similarity account but not with the semantic association or salience accounts.

One could argue that in both experiments, there might have been a confound between item type (similarity or control item) and typicality. For instance, it is possible that a winding snake is a more typical depiction of a snake than a coiled snake. Hence, the different results for the similarity and control items might have been due to differences in typicality. There are two arguments against this hypothesis. First, whereas prototypical items are normally categorized more quickly than nonprototypical items, we did not find a difference between the time to categorize similarity items and control items. Second, it would be difficult to explain why the results for similarity and control items differed only in Experiment 1 and the perceptual condition of Experiment 2, but not in the functional condition of Experiment 2.

General Discussion

In this paper, we put forward the hypothesis that the IAT provides a general measure of similarity. This implies that IAT performance can reflect not only similarity with regard to meaning or salience, but also with regard to other features. Previous research already suggested that IAT effects can be based on similarity with regard to perceptual properties of stimuli (e.g., Lasaga & Garner, 1983; Mierke & Klauer, 2003). We extended this research by demonstrating that IAT performance can depend on perceptual similarity even when only some items of the concepts are similar (i.e., photographs of round pizzas and round coins; photographs of winding rivers and winding snakes; Experiment 1). We also examined for the first time whether the test context can determine the type of similarity on which IAT performance is based. The results of Experiment 2 showed that IAT performance reflected similarity with regard to form (i.e., round or winding) when this perceptual feature was made salient during a training phase but that it reflected similarity with regard to a functional characteristic (i.e., eatable or uneatable) when this feature was made salient during the training phase.

The ideas and data that are presented in this paper have a number of theoretical implications. Most important, our similarity view provides a new perspective on the question of *what* the IAT measures. The proposal that it can measure all kinds of similarity encompasses previous proposals. It should not be contrasted with the idea that the IAT measures associations in semantic memory (e.g., Banaji, 2001; Greenwald et al., 1998), nor with the view that the IAT picks up regularities in salience (e.g., Rothermund & Wentura, 2001, 2004). Associations in semantic memory can be seen as one particular algorithmic implementation of similarity with regard to meaning (Medin et al., 1993). Likewise, stimuli can be similar or different with regard to their salience also (e.g., Kornblum & Lee, 1995; Medin et al., 1993). The similarity account leads to the insight that those previous proposals are not mutually exclusive but can in fact be regarded as two instances of the general principle that the IAT measures similarity. It is also the only perspective that is compatible with the fact that IAT effects can be based on perceptual similarity. Moreover, a similarity view leads to the prediction that it should be possible to find IAT effects that are based on yet other types of similarity, for instance, with regard to relational features (e.g., an atom and the earth are similar in that they both revolve around something; see Medin et al., 1993, p. 257).

Because there are so many features and dimensions on which stimuli can be similar or different, it is incorrect to regard similarity as a fixed entity. Rather, the extent to which someone will regard two stimuli as similar with depend on whether the features that they share are salient for that particular person in that particular context. Similarity (and thus IAT performance) depends on a constructive process that is determined by both characteristics of the individual and context. The view that IAT effects reflect similarity therefore provides a natural account of the fact that IAT effects are highly context-dependent (see Blair, 2002, for a review). In addition, a great deal is known about the dynamics of similarity (e.g., Medin et al., 1993). This knowledge can be used to increase our understanding of the IAT and related measures such as the Go-Nogo Association Test (GNAT; Nosek & Banaji, 2001) and Extrinsic Affective Simon Task (EAST; De Houwer, 2003a). For instance, research has shown that as children become older, they pay more attention to abstract conceptual properties and less to superficial properties (e.g., Gentner, 1988). This suggests that the impact of perceptual similarity on IAT effects will decrease with age.

The similarity view is also compatible with the fact that, structurally, the IAT has a lot in common with stimulus-response compatibility tasks (as do many other indirect measures of attitudes; see De Houwer, 2001, 2003b). In theories about stimulus-response compatibility effects, similarity plays a crucial role (e.g., Kornblum & Lee, 1995). That is, stimulus-response compatibility effects are thought to arise because of perceptual, structural, or semantic similarity between stimuli and responses or between stimuli. In hindsight, it thus comes as no surprise that IAT effects can be based on different forms of similarity.

Finally, the hypothesis that the IAT is a general measure of similarity leads to interesting new questions. An important question for future research is how the effects of various forms of similarity interact. One possibility is that all effects are additive. Another possibility is that the effects interact in such a way that the effect of one type of similarity can reduce or eliminate the effect of another type of similarity. For instance, it is possible that when stimuli are semantically similar, similarity at a perceptual level has little or no effect. Our results do not inform us about this issue because we did not manipulate two kinds of similarity within the same task. More research is thus needed to address this important question.

Although we believe that the similarity view has clear scientific merits, one could argue that it is also limited in certain ways. First, one could argue that it is built on the ill-specified concept of "similarity." Although we acknowledge that there are problems associated with this concept (as there are with other concepts such as "association"), it is important to realize that it has proven to be an extremely useful explanatory concept in research on a huge number of topics such as memory retrieval (Hintzmann, 1986), categorization (e.g., Nosofsky, 1986), learning (e.g., Ross, 1984), and social judgment (Smith & Zarate, 1992) to name just a few. Moreover, researchers such as Medin et al. (1993) have provided a coherent and inspiring framework for understanding the nature of similarity.

Second, the similarity view offers a perspective on the question of *what* the IAT measures but does not entail an account of *how* IAT effects come about. That is, it does not specify the processes by which similarity between items can produce IAT effects. Nevertheless, one can think of ways in which similarity could influence IAT performance. For instance, when similar concepts are assigned to the same key, this could result in an unambiguous acquired meaning of the responses and thus in less stimulus-response conflicts than when dissimilar concepts are assigned to the same key (see De Houwer, 2001, 2003b, for such a stimulus-response compatibility account of the IAT). Another possible mechanism is similarity-based retrieval of stimulus-response episodes. Research suggests that the presentation of a stimulus leads to the activation of recent memory traces that include similar stimuli (e.g., Hommel, 1998). As a result, the responses that are part of the activated memory traces are also activated. This will influence response selection and thus performance. Such similarity-based retrieval of responses would facilitate performance when similar stimuli are assigned to the same key but would hamper performance when similar stimuli are assigned to different keys.

Third, one might argue that, unlike a proper scientific theory, the similarity view is difficult to refute empirically. Because most objects and concepts are similar in some respects, one can always interpret a particular IAT effect as reflecting some kind of similarity. Although it is indeed true that one can interpret any IAT effect in terms of similarity, such interpretations are, however, testable. For instance, if one argues that an IAT effect is due to a certain similarity between the stimuli, one should be able to detect this similarity using other tasks that have been used to measure similarity (e.g., a same-different categorization task; see Lasaga & Garner, 1983) or be able to alter the IAT effect by altering the salience of a particular type of similarity. Moreover, as we pointed out above, adopting a similarity view of the IAT is scientifically productive in that it leads to new insights and predictions.

Rather than challenging the scientific merits of the similarity view, one could also raise doubts about the novelty of this account. Greenwald, Nosek, Banaji, and Klauer (2005) point out that philosophers such as Aristotle and Hume regarded similarity as a basis for the formation of an association. One could thus argue that the term "association" in a broad sense encompasses the term "similarity" and that a similarity view is not incompatible with the view that the IAT measures associations. We agree that there is little difference between a similarity view and an associative view if that two objects or concepts are associated whenever they are similar in a certain respect. However, we found no indication that the term "association," as used in the literature on the IAT, has been understood in this broad sense. Rather, there are strong indications that the term "association" was used in the same way as in semantic network models (e.g., Collins & Quinlan, 1972), namely, as referring to a relatively stable connection between two conceptual representations in semantic memory, a connection that gradually builds up on the basis of successive experiences. For instance, in their seminal paper on the IAT, Greenwald et al. (1998) used the IAT to measure attitudes and stereotypes, concepts that Greenwald et al. (2002) subsequently conceived of as connections in a semantic network. Likewise, in their review paper about the IAT, Greenwald and Nosek (2001, p. 85) defined the IAT as "a method for indirectly measuring strengths of associations between *concepts*" (italics added). Such an association view has difficulties dealing with the fact that IAT effects can be based on nonsemantic types of similarity and the fact that IAT effects are highly malleable. These observations do fit very well with the similarity view that is explicitly put forward in this paper.

The ideas and results that we present do not only have implications for the question of *what* the IAT measures, they also provide information regarding the second theoretical question: *How* do IAT effects come about? A first implication is that an adequate process model of the IAT should be able to explain that IAT effects can reflect not only associations in memory or salience asymmetries but also other kinds of similarity. A second implication is related to the role of items and labels in the IAT. In both Experiments 1 and 2, we observed that the degree of perceptual similarity between items can have an important impact on IAT effects. Together with a number of similar findings (e.g., Bluemke & Friese, in press; De Houwer, 2001, Footnote 4; Govan & Williams, 2004; Mierke & Klauer, 2003; Mitchell et al., 2003; Steffens & Plewe, 2001), our results thus suggest that process models of the IAT should at least be able to account for the impact of item properties on IAT effects. This raises doubts about the validity of certain process models of the IAT such as the relevant feature account (De Houwer, 2001). On the other hand, earlier findings showed that labels can also be important (e.g., Bluemke & Friese, 2003; De Houwer, 2001; Mitchell et al., 2003; Neumann et al., 2000). One possible way to reconcile these findings is by assuming that labels exert their influence by biasing the way in which items are encoded or attitudes are constructed. When the IAT is used to measure attitudes, it is possible that the attitude toward the items (and thus the similarity between items of different categories) is determined in part by the label that one uses because the label makes salient certain properties of the items (see above). For instance, when the label SCIENTIST is used, an exemplar such as "Einstein" will probably be perceived as positive and thus similar to exemplars of the category "positive." However, when the label FOREIGN is applied to the stimulus "Einstein," it might be perceived as negative and thus similar to exemplars of the category "negative" (see De Houwer, 2001; Mitchell et al., 2003). Although this account is compatible with the available evidence and with a similarity view, more research is needed before strong conclusions can be made about its validity. For instance, it would be interesting to examine whether the effect of labels depends on the extent to which the labels influence the encoding of the items.

Finally, our proposal that the IAT provides a general measure of similarity could also be relevant for researchers who rely upon the concept "similarity" in their work. The IAT might provide a useful measure of similarity in these areas of research. Most procedures to estimate similarity are designed to deal with rather simple stimuli or rely on explicit ratings of stimulus properties. The IAT, on the other hand, would allow one to measure similarity between fairly complex stimuli and concepts in an indirect manner.

We thank Jan Mierke, Tony Greenwald, and Klaus Rothermund for their helpful comments on earlier drafts.

Correspondence should be addressed to Jan De Houwer, Department of Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Ghent, Belgium (Tel: 0032 9 264 64 45; Fax: 0032 9 264 64 89; E-mail: Jan.DeHouwer@UGent.be).

References

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, III & J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert B. Crowder.*

Science conference series (pp. 117-150). Washington, DC: American Psychological Association.

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

Bluemke, M., & Friese, M. (in press). Do features of stimuli influence IAT effects. *Journal of Experimental Social Psychology*.

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

Collins, A. M., & Quinlan, M. R. (1972). How to make a language user. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 310-351). New York: Academic Press.

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

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

De Houwer, J. (2003b). 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* (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*, 47-59.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118,* 13-42.

Govan, C. L., & Williams, K. D. (2004). Reversing of elimination IAT effects by changing the affective valence of the stimulus items. *Journal of Experimental Social Psychology, 40*, 357-365.

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellot, 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., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197-216

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General, 134,* 420-425.

Hintzmann, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition, 5*, 183-216.

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.

Lasaga, M. I., & Garner, W. R. (1983). Effect of line orientation on various information-processing tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 215-225.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254-278.

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

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

Neumann, R., Totzke, I., Popp, C., & Fernandez, O. (2000). *Distinct attitudes despite similar targets: Comparing German students' attitudes toward "blacks" and foreigners*. Unpublished manuscript.

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

Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.

Ross, B. (1984). Remindings and their effects in learning a cognitive skill. *Cognitive Psychology, 16*, 371-416.

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

Rothermund, Klaus; Wentura, D. (2004). Underlying Processes in the Implicit Association Test: Dissociating Salience From Associations. *Journal of Experimental Psychology: General. 133*(2), 139-165.

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

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

Tversky, A. (1977). Features of similarity. *Psychological Review, 84*, 327-352.

Sommaire

Dans cet article, nous mettons de l'avant l'hypothèse que le Test d'association implicite (TAI) permet une mesure générale de la similarité. Cela suppose que le rendement au TAI reflète non seulement la similarité en ce qui a trait à la signification ou à la prégnance d'une attitude, mais aussi en ce qui touche d'autres particularités. La recherche antérieure suggère en effet que les effets du TAI peuvent être fondés sur les propriétés perceptuelles du stimuli (p. ex., Lasaga et Garner, 1983; Mierke et Klauer, 2003). Nous prolongerons cette recherche en démontrant que le rendement au TAI peut dépendre d'une similarité perceptuelle même lorsque seulement certains éléments des concepts sont semblables (p. ex. photo d'une pizza ronde et d'une pièce de monnaie ronde; photo d'une rivière sinueuse et d'un serpent sinueux; expérience 1). Nous avons également examiné pour la première fois si le contexte du test peut déterminer le type de similarité sur lequel se base le TAI. Les résultats de l'expérience 2 ont montré que le rendement au TAI reflète une similarité en ce qui touche la forme (p. ex. ronde ou sinueuse) lorsque cette particularité perceptuelle a été rendue saillante au cours d'une phase de formation, mais qu'elle reflétait une similarité en ce qui touche une caractéristique fonctionnelle (p. ex. mangeable ou non mangeable) lorsque cette particularité était rendue saillante au cours de la phase de formation.

Les idées et les données qui sont présentées dans cet article comptent un certain nombre de suppositions théoriques. De façon plus importante, notre perception de similarité présente une nouvelle perspective sur la question de ce que mesure le TAI. La proposition qu'il peut mesurer toutes sortes de similarités englobe les propositions antérieures. Il ne faudrait pas l'opposer à l'idée que le TAI mesure les associations en mémoire sémantique (p. ex., Banaji, 2001; Greenwald et al., 1998), non plus que le point de vue que le TAI reconnaît les régularités dans la saillance (par ex., Rothermund et Wentura, 2001, 2004). La perception de similarité amène à l'intuition que ces propositions antérieures ne sont pas mutuellement exclusives, mais peuvent être en fait considérées comme deux cas de principes généraux que le TAI mesure la similarité. Il s'agit aussi de la seule perspective qui est compatible au fait que les effets du TAI peuvent être fondés sur la similarité perceptuelle.

Les idées et les résultats que nous présentons ont non seulement des conséquences sur la question de ce que mesure le TAI, mais présentent aussi de l'information concernant la deuxième question théorique : Comment surviennent les effets du TAI? Une première supposition est qu'un modèle de processus adéquat du TAI devrait être en mesure d'expliquer que les effets du TAI peuvent refléter non seulement les associations en mémoire, mais aussi d'autres genres de similarités. Une deuxième supposition concerne le rôle des éléments et des étiquettes dans le TAI. Le fait que nous avons trouvé un effet du TAI qui était fondé sur une similarité perceptuelle entre des éléments (expériences 1 et 2) montre clairement que les propriétés et les éléments peuvent avoir une incidence importante sur les effets du TAI. En dernier lieu, notre proposition que le TAI fournit une mesure générale de similarité pourrait également être pertinente pour les chercheurs qui comptent sur le concept de « similarité » dans leur travail parce que le TAI pourrait fournir une mesure utile de similarité dans ces domaines de recherche.