European Journal of Social Psychology
Eur. J. Soc. Psychol. **40**, 321–337 (2010)
Published online 8 April 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/ejsp.626

Research article

Picture-IAT versus *Word*-IAT: Level of stimulus representation influences on the IAT

FRANCESCO FORONI[1]* AND TARIK BEL-BAHAR[2]
[1]*Department of Social Psychology, FPP, VU University Amsterdam, The Netherlands*
[2]*Department of Psychology, University of Oregon, USA*

Abstract

The Implicit Association Test is a paradigm designed to assess individual differences in implicit cognition. The goal of this report was to examine the reasons for discrepant effect magnitudes obtained with two presumably interchangeable versions: Picture-IAT (P-IAT) and Word-IAT (W-IAT). We show that this discrepancy is due to the relation between stimuli and referent category: the level of representation (LR) at which a stimulus represents an intended category. Experiment 1 replicates the discrepancies found in previous research. Experiments 2–4 show that increasing the LR of stimuli increases the IAT effect. LR affects the magnitude of the IAT effect even when modality and other features of the stimuli are kept constant. The utility of LR for future investigations examining the IAT paradigm is discussed. Copyright © 2009 John Wiley & Sons, Ltd.

The Implicit Association Test (IAT; Greenwald et al., 1998; see also Devine, 2001) is a reaction time (RT) method developed as an indirect measure of individual differences in the strengths of associations among concepts (e.g., categories and attributes: for reviews see: Hoffman, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, Greenwald, & Banaji, 2006). Critical concerns regarding this paradigm have been partially addressed (e.g., Dasgupta, McGhee, Greenwald, & Banaji, 2000; Greenwald & Nosek, 2001; Nosek et al., 2006), yet the IAT shows serious limitations which require more elaboration (e.g., Blanton, Jaccard, Gonzales, & Christie, 2006; De Houwer & Moors, 2007; Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007; Wentura & Rothermund, 2007).

The present paper, in line with other research (e.g., Bluemke & Friese, 2006; De Houwer, 2002; Govan & Williams, 2004; Steffens & Jelenec, 2007) focuses on the modulating effect that features of the stimuli have on the IAT effect. In this paper, we take a novel look at the relation between the category labels and the stimuli to be classified. For the development of IATs, stimuli sets have been selected to avoid influence of known confounding variables such as social desirability (e.g., Greenwald et al., 1998; Ottaway, Hayden, & Oakes, 2001). Some authors suggest that items should be selected to instantiate the concepts of interest as closely as possible in order to make participants think of the label in the desired way (e.g., De Houwer, 2002; Govan & Williams, 2004; Olson and Fazio, 2003) but different versions have been developed to assess the same construct without much work to verify convergent validity across tests with the implicit assumption that different versions are interchangeable (i.e., supposedly measuring the same underlying construct). For instance, the Ethnic IAT (developed to measure the mental associations individuals hold about White- and Black-Americans) exists as a picture version (*Picture*-IAT) and a word version (*Word*-IAT) that present the same text labels representing a category but different stimuli sets for the relevant category (pictures of individuals and stereotypical first names printed in text, respectively).

*Correspondence to: Francesco Foroni, Department of Social Psychology, FPP, VU University Amsterdam, van der Boechorststr 1, 1081 BT Amsterdam, The Netherlands. E-mail: f.foroni@psy.vu.nl

Received 8 March 2008
Accepted 5 February 2009

These two versions present discrepancies in the results: *P*-IATs, in general, are associated with smaller IAT effects sizes than *W*-IATs (e.g., Mitchell, Nosek, & Banaji, 2003; Nosek, Banaji, & Greenwald, 2002). A better understanding of this discrepancy and its meaning in terms of IAT version interchangeability is the goal of this report.

PICTURE-IAT (*P*-IAT) VERSUS WORD-IAT (*W*-IAT)

Different versions of the same IAT (e.g., Ethnic IAT), even if assumed to be interchangeable, show discrepancies that go against the idea that any given IAT version measures the *same* mental associations (Banaji, 2001; Steffens & Buchner, 2003). The idea of IAT interchangeability rests on the assumption that the stimuli of the *P*-IAT version (i.e., pictures of individuals) and of the *W*-IAT version (i.e., first names such as ''Tyrone'') hold the same relationship with the relevant category (e.g., ''Black Americans''). However, while a picture of a face is an image of *one and only one* individual (i.e., a specific exemplar), first names seem to relate differently to the category (i.e., ''Tyrone'' may represent *any individual* who is named Tyrone).[1] Mental processing of a single forename, in contrast to a single face, may require the activation of multiple exemplars or subgroup knowledge (Park, Ryan, & Judd, 1992; Rothbart & John, 1985) and may involve more categorical versus individuating, or more intergroup versus interpersonal processing biases (Brewer, Weber, & Carini, 1995; Fiske & Neuberg, 1990; Kruglanski & Orehek, 2007).

Imagine now three different sets of stimuli for the same relevant category Black-American: (a) pictures of an individual, (b) first names (e.g., Tyrone), and (c) category labels (e.g., Black-American). We assume that these sets vary in the way they represent the category from individual member to the whole category. If truly interchangeable, the different versions should produce similar and correlated results. We suggest that these versions have stimuli that differ on the level with which they represent the relevant category and that these differences leads to interpretable changes in IAT effect size.

In cognitive models of hierarchical levels of ''representation'' or ''categorization'' (Costa, Mahon, Savova, & Caramazza, 2003; Rosch, Mervis, Gray, Johnson, & Boyes-Braere, 1976) knowledge is assumed to be structured from specific to abstract: subordinate-, basic-, and superordinate- (or category-level). The levels reflect changes in the degree of inclusiveness of lower-order concepts in relation to abstract taxonomic categories.[2] These three levels in concept knowledge seem to correspond to the three sets of stimuli in the above example. The idea of a ''level of representation'' (LR) is not completely new and it has been already applied to studies of social cognition (e.g., Crisp & Hewstone, 2007; Kunda & Oleson, 1995; McGarty, 2006; Park et al., 1992; Rothbart & John, 1985). Changes on this dimension have been shown to produce meaningful effects on cognitive paradigms (e.g., Murphy & Lassaline, 1997). Moreover, a similar dimension from individual to category is implied by social cognition models of dual processes in impression formation (Brewer, 1988; Fiske & Neuberg, 1990; Kruglanski & Orehek, 2007). For these reasons, LR is an interesting candidate to test in a paradigm such as the IAT. In the next section we explore in more detail the reasons that support the view that the LR of stimuli is relevant for the IAT.

LEVEL OF REPRESENTATION AND THE IAT PARADIGM

In this paper we do not attempt to develop a theory of LR influence during dual-choice RT tasks such as the IAT. Realistic cognitive process models of IAT-related mechanisms are currently only in their early stages (Brendl, Markman, & Messner, 2001; Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; De Houwer & Moors, 2007; Klauer et al., 2007). LR is an important dimension in cognitive models of categories and mental representations. For our purposes, we assume that we are examining stimulus LR influences on some aspect of the automatic knowledge activation, stimulus

[1]It is possible that one face will make a person think of several other faces, and that a single first name (e.g., Tyrone) may lead to think about a specific person, but we believe that this is not the norm.
[2]We are aware that Rosch et al. (1976) with different 'level of representation' usually indicated the levels on which a stimulus is classified (usually keeping the stimuli constant). For instance, the picture of a kitchen chair can be classified as 'chair' (i.e., basic level) or as furniture (i.e., super-ordinate level). Nevertheless, the relation between a stimulus (i.e., kitchen chair) and the referent category (e.g., chair or furniture) is well captured by this dimension of ''level of representation''. We borrow this terminology because of the parallels and overlap between our reasoning and their model. However, in this study we use the term level of representation to refer to the degree with which a given stimulus represents the referent category.

DOI: 10.1002/ejsp

perception, decision, and response processes recruited during IAT tasks. In this paper, we try to focus our explanation of effects on such general process components.

In keeping with our views, some authors have suggested that the specific characteristics of different IAT versions might partly determine their findings (Steffens, Kirschbaum, & Glados, 2008) and that different IAT versions might not measure the exact same ''attitudes'' but different representations elicited by the relation of the specific stimuli sets to be classified and the category labels (e.g., Blair, 2002; Devos & Ma, 2008; Steffens et al., 2004; see also De Houwer, 2002; Govan & Williams, 2004). Moreover, the IAT is thought to be more sensitive to mental representations at the level of categories than at the level of the category's exemplars (Nosek et al., 2006, p. 24; Nosek, Greenwald, & Banaji, 2005). Other authors (De Houwer, 2001; Olson & Fazio, 2004) have argued that the IAT paradigm is designed to focus on the association between the target categories (as a whole) and evaluations of these categories. These considerations seems to direct research attention to the relevance of the relation between category label and stimuli. A general principle that IAT effects depend on how well the stimuli instantiate the concepts (i.e., labels) that one wants to measure has been suggested (e.g., De Houwer, 2002; Govan & Williams, 2004). When the referent category label is kept constant (e.g., African Americans), such category–stimuli relations are well captured by the stimulus LR. Thus, we expected that IAT effect size should be influenced by the LR of the stimuli and in particular should increase as stimuli increase in their representation of a category as a whole.

Stimuli at different LR seem to induce processing differences. These basic processing differences may be useful in interpreting the meaning of the IAT effect. Namely, low-LR and high-LR stimuli seem to recruit, respectively, semantic-verbal and perceptual-visual networks (Marsolek, 1999; Tanaka, Luu, Weisbrod, & Kiefer, 1999; Whatmough, Verret, Fung, & Chertkow, 2004). High-LR stimuli engage more semantic interference or processing difficulty compare to lower-LR stimuli (e.g., Nelson & Schreiber, 1992; Paivio, Walsh, & Bons, 1994; Schwanenflugel & Shoben, 1983). In contrast we assume faces are concrete and specific low-LR exemplars that are processed with a subordinate individuation bias (D'Lauro, Tanaka, & Curran, 2008; Tanaka, 2001). High-LR stimuli may be more difficult to process because high-LR stimuli tend to activate multiple lower LR exemplars *and* perceptual features related to the category (e.g., De Wilde, Vanoverberghe, Storms, & De Boeck, 2003; Heit & Barsalou, 1996; Murphy & Smith, 1982 see also fan effect: Anderson & Reder, 1999).[3]

Our general expectations are consistent with brain-based studies showing that faces (low-LR stimuli) and words (high-LR stimuli) are processed preferentially by partly separate networks responsible for perceptual and semantic processing (e.g., Henson, Price, Rugg, Turner, & Friston, 2002; Ito & Ulrand, 2005; Joyce, Cottrell, & Tarr, 2003; Puce, Allison, Asgary, Gore, & McCarthy, 1996; Tarkiainen, Cornelisen, & Salmelin, 2002; Todorov, Gobbini, Evans, & Haxby, 2007). Other work also suggests concrete concepts preferentially recruit perceptual and visual brain networks more than abstract concepts which preferentially recruit semantic and verbal brain networks (e.g., Large, Kiss, & McMullen, 1998; Sabsevitz, Medler, Seidenberg, & Binder, 2005; Vitkovitch & Underwood, 1991; West & Holcomb, 2000; Whatmough et al., 2004). Models that propose a right lateralized brain network for individuation processing (with a perceptual bias) and a left lateralized brain network for categorical processing (with a semantic-verbal bias) (Zarate, Stoever, Maclin, & Chavex, 2008) further support our view regarding strong processing differences for stimuli at different LR.

Our initial assumption is that LR, as shown for other paradigms, is a relevant dimension that influences the IAT effect. Low-LR stimuli (i.e., pictures) in the IAT should produce a smaller effect size than high-LR stimuli (i.e., first names labels). We assume that pictures of individuals are low-LR stimuli that refer to one person, are concrete, activate less extraneous knowledge, are easier to categorize, and have perceptual features that minimize semantic interference effects. In contrast first names, and especially more super-ordinate text stimuli, are high-LR stimuli that may refer to more than one exemplar, are abstract, activate more extraneous knowledge, are harder to categorize, and have text features that enhance semantic interference effects. These are idealized distinctions by which we attempt to highlight the most important differences that we believe exist between *W*-IAT and *P*-IAT versions.

[3]The semantic fan effect (Anderson & Reder, 1999) refers to the finding that as people study more facts about a concept it takes them longer to respond to questions about that concept. This research implies that a stimulus at the basic-level (first name) or superordinate-category level (group name) should refer to more than one individual, and thus more associations, than a picture of a specific individual. Thus, larger *W*-IAT effect sizes may be a function a greater semantic fan related to the word stimuli used. The smaller *P*-IAT effect sizes may be a function of the smaller semantic fan that is activated with the greater amount of specific perceptual-visual information available from images. Further, work by researchers such as Heit and Barsalou (1996) suggests that the classical distinctions regarding LR do not hold in everyday cognition, wherein task set may initiate the instantiation of perceptually-based or task-based categories. We do not address this issue in the current report, although it is associated with the similarity processing hypothesis for IAT effects (De Houwer et al., 2005).

DOI: 10.1002/ejsp

We hypothesize that larger *W*-IAT than *P*-IAT effect sizes result from a different LR of the stimuli sets in relation to the referent category. When using stimuli at different LR researchers may create IAT versions that are not be interchangeable due to the different processing induced by the different LR.

OVERVIEW

The goal of this set of studies was to address the discrepancy in IAT effect size found when using two different versions of the IAT (*P*-IAT and *W*-IAT). In the first experiment, the usual discrepancies (i.e., *P*-IAT shows smaller IAT effect) were replicated in a within-subject design in a different cultural context (The Netherlands where part of the research was conducted). Second, we tested whether different sets of stimuli (varying on LR) show differences in the magnitude of the IAT effect (Exp. 2). In explaining this discrepancy, however, one needs to exclude alternative explanations. In Experiment 3 and 4 (using social and non-social categories) LR of the stimuli was manipulated while keeping other possible sources of the discrepancy constant (e.g., stimulus modality, familiarity). This is the first time to our knowledge that LR is examined in the context of the IAT.

EXPERIMENT 1

Research has shown that, in general, *P*-IATs show smaller IAT effects than *W*-IATs (e.g., Dasgupta et al., 2000; Greenwald, 2004; Nosek et al., 2002). These comparisons are usually done between subjects and the evidence for the lack of discrepancies is generally based on null results. The goal of this experiment was twofold: first, testing whether *P*-IAT and *W*-IAT show the usual discrepancy in the IAT effect in a within-subjects design; second, testing the generalizability of the *P*-IAT/*W*-IAT discrepancy in a different cultural context (The Netherlands).[4]

Notably, the two ‘‘*Ethnic*’’ IAT versions used here are exactly the same (same labels and same stimuli for positive/negative classification) with the only exception being stimuli for the classification in the relevant categories (i.e., photos vs. names stimuli).

Method

Participants

Forty Dutch students from the VU University Amsterdam participated in the computer-based experiment in exchange for 5 Euros (*circa* 6 dollars).

Stimuli and Tasks

Each participant worked through two IATs (one *P*-IAT and one *W*-IAT) that differed only in the stimuli used in the category classification (Native/Immigrant). For the *W*-IAT, the stimulus material consisted of 24 word-stimuli (six first names stereotypically associated with Immigrants, six Dutch first names, six bad words, and six good words). For the *P*-IAT, the stimulus material consisted of 12 words (six bad and six good words) and 12 pictures (six pictures of dark-skin tone individuals and six pictures of light-skin tone individuals). Stimulus materials were taken from previously published IAT

[4]This experiment, as well as Experiment 4, was run in the Netherlands where African Americans and White Americans are not meaningful social categories. The classification Allochtoon/Autochtoon (in English: Immigrant/Native) was instead used. This distinction is overlapping with the black/white classification both in term physical distinction and in term of evaluative distinction. For the purpose of these specific experiments, the possible differences in social representation between European-Americans versus African-Americans on one hand, and Immigrant versus Indigene on the other hand, are not relevant. In fact, the negative bias among Dutch participants against immigrants of dark skin-tone is similar to the one found amongst European-American participants against African-Americans.

DOI: 10.1002/ejsp

research (Dasgupta et al., 2000; Greenwald et al., 1998; Nosek & Banaji, 2001). Stimuli were pre-tested for evaluation. Fifteen participants rated on a 9-point scale how positive or negative each stimulus was among a larger sample of stimuli. Stimuli sets were selected to be of equal evaluation across LR (ts(14) < 1.4, ps = ns.). Picture luminosity and hue (for all the experiments reported here) were equalized as much as possible using photo editing software. See the Appendix for a full list of all stimuli used in all experiments reported in this paper.

Procedure

After being introduced to the task, participants completed the IAT as in Greenwald et al. (1998) with the only exception that participants performed two versions of the IAT (in counter-balanced order across participants). Per assessment, each participant performed a total of five blocks: (1) Immigrant/Indigene discrimination, 24 trials; (2) evaluative attribute discrimination, 24 trials; (3) combined, compatible discrimination, 72 trials; (4) practice of reversed Immigrant/Indigene discrimination, 24 trials; and (5) combined, incompatible discrimination, 72 trials. Order of block presentation for compatible and incompatible blocks was counter-balanced across participants. The IAT effect was obtained by comparing performance in blocks (3) and (5). Stimuli for each block were sampled randomly without replacement from the stimulus list. Participants were explicitly instructed to respond to each stimulus as fast as possible while avoiding errors.

Design and Hypotheses

The within-subject factor, IAT-type, had two levels (*P*-IAT and *W*-IAT). We hypothesized that the participants would show the usual discrepancy with *P*-IAT, showing a smaller IAT effect than *W*-IAT.

Results and Discussion

Data reduction followed the ‘‘improved scoring algorithm’’ proposed by Greenwald et al. (see Greenwald, Nosek, & Banaji, 2003, Table 4, p. 214). The dependent measure for our analyses was the Cohen’s *d*. Both IAT versions show a significant IAT effect replicating previous research. As predicted the IAT-effect for the *P*-IAT ($d' = .18$) was smaller than for the *W*-IAT ($d' = .30$), t (39) = −2.42, $p = .01$.[5] No other effect was significant. Finally, these two IAT versions had a moderate positive correlation ($r = .44$, $p < .01$).[6] Results using raw RT data replicate in their significance those using the improved algorithm and so will not be reported in detail (see Table 1 for RTs and error rate data).[7]

This result replicates previous findings of the discrepancy between the two IAT versions in a within-subject design and in a population outside of the United States with a partially modified set of category labels. Moreover, this result confirms that the IAT effect is modulated by different stimuli sets (e.g., Bluemke, & Friese, 2006; Govan & Williams, 2004; Mitchell et al., 2003; Nosek et al., 2006; Steffens & Plewe, 2001). However, while previous research manipulated stimulus features such as valence within the same IAT version, the discrepancies reported here are between two different versions of the *Ethnic*-IAT that are the same except for the stimuli used for the classification of the ethnic categories, making this comparison particularly strong.

[5]This experiment also replicated the second discrepancy between *P*-IAT and *W*-IAT. Participants in the *P*-IAT showed faster general RTs ($M = 963$ ms.) than in the *W*-IAT ($M = 1011$ ms.), t (39) = −2.70, $p < .010$. However, this discrepancy in RTs is not the topic of this paper so it will not be discussed further.

[6]Cunningham, Preacher, and Banaji (2001) suggested that the bivariate correlations may partially underestimate the inter-relation between two IATs. However, other authors have provided evidence for different views and have suggested different conclusions about Cunningham and colleagues’ models (Steffens & Buchner, 2003).

[7]Significant IAT effects were found in all the IAT versions used in this report. In addition, in all the studies counterbalancing order of the IATs, order of the combined blocks and error rates parallel results from previous research (e.g., Greenwald et al., 1998). For instance, error rate is significantly higher in incompatible blocks (except in Experiment 3) and significantly lower in IAT involving pictures compare to words (e.g., Experiment 1 and 2). However, these variables never interact with the variable of interest. Therefore, these variables and error rate analyses will not be presented or mentioned further.

DOI: 10.1002/ejsp

Table 1. IAT effect, RTs, and percentage of errors by experiment and IAT-types

			RTs ms. (SD)			Error percentage	
IAT type	Stimuli level	d'	Compatible block	Incompatible block	Difference	Compatible block (%)	Incompatible block (%)
Exp. 1							
P-IAT	Subord.	.18	899 (179)	1026 (265)	127	4.5	7.6
W-IAT	Basic	.30	908 (174)	1115 (209)	207	6.5	9.0
Exp. 2							
P-IAT	Subord.	.14	682 (117)	742 (120)	60	5.0	6.7
F/N-IAT	Subord.	.17	836 (151)	906 (158)	70	7.2	9.2
W-IAT	Basic	.24	766 (167)	859 (132)	93	6.5	9.7
W/C-IAT	Super-ord.	.32	742 (149)	878 (142)	136	6.2	9.8
Exp. 3							
Sub.-IAT	Subord.	.45	759 (155.)	921 (204)	162	6.8	8.3
Basic-IAT	Basic	.58	705 (108)	888 (156)	183	6.3	7.0
Exp. 4							
P-IAT	Subord.	.16	790 (122)	887 (155)	97	4.8	6.3
Group-IAT	Basic	.29	824 (128)	975 (203)	151	4.9	5.8

Note: All IAT effects are significantly different from 0.; Subord. = Subordinate level; Basic = Basic level; Super-ord. = Super ordinate level.

EXPERIMENT 2

The goal of Experiment 2 was to show that sets of stimuli with different LR would produce different magnitudes of the IAT effect while sets of stimuli with equal LR would not show different magnitudes of the IAT effect. Participants in this study performed four different IATs. Each IAT presented a unique set of stimuli for the Blacks/Whites classification. Each set was created or selected to represent the relevant social categories at different LR. All other aspects across the four different IAT versions (i.e., labels, bad words, and good words) were kept constant.

Method

Participants and Material

Forty students of the University of Oregon participated in the computer-based experiment as partial fulfillment for course requirements. The stimulus material was taken from previously published IAT research (Dasgupta et al., 2000; Greenwald, 2004; Nosek et al., 2002) and supplemented with new stimuli.[8] The material consisted of six bad words and six good words common to all four IATs. The Blacks/Whites stimuli varied according to the IAT-type. For the *P*-IAT, stimuli consisted of 12 faces of unknown people (six Black Americans and six White Americans); for the *Full/Name*-IAT (*F/N*-IAT stimuli consisted in 12 names of famous people (six Black Americans and six White Americans); and, for the *W*-IAT, stimuli consisted of 12 first names (six stereotypically Black-American and six stereotypically White American). Finally, for the *Whole/Category*-IAT (*W/C*-IAT), stimuli consisted of six symbols (e.g., black triangle and white triangle) and six labels (e.g., ''African-American'' and ''European-American'').

[8]Stimuli for this experiment were either taken from previously published IAT research or generated for this research. *P*-IAT and *W*-IAT stimuli were pre-tested by other authors (Greenwald et al., 1998; Dasgupta et al., 2000); *F/N*-IAT stimuli were pre-tested by other authors (Dasgupta & Greenwald, 2001) and selected to be half positive and half negative to avoid to have a confound with evaluation; finally, stimuli for the *W/C*-IAT have been generated to be representative of the two referent categories.

DOI: 10.1002/ejsp

Procedure

Procedure, type, and sequence of blocks were identical to Experiment 1, unless otherwise specified. Participants performed four different IATs (in counter-balanced order across participants).[9] Per assessment, each participant performed a total of five blocks: (1) Black–White discrimination, 24 trials; (2) evaluative attribute discrimination, 24 trials; (3) combined, compatible discrimination, 72 trials; (4) practice of reversed Black–White discrimination, 24 trials; and (5) combined, incompatible discrimination, 72 trials. At the end of the IAT task participants rated all the stimuli on the dimension LR as a manipulation check. Participants were asked to rate the degree with which each stimulus represents the intended referent category. To do so, participants had to choose from nine possible graphical representations as responses (i.e., a circle representing the relevant category colored in different proportion). This scale was adapted from Aron, Aron, and Smollan (1992). Each one of the nine image represented a different degree of LR. The circle went from having only a colored dot within it (indicating that a stimulus represents a minimal portion of the category) to being fully colored (indicating that the stimulus represents the full category).

Design and Hypotheses

The goal of this experiment was to test the effect of 4 different sets of stimuli (that differ in their LR) on the magnitude of the IAT effect. The within-subject factor, IAT-type, presented four levels (*P*-IAT, *F/N*-IAT, *W*-IAT, and *W/C-IAT*). We expected that the magnitude of the IAT effect would be influenced by IAT-type. In particular, it was hypothesized that (a) the *P*-IAT and *F/N*-IAT would show the smallest IAT effects and would not differ from each other (since the stimuli for both represent specific individuals: i.e., subordinate or single-person LR); (b) the *W*-IAT would show a larger IAT effect (since first names represent a subgroup of the target categories: i.e., basic level), and (c) that the *W/C*-IAT would show the largest IAT-effect (since the stimuli represent the intended super-ordinate categories: i.e., category level).[10]

Results

Implicit Association Test

The magnitude of the IAT-effect was significantly influenced by the factor IAT-type, $F(3, 37) = 6.25$, $p = .002$. Examination of planned comparisons showed that the *F/N*-IAT ($M = .16$, SD = .26) did not differ from the *P*-IAT ($M = .14$, SD = .23), $F < 1$, $p =$ ns. In addition, they showed that both the *W*-IAT ($M = .24$, SD = .27) and the *W/C*-IAT ($M = .32$, SD = .25) differed from the *P*-IAT significantly ($F[1,39] = 3.90$, $p = .05$ and $F[1,39] = 18.89$, $p < .001$, respectively). *F/N*-IAT also differed from *W/C*-IAT ($F[1,39] = 8.06$, $p = .007$) but did not differ at standard level of significance from *W*-IAT ($F[1,39] = 2.33$, $p = .067$ [one-tailed]).

The correlations between the different IAT versions are reported in Table 2. The four IAT versions did not always significantly correlate with each other. To further understand the pattern of correlations, a statistical test of difference among the dependent correlations of the matrix was run using Williams' proposed formula as suggested by Steiger (1980). Table 2 also show which correlations differ from each other. Of particular interest is the fact that several comparisons between correlation in the matrix show significant differences indicating an inconsistent pattern of correlations between IAT types.[11]

[9]Four different counterbalanced orders (out of the possible 24) were chosen. Each IAT type was run equally in the first, second, third, and fourth position. We believe that this incomplete counterbalanced design does not affect the results. In fact, other authors have run several tasks without counterbalancing the order and showed no order effect (e.g., Bosson, Swann, & Pennebaker, 2000). Moreover, separate analyses on the present data based on the counterbalanced orders did not show any differences; therefore, this factor will not be discussed further.

[10]An implication of our hypothesis on the LR is that pictures of famous people and unknown people hold the same relationship with the category (i.e., they all represent individual members), making them interchangeable. On the other hand any name (e.g., "Joe Smith") may represent different people that share that name. For this reason we decide to use names of famous people. In fact, names of famous people (e.g., "Al Capone") are indeed thought as a specific individual because of its fame. Pictures of unknown people (in the *P*-IAT) and full names of famous people (in the *F/N*-IAT) according to our hypotheses should be interchangeable because representing individual people.

[11]A more definitive demonstration that different IAT versions are not measuring the same exact construct would be the case when inter-correlations between the different IAT versions (see Table 2) were lower than test–retest reliabilities of such versions. Initial evidence on this question (Foroni, 2007) is contradictory with some data supporting the interchangeability assumption of the IAT versions and other data clearly casting doubts on this assumption.

DOI: 10.1002/ejsp

Table 2. Correlations between IAT versions in Experiments 2

	IAT-type			
	Picture-IAT	*Full/Name*-IAT	*Word*-IAT	*Whole/Category*-IAT
Picture-IAT	—	$^{**}0.41^{a}$	0.12^{b}	$^{**}0.42^{a}$
Full/Name-IAT		—	$^{*}0.31^{ab}$	0.17^{b}
Word-IAT			—	0.08^{b}
Whole/Category-IAT				—

$^{*}p < .05$; $^{**}p < .01$. Note: Different letters indicate significant different correlations (all *p*s < .07) based Williams' formula to test difference between dependent correlations as suggested by Steiger (1980).

Manipulation Check

Average ratings for the four different sets of stimuli (four different IAT versions) were analyzed to parallel the analyses on the IAT effect. As expected, the results on the ratings parallel those on the IAT effect. The factor IAT-type was significant (F [3,37] = 9.22, $p < .001$) for the rating scores. The four stimuli sets were seemingly perceived as representing the referent category at different degrees. Moreover, planned contrasts showed that the stimuli represent the referent category linearly, increasing from *P*-IAT to *F/N*-IAT, *W*-IAT, and *W/C*-IAT (F [1,39] = 4.54, $p = .039$).

Discussion

The results of Experiment 2 are in line with previous research that shows that the IAT is sensitive to different types of stimuli (e.g., Govan & Williams, 2004). The magnitude of the IAT effect varied across IAT-types according to the LR of the stimuli used (as supported by the results of the manipulation check). These results demonstrate that the IAT is sensitive to stimulus type, and in particular, the more the stimuli represent the category as a whole, the larger the IAT effect. When stimuli are at the same LR (as in the case of the *P-IAT* and *F/N-IAT*), they do not show a difference in the magnitude of the IAT effect. Finally, the correlations among versions are not always significant and, in some cases, are significantly different from each other. This inconsistent pattern of correlations cast some doubts on the idea that all IAT versions are *necessarily* interchangeable.

Experiment 2 also suggests that the stimulus modality (i.e., faces vs. text) is not as relevant as LR. However, Experiment 2 presents some potential limitations. The stimuli were selected from previous IAT research to be at different LR, but were pre-tested for evaluation and not for LR. The manipulation check regarding LR was done after the IATs, but suggests that our stimuli sets were readily perceived at the intended LR. Nevertheless, it would be important to pre-test the stimulus material in a separate experimental sessions and using a separate sample. Additionally, while differing on LR, the stimuli also differed on other dimensions (e.g., modality and familiarity). For instance, in the *W/C*-IAT the stimuli included both graphical-facial items and text items; while the other versions use either faces (i.e., *P*-IAT) or text (i.e., *F/N*- and *W*-IAT) items. The *F/N*-IAT presented names of famous people, while *P*-IAT and *W*-IAT presented stimuli that are either unspecified (i.e., first names) or unfamiliar (picture of an unknown person). One could argue that modality or favorability (and not LR) is responsible for the difference in IAT effect found in Experiment 2. We believe this is not the case. First, different modalities did not show significantly different IAT-effects (i.e., the effect found with *F/N*-IAT is not different from the one assessed by *P*-IAT). Second, the ethnic versions of the *P*-IAT and *W*-IAT in Experiment 1 presented stimuli with similar levels of familiarity and favorability, but nevertheless they showed significantly different IAT effects. However, to rule out these alternative possibilities we needed to show that manipulating the LR while keeping stimulus modality and other factors constant would again influence the IAT effect. The goal of Experiment 3 was to directly test this hypothesis.

EXPERIMENT 3

The results of Experiment 2 suggest that stimulus modality (i.e., faces vs. text) does not explain differences in IAT-effect and suggest that, in contrast, LR may be an important dimension. To test this conclusion directly and rule out possible

DOI: 10.1002/ejsp

alternative explanations (e.g., evaluation and familiarity) two *W*-IATs were implemented using stimulus material from research on the level of categorization that was pre-tested and selected to be different on the LR but equal in term of familiarity, evaluation, and other factors (e.g., Job, Rumiati, & Lotto, 1992). In addition, this material was made up of only text items, thus also excluding modality as an alternative explanation of the differences in IAT effect size. Participants performed two IATs that differed only on the stimuli sets used for the categorical classification: Musical Instruments and Weapons. Stimuli for these two versions were selected to be either at the subordinate-level (*Subordinate*-IAT) or at the basic-level (*Basic*-IAT).

Method

Participants and Material

Thirty-two Dutch students of the VU University Amsterdam participated in the computer-based experiment in exchange for 5 Euros (*circa* 6 dollars). The stimulus material consisted of four bad words and four good words common to both IATs while the stimuli for the classification into ''Musical-Instruments''/''Weapons'' varied according to the IAT-type. For the *Subordinate*-IAT, stimuli consisted of four musical instruments (e.g., drum) and four weapons (e.g., knife). For the *Basic*-IAT, stimuli consisted of the names of the correspondent higher order category of musical instruments (e.g., percussions for drum), and of weapons (e.g., blade for knife). The material was adapted from previous research (Greenwald et al., 1998; Job et al., 1992) and was pre-tested. Fifteen participants rated on a 9-point scale the positive and negative valence of each stimulus word within each of the four stimulus sets. The only significant difference was between musical instruments ($M = 5.42$; SD $= 1.05$) that were rated in general as more positive than weapons ($M = 3.16$; SD $= 1.30$), $F\,(1,14) = 18.31$, $p = .001$. In addition, stimuli were also pre-tested on LR (using the same response scale from the manipulation check of Experiment 2). The stimuli of the *Subordinate*-IAT showed the expected significant lower LR of the category in contrast to the corresponding stimuli from the *Basic*-IAT for both weapons ($t[35] = 7.82$, $p < .001$) and musical instruments ($t[30] = 1.60$, $p = .06$).

Procedure

Procedure, type, and sequence of blocks were identical to Experiment 1, unless otherwise specified. Participants performed two different IATs. Per assessment, each participant performed a total of five blocks: (1) Musical-Instruments/Weapons discrimination, 24 trials; (2) evaluative attribute discrimination, 24 trials; (3) combined, compatible discrimination, 48 trials; (4) practice of reversed Musical-Instruments/Weapons discrimination, 24 trials; and (5) combined, incompatible discrimination, 48 trials.

Design and Hypotheses

The within-subject factor, IAT-type, presented two levels (*Subordinate*-IAT and *Basic*-IAT). We hypothesized that the *Subordinate*-IAT would show a significantly smaller IAT effect than the *Basic*-IAT.

Results and Discussion

Four participants displayed high error rate (higher than 25%), and therefore, were not included in the analyses. However, including these participants did not change the results. The IAT effect was significantly influenced by the factor IAT-type, $t\,(27) = 1.92$, $p = .033$ (one-tailed). As predicted participants in the *Subordinate*-IAT ($M = .45$, SD $= .54$) showed a significantly smaller IAT effect than in the *Basic*-IAT ($M = .58$, SD $= .31$). In this experiment, these two non-social IAT versions show a positive correlation, $r = .77$, $p < .001$.

DOI: 10.1002/ejsp

Basic-level stimuli produced larger IAT effects than subordinate-level stimuli. The IAT effect using non-social categories varied across IAT-types according to the LR of the stimuli while keeping the stimulus modality constant (i.e., text). In this experiment, only stimuli with clearly defined LR were used, allowing for a strong test of our hypotheses.[12] In addition, the stimuli were only text items providing evidence against alternative explanations such as "difference in stimulus modality." Previous research on the interplay between linguistic categorization and LR has relied almost exclusively on non-social categories (e.g., Costa et al., 2003; Large et al., 1998; Rosch et al., 1976; but see Cantor & Mischel, 1977; Carlston & Smith, 1996; John, Hampson, & Goldberg, 1991; Stangor & Lange, 1994). One of the reasons for this may be that even if mental categories referring to various social groups have a clear distinction between different levels (i.e., subordinate, basic, and super-ordinate), these levels may vary from individual to individual as compared to non-human object categories. Our hypothesis regarding LR also applies to social categories (as suggested by the results of Experiment 2). It would be important to replicate the findings of Experiment 3 using IATs involving social categories. This was the goal of Experiment 4.

EXPERIMENT 4

In Experiment 4, the LR hypothesis was tested in the domain of social category with pictures of individuals (i.e., again keeping stimulus modality constant). Participants performed two different *P*-IATs. Each IAT presented a unique set of stimuli for the Immigrant/Native classification representing the categories in different degrees. The first IAT (*Individual*-IAT: low LR) presented picture of single individuals, while the second presented pictures of groups of individuals (*Group*-IAT: high LR).

Method

Participants and Material

Forty Dutch students of the VU University Amsterdam participated in the computer-based experiment in exchange for 5 Euros (*circa* 6 dollars). The stimulus material consisted of six bad words and six good words common to both IATs. The Immigrant/Native stimuli varied according to the IAT-type. For the *Individual*-IAT, stimuli consisted of 12 faces of unknown individuals (six dark-skin tone individuals and six light-skin tone individuals); for the *Group*-IAT stimuli consisted of 12 pictures of groups of people of homogenous ethnicity (six groups of dark-skin tone individuals and six groups of light-skin tone individuals). The groups of "Immigrant" and "Native" people were equated for number of people and activity (e.g., summer-school participants, office picture, graduation, etc.). Pictures from the four stimuli sets were rated on valence by 20 participants on a 9-point Likert scale. The only significant difference was between picture of individuals ($M = 5.30$; $SD = 1.15$) that were rated as less positive than picture of groups ($M = 5.71$; $SD = 0.94$), $F(1,19) = 16.13$, $p < .01$. In addition, the images were pre-tested on the degree to which they represent the intended IAT category. The stimuli of the *Individual*-IAT showed the expected significantly lower LR in contrast to the corresponding stimuli in the *Group*-IAT for both the dark-skinned individuals ($t[47] = -1.90$, $p = .03$) and for the light-skinned individuals ($t[47] = -2.39$, $p = .02$). Stimulus material for the two IATs had the same modality (i.e., visual-facial), and similar level of familiarity (i.e., unknown people) and favorability (i.e., mildly positive).

Procedure

Procedure, type, and sequence of blocks were identical to Experiment 1, unless otherwise specified. Participants performed two different IATs. Per assessment, each participant performed a total of five blocks: (1) Immigrant/Native discrimination, 24 trials; (2) evaluative attribute discrimination, 24 trials; (3) combined, compatible discrimination,

[12]Using different categories (i.e., "animals" and "weapons") the same results were obtained (Foroni, 2007).

DOI: 10.1002/ejsp

48 trials; (4) practice of reversed Immigrant/Native discrimination, 24 trials; and (5) combined, incompatible discrimination, 48 trials.

Design and Hypotheses

The within-subject factor, IAT-type, presented two levels (*Individuals*-IAT and *Group*-IAT). It was hypothesized that *Individual*-IAT would show a significantly smaller IAT effect than the *Group*-IAT.

Results and Discussion

The magnitude of the IAT effect was significantly influenced by the factor IAT-type, $t(39) = 2.86$, $p = .007$. As predicted participants in the *Individual*-IAT ($M = .16$, SD = .21) showed a significantly smaller IAT effect than in the *Group*-IAT ($M = .29$, SD = .20). Again, the IAT effect varied across IAT-types according to LR of the stimuli. In addition, these two IAT versions were not significantly correlated ($r = .13$, ns.) lending further support to the claim that not all the versions of the IAT are necessarily interchangeable because they are not necessarily assessing the same construct. These two *P*-IATs seem to elicit and capture different aspects of automatic associations regarding the categories. The lack of correlation between these two IATs may suggest that they are tapping partially separate processes. This suggestion includes the possibility that pictures of groups have unique properties, such as priming approach or avoidance, or threat, more than single faces. To our knowledge, no other study has so far compared pictures of groups and individuals so any *post hoc* alternative explanations remain at this point highly speculative and they will be discussed further in the general discussion.

GENERAL CONCLUSIONS

When comparing the results obtained with two different IAT versions (i.e., *W*-IAT and *P*-IAT), previous authors noted that *P*-IAT shows faster RTs and smaller IAT effects (e.g., Dasgupta et al., 2000; Nosek et al., 2002). These discrepancies were replicated in a within-subject design (Exp. 1). In the remaining experiments, we showed that the discrepancy in effect size may be attributable to the different levels of representation of the stimuli sets (Exp. 2–4). Differences in the results obtained using the *W*-IAT and the *P*-IAT might thus be explained in terms of the primary feature on which the corresponding stimuli sets differ: the LR of stimuli exemplars in relation to the specific intended category. In our studies the LR level was manipulated while minimizing the impact of other stimulus features. Other dimensions of the stimuli were also controlled for: familiarity and modality (Exp. 3 and 4) and valence (Exp. 1, 3, and 4). Experiments 2–4 are consistent with previously documented effects of stimuli on the IAT (e.g., Bluemke, & Friese, 2006; Govan & Williams, 2004; Nosek et al., 2005). Our results confirm that IAT effects are not only functions of the category labels (as suggested by De Houwer [2001]), that here were kept constant, but they are also function of the stimuli used (for a similar position, see Nosek et al. [2006]).

The correlation pattern across IAT versions was inconsistent making strong conclusions difficult. Unfortunately, this is a known problem with the IAT paradigm (e.g., Steffens & Bucher, 2003). In Experiment 2, for example, the inter-version correlation is significant but low ($r = .21$), and correlations range from 0.08 to .4 with a non-significant correlation between the two versions of interest: *P*-IAT and *W*-IAT ($r = .12$). In addition, the correlations in this matrix do significantly differ from each other. Moreover, in Experiment 4 *Individual*-IAT and *Group*-IAT also showed a non-significant correlation ($r = .13$). Further research that compares IAT versions inter-correlation and test/re-test reliabilities is needed to better understand the shared and non-shared construct and method variances across IAT types (e.g., Smith-McLallen, Johnson, Dovidio, & Pearson, 2006; Steffens & Jelenec, 2007). However, our results clearly pose some doubts on interchangeability assumptions and instead suggest that different IAT versions may assess different sets of associations regarding the same social category, or alternatively, non-overlapping category evaluations.

This report does not speak directly to the validity of the IAT as a measure to assess individual differences in implicit cognition (see De Houwer et al., in press), but rather focuses on potential influences of LR on the overall magnitude of IAT

DOI: 10.1002/ejsp

effects. However, taken together, the results support the claim that the IATs are not necessarily interchangeable just because using the same referent category labels: the IAT paradigm should be considered a family of instruments where different versions are not necessarily assessing the same construct unless shown to do so with appropriate theoretical and psychometric development (Banse, Seise, & Zerbes, 2001; De Houwer & Moors, 2007; Steffens & Bucher, 2003). From our point of view, differences and similarities between IAT versions need to be tested and stimulus features such as LR of stimuli need to be taken into consideration in order to explain existing and future differences between IAT types. LR may serve as a useful window for taking a new look at mental processes underlying the IAT.

We have alluded throughout the paper that there are differences between stimuli at different LR that have been established outside of the IAT literature, and ignored to date by IAT researchers. We suggested that LR may influence the IAT by inducing processing differences and we tested the expectation that LR directly influences the magnitude of the IAT effect. Different IAT effects found with distinct IAT versions (e.g., *P*-IAT vs. *W*-IAT) may be indicative of different types of task engagement, cognitive processes, or emotional processes. As detailed in the introduction, pictures, and concrete words may engage visual and perceptual systems most, while words, and more abstract words specifically, may engage semantic and long-term memory systems most (e.g., Laeng, Zarrinpar, & Kosslyn, 2003; Whatmough et al., 2004). Thus different stimuli may lead to different types and amounts of interference during the IAT tasks instead of being indicative of a different magnitude of the same underlying dimension (i.e., implicit evaluations of ethnic groups).

Further general support for the hypotheses that different IATs may access partly separate types of processing was provided by correlational evidence (Foroni & Schubert, 2007). These authors examined whether three versions of the IAT (*Picture*-IAT, *Word*-IAT, and *Group*-IAT) were differentially associated with an approach/avoidance task (from Castelli, Zogmaister, Smith, & Arcuri, 2004). The authors hypothesized that the *P*-IAT would be significantly correlated with performance on the approach/avoidance task while the other two versions would not. The results confirmed their hypothesis. *P*-IAT was moderately but significantly correlated with approach/avoidance tendency while the other two versions were not. These results, albeit correlational and preliminary, are in line with the idea that different IAT versions may be associated with different task sets (see also Devos & Ma, 2008). These results may be explained by the P-IAT's preferential activation of emotion, response, and motor processes (e.g., Carver & Scheier, 2002; Ellsworth & Scherer, 2003; Frijda, 2000). By virtue of using more concrete, specific, and possibly more socially and emotionally relevant stimuli, *P*-IAT stimuli may be experienced as more realistic and engaging real-world stimuli, leading to greater overall engagement of participant's behavioral approach-avoid tendencies rather than the semantic and categorical thinking which *W*-IATs were built to access.

Further testing and more detailed research should verify if such difference in correlation is a significant indication of separate types of processing. It remains speculative why, at a cognitive processes level, LR should have any influence on IAT. The IAT category-attribute labels (e.g. White-Good or White-Bad) probably serve as attractors in a cognitive decisional space (Klauer et al., 2007) that allow for easier classification of stimuli into compatible versus incompatible combinations. It is possible that the compatible category-attribute labels are experienced in a manner similar to what we expect for high-LR stimuli, as broader, more inclusive, more semantic, and less perceptual and specific that low-LR stimuli. Higher LR stimuli may increase IAT effects size by facilitating the classification of stimuli into compatible categories. It is also possible that higher-LR stimuli also create more interference, and longer decision time, for classification into incompatible category-attribute combinations. In this sense, LR effects, if any, maybe influencing IAT effects mainly at the level of ease and speed of information accumulation (Klauer et al., 2007).

In this report, the LR hypothesis was also tested against an alternative hypothesis (stimulus modalities or cross-modality interference). The results clearly excluded this alternative hypothesis (as well as other dimensions such as familiarity and valence). Instead, these results support our view that LR modulates the magnitude of the IAT effect (see Table 1 for a summary of the IAT effects across experiments). First, in Experiment 2 the *P*-IAT and the *Full/Name*-IAT (i.e., different modalities) do not show different IAT effects. Second, in Experiments 3 and 4 stimulus modality was kept constant and LR was manipulated leading, nevertheless, to consistent differences in IAT effects. Nosek et al. (Supplementary Study A, cited in Nosek et al., 2005) argued that the use of pictures versus words does not necessarily alter the IAT. Our results suggest instead that it does; however, we believe that the difference is not due to stimulus modality *per se* (pictorial vs. verbal), but instead is due to the distinct LR relation that the stimuli hold with the intended category.

In order to test the LR hypothesis and be sure to manipulate the stimuli and not the category of reference, we always kept constant the category labels (e.g., Black Americans) and we manipulated only the LR of the stimuli. However, one could wonder if it is the absolute LR of the stimuli that matters or if it is the relative distance between stimuli LR and LR of

DOI: 10.1002/ejsp

the category label. Some authors have suggested a general principle: IAT effects depend on how well the stimuli instantiate the concepts that one wants to measure (De Houwer, 2002; Govan & Williams, 2004; Olson & Fazio, 2003; Steffens et al., 2008). According to this view superordinate ordinate level labels would be best represented by superordinate level stimuli and, thus, the present results could be the result of an increase ''match'' or ''similarity'' between LR of the labels and the LR of the stimuli.[13] This is a very interesting possibility that is in line with previous other research results and theorizations (e.g., De Houwer, Geldof, & De Bruycker, 2005). This view would also suggests that if the referent concept label is at a subordinate level instead of at the superordinate level (like in some IAT versions that juxtapose two individuals, e.g., McCain–Obama 2008 U.S. presidential elections IAT), then, the IAT effect would be larger with stimuli at the subordinate level. Our current data cannot disentangle these two alternative interpretations. In the present research we were interested in manipulating the stimulus LR while keeping the category of reference constant. For this reason currently both interpretations are plausible. This study, however, opens the possibility for such a clarification that will increase our understanding of the IAT paradigm. It is our belief that, at minimum, the IAT emphasis or bias on how well stimuli ''match'' category labels is not the same as the category-level bias in IAT processing, whereas the way that the IAT method was built emphasizes category level cognition.

The present research showed that LR of the stimuli (absolute or relative to the label LR) is an important dimension that has a significant impact onto the IAT paradigm. However these studies have some limitations that future research should try to address. For instance, as in any study where stimuli features are manipulated by means of pre-testing, one cannot be completely certain that the only dimension on which the stimuli differ is the intended one (i.e., here the LR). In fact, *post hoc* alternative explanations for the results and an alternative dimension (or combination of dimensions) on which the stimuli sets may differ can be proposed. However, the logical rational of the LR, its theoretical support, and the careful pre-testing on other known relevant dimensions (e.g., familiarity, evaluation) make LR a valid candidate to partly explain the discrepancies between IAT types.

CONCLUDING REMARKS

This report continues the exploration of the specific parameters or susceptibilities of the IAT effect. More work needs to be done modeling processes that may be at the core of the IAT paradigm. However, the link between the discrepancies using different IAT versions and LR opens an important avenue between social cognitive research and a large body of cognitive and neuropsychological research on LR. We may thus gain a better understanding of the IAT task and its true value in the assessment of implicit social cognition.

ACKNOWLEDGEMENTS

The authors are grateful to Jonathan Cook, Lew Goldberg, Bertram Malle, Myron Rothbart, and Gün Semin, for their valuable suggestions and comments on an earlier draft of this paper.

REFERENCES

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*, 186–197.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*, 596–612.

Banaji, M. R. (2001). Implicit attitude can be measured. In H. L. Rodriger, III & J. S. Nairne (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder. Science conference series* (pp. 117–150). Washington, DC: American Psychological Association.

[13]We would like to thank Jan De Houwer for suggesting this alternative interpretation.

DOI: 10.1002/ejsp

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

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

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology*, *42*, 192–212.

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*, 163–176.

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

Brendl, 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 & Social Psychology*, *81*, 760–773.

Brewer, M. B. (1988). A dual process model of impression formation. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Advances in social cognition* 1(pp. 1–36). Hillsdale, NJ: Erlbaum.

Brewer, M. B., Weber, J. G., & Carini, B. (1995). Person memory in intergroup contexts: Categorization versus individuation. *Journal of Personality and Social Psychology*, *69*, 29–40.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, *35*, 38–48.

Carlston, D. E., & Smith, E. R. (1996). Principles of mental representation. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 184–210). New York: Guilford Press.

Carver, C. S., & Scheier, M. F. (2002). Control processes and self-organization as complementary principles underlying behavior. *Personality and Social Psychology Review*, *6*, 304–315.

Castelli, L., Zogmaister, C., Smith, E. R., & Arcuri, L. (2004). On the automatic evaluation of social exemplars. *Journal of Personality and Social Psychology*, *86*, 373–387.

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: The Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–487.

Costa, A., Mahon, B., Savova, V., & Caramazza, A. (2003). Level of categorisation effect: A novel effect in the picture-word interference paradigm. *Language and Cognitive Processes*, *18*, 205–233.

Crisp, R. J., & Hewstone, M. (2007). Multiple social categorization. In M. P. Zanna (Ed.), *Advances in experimental social psychology* 39(pp. 163–254). Orlando, FL: Academic Press.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–170.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality & Social Psychology*, *81*(5), 800–814.

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

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

De Houwer, J. (2002). The implicit association test as a tool for studying dysfunctional associations in psychopathology: Strengths and limitations. *Journal of Behavior Therapy and Experimental Psychiatry*, *33*, 115–133.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes* C (pp. 179–194. New York, NY: Guilford Press.

De Houwer, J., Geldof, T., & De Bruycker, E. (2005). The Implicit association test as a general measure of similarity. *Canadian Journal of Experimental Psychology*, *59*, 228–239.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (in press) Implicit measures: A normative analysis and review. *Psychological Bulletin.*

De Wilde, E., Vanoverberghe, V., Storms, G., & De Boeck, P. (2003). The instantiation principle reevaluated. *Memory*, *11*, 533–548.

Devine, P. G. (2001). Implicit prejudice and stereotyping: How automatic are they? *Journal of Personality and Social Psychology*, *81*, 757–759.

Devos, T., & Ma, D. S. (2008). Is Kate Winslet more American than Lucy Liu? The impact of construal processes on the implicit ascription of a national identity. *British Journal of Social Psychology*, *47*, 191–215.

D'Lauro, C., Tanaka, J. W., & Curran, T. (2008). The preferred level of face categorization depends on discriminability. *Psychonomic Bulletin & Review*, *15*, 623–629.

Ellsworth, P. C., & Scherer, K. R. (2003). Appraisal processes in emotion. In R. J. Davidson, H. Goldsmith, & K. R. Scherer (Eds.), *Handbook of the affective sciences* (pp. 572–595). New York and Oxford: Oxford University Press.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. *Advances in Experimental Social Psychology*, *23*, 1–74.

Foroni, F. (2007). LR of stimuli effects in the IAT paradigm. Unpublished data.

Foroni, F., & Schubert, T. W. (2007). Levels of stimuli representation in the IAT paradigm and IAT's differential predicting validity. Unpublished data.

Frijda, N. H. (2000). The psychologist point of view. In M. Lewis, & J. H. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 59–74) New York, NY: Guilford Press.

DOI: 10.1002/ejsp

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, *40*, 357–365.

Greenwald, A. G. (2004). *Revised top 10 list of things wrong with the IAT.* Paper presented at the Attitudes Pre-conference of the 5th annual meeting of the Society of Personality and Social Psychology, Austin, TX.

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

Greenwald, A. G., 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: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Heit, E., & Barsalou, L. W. (1996). The instantiation principle in natural categories. *Memory*, *4*, 413–451.

Henson, R. N. A., Price, C. J., Rugg, M. D., Turner, R., & Friston, K. J. (2002). Detecting latency differences in event-related BOLD responses: Application to words versus non-words and and initial versus repeated face presentations. *NeuroImage*, *15*, 83–97.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*, 1369–1385.

Ito, T. A., & Urland, G. R. (2005). The influence of processing objectives on the perception of faces: An ERP study of race and gender perception. *Cognitive, Affective, and Behavioral Neuroscience*, *5*, 21–36.

Job, R., Rumiati, R., & Lotto, L. (1992). The picture superiority effect in categorization: Visual or semantic? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1019–1028.

John, O. P., Hampson, S. E., & Goldberg, L. R. (1991). The basic level in personality-trait hierarchies: Studies of trait use and accessibility in different contexts. *Journal of Personality and Social Psychology*, *60*, 348–361.

Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*, 1609–1624.

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the implicit association test: A diffusion-model analysis. *Journal of Personality and Social Psychology*, *93*, 353–368.

Kruglanski, A. W., & Orehek, E. (2007). Partitioning the domain of social inference: Dual mode and systems models and their alternatives. *Annual Review of Psychology*, *58*, 291–316.

Kunda, Z., & Oleson, K. C. (1995). Maintaining stereotypes in the face of disconfirmation: Constructing grounds for subtyping deviants. *Journal of Personality and Social Psychology*, *68*, 565–579.

Laeng, B., Zarrinpar, A., & Kosslyn, S. M. (2003). Do separate processes identify objects as exemplars versus members of basic-level categories? Evidence from hemispheric specialization. *Brain and Cognition*, *53*, 15–27.

Large, M.-E., Kiss, I., & McMullen, P. A. (1998). Electrophysiological correlates of objects categorization: Back to basics. *Cognitive Brain Research*, *20*, 415–426.

Marsolek, C. J. (1999). Dissociable neural subsystems underlie abstract and specific object recognition. *Psychological Science*, *10*, 111–118.

McGarty, C. (2006). Hierarchies and minority groups: The roles of salience, overlap, and background knowledge in selecting meaningful social categorizations from multiple alternatives. In R. J. Crisp, & M. Hewstone (Eds.), *Multiple social categorization: Processes, models and applications*. Hove, E. Sussex: Psychology Press.

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

Murphy, G., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts, & D. R. Shanks (Eds.), *Knowledge, concepts and categories. Studies in cognition* (pp. 93–131). Cambridge, MA: The MIT Press.

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning & Verbal Behavior*, *21*, 1–20.

Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language*, *31*, 237–260.

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

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

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin*, *31*, 166–180.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2006). The implicit association test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Social Psychology and the Unconscious: The Automaticity of Higher Mental Processes* (pp. 265–292). New York, NY, USA: Psychology Press.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, *14*, 636–639.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the implicit association test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653–667.

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

DOI: 10.1002/ejsp

Paivio, A., Walsh, M., & Bons, T. (1994). Concreteness effects on memory - when and why. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20*, 1196–1204.
Park, B., Ryan, C. S., & Judd, C. M. (1992). Role of meaningful subgroups in explaining differences in perceived variability for in-groups and out-groups. *Journal of Personality and Social Psychology*, *63*, 553–567.
Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional MRI study. *Journal of Neuroscience*, *16*, 5205–5215.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braere, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.
Rothbart, M., & John, O. P. (1985). Social categorization and behavioral episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues*, *41*, 81–104.
Sabsevitz, D. S., Medler, D. A., Seidenberg, M., & Binder, J. R. (2005). Modulation of the semantic system by word imageability. *Neuroimage*, *27*, 188–200.
Schwanenflugel, P. J., & Shoben, E. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *9*, 82–102.
Smith-McLallen, A., Johnson, B. T., Dovidio, J. F., & Pearson, A. R. (2006). Black and white: The role of color bias in implicit race bias. *Social Cognition*, *24*, 46–73.
Stangor, C., & Lange, J. (1994). Mental representations of social groups: Advances in understanding stereotypes and stereotyping. *Advances in Experimental Social Psychology*, *26*, 357–416.
Steffens, M. C., Kirschbaum, M., & Glados, P. (2008). Avoiding stimulus confounds in implicit association tests by using the concepts as stimuli. *British Journal of Social Psychology, 47*, 217–243.
Steffens, M. C., & Buchner, A. (2003). Implicit association test: Separating transsituationally stable and variable components of attitudes toward gay men. *Experimental Psychology*, *50*, 33–48.
Steffens, M. C., & Jelenec, P. (2007). On the generality of stimulus effects in implicit association tests: A two-factor model of reaction time differences in IATs. Manuscript in preparation.
Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the implicit association test. *Zeitschrift für Experimentelle Psychologie*, *48*, 123–134.
Steffens, M. C., Lichau, J., Still, Y., Jelenec, P., Anheuser, J., Goergens, N. K., et al. (2004). Individuum oder Gruppe, Exemplar oder Kategorie? Ein Zweifaktorenmodell zur Erklärung der Reaktionszeitunterschiede im Implicit Association Test (IAT). *Zeitschrift für Psychologie*, *212*, 57–65.
Steiger, J. H. (1980). Test for comparing elements of a correlation Matrix. *Psychological Bulletin*, *87*, 245–251.
Tanaka, J. W. (2001). The entry point of face recogntion: Evidence for face expertise. *Journal of Experimental Psychology: General*, *130*, 534–543.
Tanaka, J., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *NeuroReport*, *10*, 829–835.
Tarkiainen, A., Cornelissen, P. L., & Salmelin, R. (2002). Dynamics of visual feature analysis and object-level processing in face versus letter-string perception. *Brain*, *125*, 1125–1136.
Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*, 163–173.
Vitkovitch, M., & Underwood, G. (1991). Hemispheric differences in the processing of pictures of typical and atypical semantic category members. *Cortex*, *27*, 475–480.
Wentura, D., & Rothermund, K. (2007). Paradigms we live by. A plea for more basic research on the IAT. In B. Wittenbrink, & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 195–215). New York, NY: Guilford Press.
West, W. C., & Holcomb, P. J. (2000). Imaginal, semantic, and surface-level processing of concrete and abstract words: An electrophysiological investigation. *Journal of Cognitive Neuroscience*, *12*, 1024–1037.
Whatmough, C., Verret, L. D., Fung, D., & Chertkow, H. (2004). Common and contrasting areas of activation for abstract and concrete concepts: An H215O PET study. *Journal of Cognitive Neuroscience*, *16*, 1211–1226.
Zárate, M. A., Stoever, C. J., MacLin, M. K., & Chavez, C. J. (2008). Neurocognitive underpinnings of face perception: Further evidence of distinct person and group perception processes. *Journal of Personality and Social Psychology*, *94*, 108–115.

DOI: 10.1002/ejsp

APPENDIX

Stimulus material for all experiments. Dutch translations are in parentheses. Actual experimental material is in upper case. Material was adapted from Dasgupta and Greenwald (2001); Greenwald et al. (1998); Job et al. (1992); Nosek and Banaji (2001).

	Immigrant (ALLOCHTON)	Native (AUTOCHTON)	Good (GOEDE)	Bad (SLECHTE)
Exp. 1				
Picture-IAT	Six pictures of unknown dark skin-tone persons	Six pictures of unknown light skin-tone persons	Smiling (LACHEN), Pleasure (PLEZIER), Happy (BLIJ), Friendly (VRIENDELIJK), Cheerful (VROLIJK), Loving (LIEFDEVOL)	Horrible (VRESELIJK), Angry (BOOS), Destroy (VERWOESTEN), Brutal (BEESTACHTIG), Tragic (TRAGISCH), Poison (VERGIF)
Word-IAT	MAARTEN, SANDER HAROLD, PAUL JAN, PETER	TYRONE, MALIK, ASHAUN, LAMAR, ACHMET, MOHAMMED		
	BLACK AMERICAN	WHITE AMERICAN	GOOD	BAD
Exp. 2				
Picture-IAT	Six pictures of unknown African Americans	Six pictures of unknown European Americans	SMILING, PLEASURE, HAPPY, FRIENDLY, CHEERFUL, LOVING	HORRIBLE, ANGRY, DESTROY, BRUTAL, TRAGIC, POISON
Full/Name-IAT	MARTIN LUTHER KING, DENZEL WASHINGTON, EDDIE MURPHY, O.J. SIMPSON, MIKE TYSON, LOUIS FARRAKHAN	JOHN F.KENNEDY,TOM CRUISE,JIM CARREY,TED BUNDY,CHARLES MANSON,AL CAPONE		
Word-IAT	RASHAUN, LAMAR, LEROY, THEO, TYRONE, MALIK	BRANDON, MATTHEW, ANDREW, PETER, JUSTIN, STEPHEN		
Whole/ Category-IAT	BLACKS, AFRICAN AMERICANS, OPPRESSED and three Black symbols: black smile-face, black-square, black-triangle	WHITES, EUROPEAN AMERICANS, DOMINANT and three White symbols: white smile-face, white-square, white-triangle		
	Weapons (WAPENS)	Musical instrument (MUZIEKINSTRUMENTEN)	Good (GOEDE)	Bad (SLECHTE)
Exp. 3				
Subordinate-IAT	Beretta (BERETTA), Carabine (KARABIJN), Mine (MIJN), Knife (MES)	Drum (TROMMEL), Trombone (SCHUIFTROMPET), Violin (VIOOL), Guitar (GITAAR)	Smiling (LACHEN), Pleasure (PLEZIER), Happy (BLIJ), Friendly (VRIENDELIJK)	Angry (BOOS), Destroy (VERWOESTEN), Tragic (TRAGISCH), Poison (VERGIF)
Basic-IAT	Handgun (PISTOOL), Shotgun (GEWEER), Bomb (BOM), Blade (STEEK WAPEN)	Percussions (SLAGINSTRUMENT), Winds (BLAASINSTRUMENT), Ironing Instruments (STRIJKINSTRUMENT), Stringed Instruments (TOKKELINSTRUMENT)		
	Foreigners (ALLOCHTON)	Native (AUTOCHTON)	Good (GOEDE)	Bad (SLECHTE)
Exp. 4				
Individuals-IAT	Six pictures of unknown dark skin-tone persons	Six pictures of unknown light skin-tone persons	As in Experiment 1	As in Experiment 1
Group-IAT	Six pictures of groups of unknown dark skin-tone persons	Six pictures of groups of unknown light skin-tone persons		