Research Article

Positive Associations Primacy in the IAT

A Many-Facet Rasch Measurement Analysis

Pasquale Anselmi, Michelangelo Vianello, and Egidio Robusto

Department of Applied Psychology, University of Padua, Italy

Abstract. Two studies investigated the different contribution of positive and negative associations to the size of the Implicit Association Test (IAT) effect. A Many-Facet Rasch Measurement analysis was applied for the purpose. Across different IATs (Race and Weight) and different groups of respondents (White, Normal weight, and Obese people) we observed that positive words increase the IAT effect whereas negative words tend to decrease it. Results suggest that the IAT is influenced by a positive associations primacy effect. As a consequence, we argue that researchers should be careful when interpreting IAT effects as a measure of implicit prejudice.

Keywords: positive associations primacy, Implicit Association Test, implicit prejudice, Rasch models

The present article investigates the different contribution of positive and negative associations to the size of the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998) effect for measures of prejudice. According to the literature, the "positive associations primacy" effect represents the superiority of the associations of the concepts with the attribute *Good*, relative to the associations with *Bad*. The prominent role of positive associations on implicit measures has been observed by Sriram and Greenwald (2009) in the Brief Implicit Association Test (BIAT). In the BIAT, respondents indicate whether the stimuli (presented one at a time) belong to one of the two concepts (e.g., *Pepsi* or *Good*) or not. The BIAT effect is computed by comparing two blocks of trials that share the attribute (e.g., *Good*) but differ in the concepts (e.g., *Pepsi* or *Coke*). The authors found that, when *Good* (and not *Bad*) was the attribute shared in the two blocks, the measures were internally consistent, predictively valid, and strongly convergent with parallel self-report measures and standard IAT measures. Analogous results have been found by Bar-Anan, Nosek, and Vianello (2009) in the Sorting Paired Features task. In the present article, we investigate whether the positive associations primacy influences IAT effects as well. The IAT is the most popular implicit measure of attitude and prejudice (see Lane, Banaji, Nosek, & Greenwald, 2007 for a review). It assesses the relative strength of the associations between two concepts (e.g., *White People* and *Black People*) and two attributes (e.g., *Good* and *Bad*).

Previous research on the influence of stimuli has found that the IAT effect is influenced by (a) *the valence of concept stimuli*: Govan and Williams (2004) found that using good insects and bad flowers as stimuli changes the direction of the IAT effect; (b) *multiply categorizable social targets*: having participants focusing on one category membership (e.g., black) rather than another (e.g., young), Mitchell, Nosek, and Banaji (2003) obtained different IAT effects; and (c) both targets' and attributes' *cross-category associations* (e.g., use of an item such as "warm" or "aggressive" in a gender IAT can induce confusion on their belongings; Bluemke & Friese, 2006; Steffens & Plewe, 2001). In summary, previous literature pointed out that any time the stimuli, the context, or the category labels used in an IAT stress or accentuate one cross-category association rather than another, the IAT effect changes. Yet, no study, to our knowledge, has tried to evaluate independently the different strengths of the positive and negative associations that are present in a standard IAT. Furthermore, we conducted our studies on IAT procedures that use well-constructed, validated, and widely used sets of stimuli.

In this article, the analysis is conducted by identifying the category of stimuli whose speed of categorization changes the most between the two associative conditions. Consider, for example, a measure of implicit preference for whites relative to blacks in a Race IAT. We argue that if positive (rather than negative) words are categorized more quickly in the condition White-Good/Black-Bad than in the condition Black-Good/White-Bad, then they are the stimuli that mostly contribute to the IAT effect and that a positive associations primacy can be observed. By drawing on Sriram and Greenwald (2009), and Bar-Anan et al. (2009), we hypothesized that positive associations increase the IAT effect whereas negative associations decrease it. We tested this prediction in Study 1. In Study 2 we investigated the influence of positive and negative associations when respondents who belong to one of the two groups represented by the IAT do not implicitly prefer their own group,

DOI: 10.1027/1618-3169/a000106

as well as when they belong to a third – unrepresented – group. The analyses were performed on data collected on the Project Implicit Italian demonstration website (https://implicit.harvard.edu/implicit/italy). The data were analyzed through a Many-Facet Rasch Measurement (MFRM), a probabilistic model belonging to the family of Rasch models.

Many-Facet Rasch Measurement

The Simple Logistic Model (Rasch, 1960/1980) is the basic Rasch model for the transformation of ordinal observations into linear measures. It is meant for dichotomous data and expresses, according to a logistic distribution, the probability of a response as a function of the respondent's ability and the item's difficulty. The more able the respondent (resp. less), and the less difficult the item (resp. more), then the greater the probability of a correct response[1] (resp. incorrect).

Among the many extensions to the basic model (see, e.g., Andrich, 1978, 2005; Cristante & Robusto, 1999, 2007; Fischer & Molenaar, 1995; Masters, 1982), the MFRM (Linacre, 1989) particularly suits the analysis of the IAT, since it allows the inclusion of other sources of systematic variability (facets), in addition to respondents' ability and item difficulty, accounting for the likelihood of a response. Consequently, in a MFRM analysis of the IAT, we introduce a facet accounting for the associative condition the item is presented in. Furthermore, a parameter concerning the response scale $k = \{0, 1, \ldots, m\}$ given by a discretization of response times is considered. Therefore, the model takes on the following form:

$$\ln \frac{P_{nijk}}{P_{nij(k-1)}} = \beta_n - \delta_i - \gamma_j - \tau_k, \tag{1}$$

where P_{nijk} (resp. $P_{nij(k-1)}$) is the probability that respondent n would respond to stimulus i in condition j at speed k (resp. $k-1$); β_n is the ability (speed) of respondent n; δ_i is the difficulty (speed of categorization) of stimulus i; γ_j is the ease of condition j; τ_k is the impediment of response k relative to $k-1$.

Respondents, stimuli, and conditions are facets. As far as responses to an IAT are concerned, the model specifies that the probability that a respondent n gives a response k rather than $k-1$ to stimulus i in condition j depends on the additive effects of the speed of the respondent (β_n), the speed of categorization of the stimulus (δ_i), the ease of the condition (γ_j), and the impediment of giving the response k rather than $k-1$ (τ_k).

We have used the MFRM to analyze the IAT because of the many benefits that derive from its general and specific features (see also Robusto, Cristante, & Vianello, 2008). With respect to the former, we underline that: (a) being a Rasch model, the MFRM shares its characteristics of stochastic independence, specific objectivity, linearity, and measurement unit (for a discussion of these properties see Andrich, 1988; Bond & Fox, 2001; Fischer & Molenaar, 1995); (b) all facets are located on the same latent trait, allowing comparisons between their elements; (c) if the data fit the model, the elements of each facet are estimated independently of one another (Linacre & Wright, 2002); and (d) fit statistics provide detailed tests of the adherence of data to the model and give an important contribution to the interpretation of the results. With respect to the latter, the MFRM provides bias analysis, that is, the analysis of the interactions between elements of different facets (see Linacre, 2009a, for details). In particular, the Differential Stimulus Functioning (DSF) shows whether the speed of categorization of the stimuli changes according to the associative condition they are presented in.

For a detailed description of indexes and statistics used in MFRM, we refer the interested reader to Myford and Wolfe (2003), and to Vianello and Robusto (2010) for their meaning and use in implicit measures. Listed below are some important indexes for the analyses that follow.

1. The *Infit* and *Outfit* statistics evaluate the fit of the data to the model. If in the range from .5 to 2, they express good fit (Linacre, 2009a).
2. The *Separation Ratio* (G) represents a measure of the spread of the estimates relative to their precision. It ranges from 1 to infinity. $G = 2$, for instance, means that the dispersion in the measures of the elements in the facet is two times greater than the imprecision in their estimates (Wright, 1996).
3. The *Separation Reliability* (R) shows how reproducibly different the measures are. It ranges between 0 and 1. If R is close to 1, there is a high probability that the elements of the facet with high measure estimates actually have higher measures than those with low measure estimates (Linacre, 2009a). If $R < .5$, it is likely that the value of G is completely due to measurement error.
4. The Fixed (all-same) *chi-square* tests the hypothesis that the elements of a facet have the same logit in relation to the measurement error (*SE*).

Study 1

Study 1 assesses implicit prejudice toward black people (vs. white people) in a group of white respondents. The main goal is to investigate whether the well-known tendency of white respondents to prefer their own racial group (see, e.g., Nosek et al., 2007) is influenced by a positive associations primacy.

[1] Traditional applications of Rasch models relate to ability tests and, as a consequence, they use a language that refers to "able" or "not able" respondents, and "easy" or "difficult" items.

Method

Respondents

From the 2,332 respondents that completed the Race IAT in the period from March 2006 to September 2008, 946 provided complete data, and 880 provided data that were interpretable according to the data reduction criteria for Internet research suggested by Nosek, Banaji, and Greenwald (2002). The 880 respondents reported to be white. Their mean age was 28.29 (*SD* = 9.74, range from 12 to 70), and 43% were female.

Materials and Procedure

The Race IAT used the category labels *White People*, *Black People*, *Good*, and *Bad*. The stimuli consisted of 12 morphed faces of white and black people (Nosek et al., 2007) and 16 words with positive and negative meanings. The seven-block IAT employed for the procedure (Greenwald, Nosek, & Banaji, 2003) presented the stimuli in the center of the computer screen in an alternating fashion, and respondents had to categorize them by pressing, as quickly and accurately as possible, the response key *E* or *I*. A red cross appeared in the event of a mistake. The order of the associative conditions was counterbalanced across the respondents. Demographic data were assessed by a questionnaire that preceded or followed the IAT in a counterbalanced order.

MFRM Analysis

The MFRM analysis was performed using the computer program FACETS 3.66.0 (Linacre, 2009a). The three-facet model (respondents, conditions, stimuli) for polytomous responses that is formally represented in Equation 1 was applied to the responses given to the stimuli in the two associative conditions. Estimates of conditions and stimuli were centered around zero. DSF was investigated by measuring the interaction between facets stimuli and conditions.

Responses smaller than 300 ms and greater than 10,000 ms were excluded from the analysis, and response times were discretized according to the terziles computed on the $i \times j \times k$ complete data matrix, where $i = 880$ is the number of respondents, $j = 2$ is the number of conditions, and $k = 28$ is the number of stimuli. Therefore, our dependent variable (DV) identifies short (DV = 3), medium (DV = 2), and long (DV = 1) response times, respectively.

To test the unidimensionality of the Rasch measure, the stimuli representing the categories *White People* and *Good* have been separated from those representing the categories *Black People* and *Bad*, and person estimates β have been computed on the two resulting subsets. A Pearson's correlation between the β was computed, as well as the latent (disattenuated) correlation $lr = r/\sqrt{R_{WG} \times R_{BB}}$, where R_{WG} and R_{BB} are the reliabilities of the White-Good and Black-Bad subsets, respectively (see Linacre, 2009b). If lr is close to 1, the measure is substantively unidimensional.

Results

The correlation between the β estimates of the White-Good and Black-Bad subsets suggests that the implicit measure of racial attitude is unidimensional ($r = .88$, $p < .01$; $R_{WG} = .89$, $R_{BB} = .89$; $lr = .99$).

Fit indexes were excellent for the conditions ($.97 \leq$ Infit/Outfit ≤ 1.05) and the stimuli ($.90 \leq$ Infit/Outfit ≤ 1.10). Only 12 of the 880 respondents (1.36%) have Outfit or Infit above 2.

Differences in respondents' speeds were significant (β ranges from -3.94 to 3.52; $\chi^2_{(879)} = 11586.9$, $p < .001$), reproducible ($R = .94$), and four times greater than the imprecision in their estimates ($G = 2.89$). The condition White-Good and Black-Bad (WGBB) was easier than the condition Black-Good and White-Bad (BGWB; $\gamma_{WGBB} = .53$; $\gamma_{BGWB} = -.53$; $\chi^2_{(1)} = 6084.6$, $p < .001$; $R = 1$; $G = 55.15$), meaning that, at the group level, white respondents implicitly preferred white people to black people. The distance of 1.06 between the locations of conditions WGBB and BGWB on the latent trait represents the size in logits of the IAT effect.

Table 1 provides overall (i.e., across the two conditions) and local (i.e., in each condition) logits of each stimulus. The logit of a stimulus in a condition is computed by adding a bias measure to the overall speed of categorization of the same stimulus (δ) if the response to the latter is faster in a given condition than overall and by subtracting it if the response to the stimulus is slower.

The stimuli were categorized with different speeds (δ ranges from $-.62$ to .79; $\chi^2_{(27)} = 3382.2$, $p < .001$; $R = .99$; $G = 11.13$). By considering the local logits of the stimuli, we found that the overall speed of categorization of some stimuli significantly changed according to the condition they were presented in. Table 1 provides the DSF across conditions. The t values in the table test the null hypothesis that the difference between the local logits of the stimuli is equal to zero. The stimuli that are associated with a significant positive t increase the size of the overall IAT effect. The stimuli that are associated with a significant negative t reduce the overall IAT effect. Compared with their overall speed of categorization, seven out of eight positive words (laughter, pleasure, glory, peace, happy, joy, love) were categorized faster in the condition White-Good/Black-Bad and slower in the condition Black-Good/White-Bad. If these seven positive words were only considered in a new MFRM analysis, the size of the IAT effect would *increase* from 1.06 to 1.36. *Evil* was the only negative word that has been categorized faster in the condition White-Good/Black-Bad and slower in the condition Black-Good/White-Bad (hence it was the only negative word that increased the IAT effect). Compared with their overall speed of categorization, three out of eight negative words (agony, failure, despicable) were categorized faster in the condition Black-Good/White-Bad and slower in the condition White-Good/Black-Bad. If these three negative words were only considered in a new MFRM analysis, the size of the IAT effect would *decrease* from 1.06 to .82. With the exception of the black face *Bf6*, the overall speed of categorization

Table 1. Speed of categorization of the stimuli across the conditions and in each condition of the Race IAT (880 respondents)

	Across conditions			White-Good Black-Bad			Black-Good White-Bad				
Stimulus	*ORS*	δ	*SE*	*ORS*	*LM*	*SE*	*ORS*	*LM*	*SE*	*t*	*df*
Laughter	3,242	−.36	.04	1,886	−.23	.05	1,356	−.51	.05	3.74***	1,758
Pleasure	3,320	−.26	.04	1,928	−.13	.05	1,392	−.40	.05	3.75***	1,758
Glory	3,342	−.23	.04	1,934	−.11	.05	1,408	−.36	.05	3.35***	1,758
Peace	3,454	−.09	.04	1,990	.03	.05	1,464	−.20	.05	3.20**	1,758
Happy	3,396	−.16	.04	1,955	−.06	.05	1,441	−.26	.05	2.79**	1,757
Joy	3,312	−.27	.04	1,909	−.18	.05	1,403	−.37	.05	2.63**	1,757
Love	3,427	−.12	.04	1,967	−.03	.05	1,460	−.22	.05	2.55*	1,758
Evil	3,289	−.30	.04	1,889	−.22	.05	1,400	−.38	.05	2.20*	1,757
Wf6	3,795	.35	.04	2,137	.43	.05	1,658	.29	.05	1.96	1,758
Wonder	3,070	−.58	.04	1,763	−.52	.05	1,307	−.66	.06	1.80	1,757
Wf4	3,683	.21	.04	2,080	.27	.05	1,603	.15	.05	1.60	1,758
Wf5	3,701	.23	.04	2,084	.28	.05	1,617	.19	.05	1.22	1,757
Bf1	4,121	.79	.04	2,262	.83	.06	1,859	.77	.05	0.76	1,758
Horrible	3,323	−.25	.04	1,884	−.24	.05	1,439	−.27	.05	0.41	1,757
Wf1	3,598	.10	.04	2,014	.09	.05	1,584	.10	.05	−0.20	1,758
Bf3	4,047	.69	.04	2,217	.67	.06	1,830	.70	.05	−0.31	1,758
Terrible	3,381	−.18	.04	1,898	−.20	.05	1,483	−.15	.05	−0.76	1,757
Wf2	3,654	.17	.04	2,025	.12	.05	1,629	.21	.05	−1.35	1,758
Bf2	3,915	.51	.04	2,142	.44	.05	1,773	.56	.05	−1.62	1,758
Bf5	4,089	.75	.04	2,218	.68	.06	1,871	.80	.05	−1.59	1,758
Nasty	3,288	−.30	.04	1,835	−.36	.05	1,453	−.23	.05	−1.71	1,758
Pain	3,259	−.33	.04	1,817	−.39	.05	1,442	−.26	.05	−1.79	1,756
Wf3	3,730	.27	.04	2,055	.20	.05	1,675	.33	.05	−1.84	1,757
Bf4	3,980	.60	.04	2,167	.52	.06	1,813	.66	.05	−1.90	1,758
Bf6	3,731	.27	.04	2,029	.13	.05	1,702	.39	.05	−3.69***	1,758
Despicable	3,064	−.59	.04	1,671	−.75	.05	1,393	−.40	.05	−4.82***	1,758
Failure	3,293	−.29	.04	1,786	−.47	.05	1,507	−.09	.05	−5.42***	1,758
Agony	3,044	−.62	.04	1,652	−.80	.05	1,392	−.40	.05	−5.41***	1,758

Note. *ORS* = observed raw scores; *LM* = local measure; Wf = white face; Bf = black face. The *t* values test the hypothesis that the difference between the local δ is equal to zero. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

of faces did not change significantly between the two associative conditions.

Discussion

In line with previous literature, this study showed that white respondents implicitly preferred whites to blacks. Furthermore, we found evidence of a strong positive associations primacy, suggesting that the IAT effect analyzed in this study would be better interpreted as implicit ingroup favoritism rather than implicit prejudice.

In this study the respondents belonged to one of the two groups recalled by the IAT category labels. Popa-Roch and Delmas (2010) recently found that, in order to deal with the speed-accuracy constraints of the task, respondents to the IAT use self-relevant heuristics related to their membership to one of the two concept categories. Hence, a possible interpretation of our findings might trace back the positive associations primacy effect we observed to the effect of group membership, rather than preference. To shed light on this concern, in the following study we explore if a positive associations primacy is observed when the respondents do not implicitly prefer the group they belong to, as well as when they do not belong to one of the groups represented by the category labels of the IAT procedure.

Study 2

In this study we are interested in exploring whether a positive associations primacy influences the IAT effect when respondents belong to one of the two groups represented in the IAT and do not implicitly prefer their own group, as well as when they belong to a third – not represented – group. To this end, we measured implicit attitudes toward thin and fat people both in normal weight and obese respondents (Nosek et al., 2007).

In a pilot study, we tested the assumption that normal weight respondents actually form a third – not represented – group. We asked participants to explicitly declare their identification with both groups (How much do you identify with people who are fat? – How much do you identify with people who are thin?) Then we divided our sample into normal weight and obese respondents according to their Body Mass Index (BMI = kg/m^2, see, e.g., Eknoyan, 2008). Results suggest that normal weight respondents ($18.5 \leq$ BMI < 25, N = 165) identify with thin people significantly less than thin respondents (BMI < 18.5, N = 36; $t_{(171)} = -4.91$, $p < .001$), and that they identify with obese people significantly less than obese respondents (BMI $\geq$ 30, N = 13; $t_{(161)} = -5.53$, $p < .001$). The sample of this pilot study does not differ from that of the following study with respect to age and gender.

Method

Respondents

From the 589 respondents that completed a Weight IAT in the period from March 2006 to September 2008, 547 provided complete data, and 510 provided interpretable data (see Nosek et al., 2002). Within the whole sample, normal weight (N = 331; mean age 26.16 ± 8.07; range from 12 to 56; 64.5% female) and obese (N = 43; mean age 34.30 ± 10.01; range from 17 to 57; 58.5% female) respondents were selected for the analyses.

Materials and Procedure

The Weight IAT used the category labels *Thin People*, *Fat People*, *Good*, and *Bad*. The stimuli consisted of 10 morphed faces of thin and fat people (Nosek et al., 2007) and 16 words with positive and negative meanings. The procedure is the same as in Study 1.

MFRM Analysis

A MFRM analysis was performed separately on normal weight and obese respondents. The analysis strategy and the preliminary data reduction procedure are the same as Study 1. The unidimensionality of the measure was tested on aggregate data of all respondents.

Results

The correlation between the two sets of β parameter estimates (positive and negative words) suggested that the latent trait is unidimensional ($r = .78, p < .01; R_p = .81, R_n = .82; lr = .96$).

Normal Weight Respondents

Fit indexes were good for the conditions ($.99 \leq$ Infit/Outfit ≤ 1.01) and the stimuli ($.83 \leq$ Infit/Outfit ≤ 1.15) facets. Only 1 out of 331 respondents had Infit and Outfit values greater than 2. Differences in respondents' speeds were significant (β ranges from -2.45 to 2.74; $\chi^2_{(330)} = 2343.4$, $p < .001$), reproducible (R = .89), and almost three times greater than the imprecision in their estimates (G = 2.89). The condition Thin-Good/Fat-Bad (TGFB) was easier than the condition Fat-Good/Thin-Bad (FGTB; $\gamma_{\mathrm{TGFB}} = .39$; $\gamma_{\mathrm{FGTB}} = -.39$; $\chi^2_{(1)} = 773.7$, $p < .001$; $R = 1$; $G = 19.67$), meaning that, at the group level, normal weight respondents implicitly preferred thin people to fat people. The stimuli were categorized with different speeds (δ ranges from $-.38$ to .31, $\chi^2_{(15)} = 114.2$, $p < .001$; $R = .86$; $G = 2.50$). Compared with their overall speed of categorization, three out of eight positive words (joy, pleasure, happy) were categorized faster in the condition Thin-Good/Fat-Bad and slower in the condition Fat-Good/Thin-Bad, whereas two out of eight negative words (agony, terrible) were categorized faster in the condition Fat-Good/Thin-Bad and slower in the condition Thin-Good/Fat-Bad (see Table 2). Hence, in normal weight respondents, positive words tended to increase the IAT effect and negative words tended to decrease it.

Obese Respondents

Fit indexes were good for conditions ($.99 \leq$ Infit/Outfit ≤ 1.01) and stimuli ($.64 \leq$ Infit/Outfit ≤ 1.17) facets, and they were acceptable for the respondents ($.67 \leq$ Infit/Outfit ≤ 1.64) facet. Differences in respondents' speeds were significant (β ranges from -3.23 to 1.55; $\chi^2_{(42)} = 311.3$, $p < .001$), reproducible (R = .89), and almost three times greater than the imprecision in their estimates (G = 2.89). For obese respondents as well the condition Thin-Good/Fat-Bad was easier than the condition Fat-Good/Thin-Bad ($\gamma_{\mathrm{TGFB}} = .26$; $\gamma_{\mathrm{FGTB}} = -.26$; $\chi^2_{(1)} = 41.1$, $p < .001$; $R = .95$; $G = 4.42$). At the group level, obese respondents implicitly preferred thin people to fat people. The stimuli were categorized with comparable speeds (δ ranges from $-.54$ to .54, $\chi^2_{(15)} = 31.6$, $p < .01$; $R = .51$; $G = 1.03$). Compared with their overall speed of categorization, two out of eight positive words (happy, pleasure) were categorized faster in the condition Thin-Good/Fat-Bad and slower in the condition Fat-Good/Thin-Bad. The speed of categorization of the negative word *failure* increased in the condition Fat-Good/Thin-Bad and decreased in the condition Thin-Good/Fat-Bad (see Table 2). Therefore, the implicit preference toward thin people in obese respondents was mostly guided by two positive words.

Discussion

This study highlighted that both normal weight and obese respondents implicitly preferred thin people to fat people. In both samples of respondents, the preference for thin people was mainly based on a subset of positive words. Negative words sometimes lowered the IAT effect. In line with Study 1, we found strong evidence of a positive associations primacy effect. Furthermore, this study showed that positive associations primacies do not depend on respondents'

Table 2. Speed of categorization of the positive and negative words across the conditions and in each condition of the weight IAT

	Across conditions			Thin-Good Fat-Bad			Fat-Good Thin-Bad				
Word	*ORS*	δ	*SE*	*ORS*	*LM*	*SE*	*ORS*	*LM*	*SE*	*t*	*df*
Normal weight respondents (N = 331)											
Joy	1,380	.12	.06	788	.36	.08	592	−.09	.08	3.91**	660
Pleasure	1,350	.03	.06	763	.19	.08	587	−.13	.08	2.74**	660
Happy	1,363	.07	.06	765	.20	.08	598	−.06	.08	2.25*	660
Love	1,435	.31	.06	793	.39	.08	642	.23	.08	1.40	659
Peace	1,379	.13	.06	758	.16	.08	621	.10	.08	0.50	659
Horrible	1,359	.06	.06	746	.08	.08	613	.04	.08	0.34	660
Laughter	1,334	−.02	.06	731	−.02	.08	603	−.02	.08	0.04	660
Glory	1,357	.05	.06	741	.04	.08	616	.06	.08	−0.11	660
Nasty	1,327	−.04	.06	725	−.06	.08	602	−.03	.08	−0.24	660
Wonder	1,220	−.38	.06	670	−.40	.08	550	−.35	.08	−0.37	659
Despicable	1,263	−.24	.06	691	−.27	.08	572	−.21	.08	−0.49	659
Evil	1,356	.05	.06	730	−.03	.08	626	.12	.08	−1.29	660
Failure	1,348	.02	.06	725	−.06	.08	623	.10	.08	−1.41	660
Pain	1,311	−.09	.06	701	−.19	.08	610	.02	.08	−1.91	659
Terrible	1,357	.05	.06	721	−.08	.08	636	.18	.08	−2.36*	660
Agony	1,303	−.12	.06	688	−.29	.08	615	.05	.08	−3.03**	660
Obese respondents (N = 43)											
Happy	163	.19	.16	97	.71	.23	66	−.39	.25	3.23**	84
Pleasure	158	.07	.16	92	.45	.22	66	−.39	.25	2.51*	84
Peace	159	.09	.16	88	.26	.22	71	−.09	.24	1.07	84
Horrible	158	.07	.16	87	.21	.22	71	−.09	.24	0.92	84
Glory	159	.09	.16	87	.21	.22	72	−.04	.23	0.75	84
Despicable	153	−.06	.16	82	−.04	.22	71	−.09	.24	0.16	84
Laughter	155	−.01	.16	83	.01	.22	72	−.04	.23	0.15	84
Love	177	.54	.16	94	.55	.22	83	.53	.22	0.07	84
Nasty	156	.01	.16	82	−.04	.22	74	.07	.23	−0.34	84
Pain	151	−.12	.16	79	−.19	.22	72	−.04	.23	−0.47	84
Wonder	136	−.54	.17	71	−.61	.24	65	−.45	.25	−0.47	84
Joy	141	−.39	.17	72	−.55	.23	69	−.21	.24	−1.04	84
Agony	158	.07	.16	80	−.14	.22	78	.28	.23	−1.31	84
Terrible	158	.07	.16	79	−.19	.22	79	.33	.22	−1.63	84
Evil	149	−.17	.16	74	−.45	.23	75	.13	.23	−1.76	84
Failure	159	.09	.16	78	−.24	.23	81	.43	.22	−2.11*	84

Note. *ORS* = observed raw scores; *LM* = local measure. The t values test the hypothesis that the difference between the local δ is equal to zero. $^{*}p < .05$. $^{**}p < .01$.

membership of one of the IAT target categories, nor do they depend on implicit preference for the ingroup.

General Discussion

This article investigated the importance of positive and negative words in influencing the size of the IAT effect in measures of prejudice and attitude. Across two studies we found that responses to negative words decreased the IAT effect, whereas responses to positive words increased it. Positive associations seem to have a prominent role in IAT measures of attitude and prejudice. They drive the IAT effect both when respondents do not belong to one of the two groups represented in the IAT (normal weight respondents in Study 2), as well as when they prefer the group they belong to (white respondents in Study 1) or when they prefer another group (obese respondents in Study 2). Hence, the positive associations primacy effect cannot be fully accounted for by the positivity that characterizes the self-relevant heuristics that participants tend to use in order to deal with the speed-accuracy constraints of the task (Popa-Roch & Delmas, 2010).

Theoretical and Methodological Implications

This article provides initial evidence that positive associations are more important than negative associations in determining the size of the IAT effect.

Our results have an important and direct theoretical implication for the operational definition of implicit prejudice. Previous work used the IAT effect as a measure of implicit prejudice toward a number of different social groups or minorities (see, e.g., Hugenberg & Bodenhausen, 2003; Rudman, Ashmore, & Gary, 2001; Rudman, Greenwald, Mellott, & Schwartz, 1999), and evidence is also available suggesting that the resulting measure of implicit prejudice has good convergent and discriminant validity (see, e.g., Gawronski, 2002). However, our results limit the generalizability of this interpretation (see also van Ravenzwaaij, van der Maas, & Wagenmakers, 2010). Given that the construct of prejudice has by definition a negative component, without accurate knowledge of the meaning of the IAT effect in a specific social group we may have misleadingly interpreted the effect found, for example, in Study 1. In this case, a Race IAT, which is typically used to measure racial prejudice, provided a measure of implicit ingroup favoritism, rather than outgroup derogation or prejudice.

Furthermore, our studies showed a way in which the IAT effect can be meaningfully decomposed into positive and negative associations. This result contributes to the debate on the relative nature of the IAT effect. Indeed, the IAT is thought of as providing a relative implicit measure of an object (e.g., blacks) compared with another object (e.g., whites). By design, the IAT procedure cannot yield completely independent responses to one of the four categories of stimuli involved in the technique. Furthermore, the relative failure of previous attempts to separate its components by analyzing subsets of responses (Nosek, Greenwald, & Banaji, 2005) has led researchers to accept the relative nature of the IAT as an unquestionable truth. Yet, among the many advantages of the IAT (within which its psychometric properties play a prominent role), the relative nature of its effect is perhaps its main limit. As a result, many other implicit techniques have been developed and presented to overcome this limitation. Some of them are variants of the IAT (e.g., the Single Category IAT, Karpinski & Steinman, 2006, or the Single Target IAT, Bluemke & Friese, 2008; Wigboldus, Holland, & van Knippenberg, 2004) whereas others are completely different (e.g., the Go/No-Go Association Task, Nosek & Banaji, 2001, and the SPF task, Bar-Anan et al., 2009). Unfortunately, more work is needed on these alternatives before they can achieve the same level of validity and reliability as the IAT. Hence this situation results in researchers being compelled to a compromise between the psychometric properties of the implicit techniques they use and the range of research questions that can be addressed, only because the IAT does not always seem to fit the research question at hand (as was previously the case for unipolar constructs). Our opinion is that the relative nature of the IAT effect should not be seen by researchers as an insurmountable limit. While it is true that complete independence between the different components will never be achieved, it is also true that in this article we have been able to identify the separate contributions of good and bad words to the overall IAT effect, highlighting which was dominant in each measure. Indeed, by means of the Rasch model we obtained formally independent and fully comparable measures of each single association. This is a preliminary result that should be thoroughly investigated before claiming that the IAT effect can be effectively decomposed. Yet, if a decomposition was somehow possible, as our studies suggest, then researchers would no longer be compelled to reach a compromise between the validity and reliability of their measurement and the research questions that can be addressed. We hope that future research will be able to clarify this important point.

Finally, the article proposes a model that well suits the analysis of the IAT. The MFRM provides many advantages over alternative scoring procedures. One first point to consider is that, when the data fit the model, the MFRM produces, for each element of each facet, a measure that does not depend on the elements of the other facets. The MFRM also provides detailed fit indexes of each element (e.g., stimuli, respondents, and conditions of association), of their spread along the continuum of possible scores (*G*), and of the reproducibility of their rank ordering (*R*). Yet, the model presented in the article specifies only a respondent parameter β_n (see Equation 1) and can be considered as a unidimensional item component model.[2] With this model it is implicitly assumed that the ease of the two associative conditions, and therefore the size of the IAT effect, does not differ across the respondents. Since we were interested in group-level effects, interindividual differences have not been considered in the present work. Otherwise, we could have either computed an interaction analysis between respondents and conditions or we could have modeled individual differences by specifying a multidimensional facet model (Rost & Carstensen, 2002). Such a model allows a speed parameter β_{nj} to be estimated for each respondent n and each associative condition j.

Limitations and Future Research

We believe that our results show converging evidence that the size of the IAT effect mostly reflects the influence of positive associations. However, the conditions for positive associations primacy to occur still have to be pinned down. Also, the size of the IAT effect is not the only criterion to determine whether positive associations are more important than negative. Other aspects of measurement might be taken into account in the future, such as validity, reliability, and predictive validity. We found evidence of a positive associations primacy influencing the IAT, but we lack information about the causes of this effect. Future experimental studies that manipulate valence might be able to shed light on this issue. For example, there is some evidence that negative stimuli are processed faster and more accurately than positive stimuli (Dijksterhuis & Aarts, 2003, but see Labiouse, 2004 for contrasting evidence). It might be worth investigating whether the relationship between valence, processing

[2] We thank an anonymous reviewer for their useful comments on the linkage between MFRM models and item component models, and on the implications that follow.

speed, and accuracy is at the origin of the positive associations primacy effect in the IAT. The relationship with the "reverse priming effect" (slower responses to evaluatively matched rather than mismatched items) might also be worth investigating: Glaser and Banaji (1999) found that when primes were extreme in valence (extremely positive or extremely negative), they reduced the accessibility of a subsequent congruent target. Reverse priming effects might be at the origin of positive associations primacies. Lastly, it might also be interesting to investigate experimentally whether there is an interaction between the positive associations primacy effect in the IAT and the effects of stimuli's cross-category associations (Steffens & Plewe, 2001).

One methodological limit regards Study 2, in which the number of obese participants was rather small. The joint maximum likelihood estimates (Wright & Panchapakesan, 1969) of model parameters, which are computed in a MFRM analysis, exhibit some estimation bias (i.e., departure of estimates from their "true" values) in small samples (Linacre, 2009b). Hence, the low number of obese respondents requires caution in generalizing the results that were obtained from them.

In summary, results of our studies suggest that (a) researchers should be careful when interpreting the IAT effects as if they were equally influenced by positive and negative associations, hence IAT effects should not be interpreted as unambiguous measures of implicit prejudice and that (b) positive and negative associations involved in the IAT effects might be effectively decomposed.

Acknowledgments

The authors thank the Project Implicit (PI) team for its support and PI's principal investigators for their approval on data reporting. We are also grateful to Brian Nosek for helpful comments on an earlier draft.

References

Andrich, D. (1978). A rating scale formulation for ordered response categories. *Psychometrika, 43*, 561–573.

Andrich, D. (Ed.). (1988). *Rasch models for measurement*. Beverly Hills, CA: Sage.

Andrich, D. (2005). The Rasch model explained. In S. Alagumalai, D. D. Durtis, & N. Hungi (Eds.), *Applied Rasch measurement: A book of exemplars* (pp. 308–328). Dordrecht, The Netherlands: Springer-Kluwer.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task. A measure of association strengths. *Experimental Psychology, 56*, 329–343.

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

Bluemke, M., & Friese, M. (2008). Reliability and validity of the Single-Target IAT (ST-IAT): Assessing automatic affect towards multiple attitude objects. *European Journal of Social Psychology, 38*, 977–997.

Bond, T. G. & Fox, C. M. (Eds.). (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.

Cristante, F., & Robusto, E. (1999). Assessing dependence among subjects' responses. *Mathematical Social Sciences, 38*, 259–274.

Cristante, F., & Robusto, E. (2007). Assessing change with the extended logistic model. *British Journal of Mathematical and Statistical Psychology, 60*, 367–375.

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science, 14*, 14–18.

Eknoyan, G. (2008). Adolphe Quetelet (1796–1874) – the average man and indices of obesity. *Nephrology Dialysis Transplantation, 23*, 47–51.

Fischer, G. H. & Molenaar, I. W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York, NY: Springer-Verlag.

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology, 49*, 171–180.

Glaser, J., & Banaji, M. R. (1999). When fair is foul and foul is fair: Reverse priming in automatic evaluation. *Journal of Personality and Social Psychology, 77*, 669–687.

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

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.

Hugenberg, K., & Bodenhausen, G. V. (2003). Facing prejudice: Implicit prejudice and the perception of facial threat. *Psychological Science, 14*, 640–643.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*, 16–32.

Labiouse, C. L. (2004). Is there a real preferential detection of negative stimuli? A comment on Dijksterhuis and Aarts (2003). *Psychological Science, 15*, 364–365.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV. What we know (so far) about the method. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 59–102). New York, NY: Guilford Press.

Linacre J. M. (Ed.). (1989). *Many-facet Rasch measurement*. Chicago, IL: MESA Press.

Linacre, J. M. (2009a). *Facets Rasch measurement computer program (Version 3.66.0) [Computer software]*. Chicago, IL: Winsteps.com.

Linacre, J. M. (2009b). *Winsteps (Version 3.68.0) [Computer software]*. Chicago, IL: Winsteps.com.

Linacre, J. M., & Wright, B. D. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement, 3*, 486–512.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174.

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

Myford, C. M., & Wolfe, E. W. (2003). Detecting and measuring rater effects using many facet Rasch measurement: Part I. *Journal of Applied Measurement, 4*, 386–422.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 161–176.
Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 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., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., . . ., Banaji, M. R. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology, 18*, 36–88.
Popa-Roch, M., & Delmas, F. (2010). Prejudice Implicit Association Test effects. The role of self-related heuristics. *Zeitschrift für Psychologie, 218*, 44–50.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment test. Copenhagen: Danish Institute for Educational Research. Reprinted, 1980*. Chicago, IL: The University of Chicago Press.
Robusto, E., Cristante, F., & Vianello, M. (2008). Assessing the impact of replication on implicit association test effect by means of the extended logistic model for the assessment of change. *Behavior Research Methods, 40*, 954–960.
Rost, J., & Carstensen, C. H. (2002). Multidimensional Rasch measurement via item component models and faceted designs. *Applied Psychological Measurement, 26*, 42–56.
Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81*, 856–868.
Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition, 17*, 437–465.
Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology, 56*, 283–294.
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.
van Ravenzwaaij, D., van der Maas, H. L. J., & Wagenmakers, E. J. (2010). Does the Name – Race Implicit Association Test measure racial prejudice? *Experimental Psychology*. Advance online publication. doi: 10.1027/1618-3169/a000093.
Vianello, M., & Robusto, E. (2010). The many-facet Rasch model in the analysis of the Go/No-Go association task. *Behavior Research Methods, 42*, 944–956.
Wigboldus, D. H. J., Holland, R. W., & van Knippenberg, A. (2004). *Single target implicit associations.* Unpublished manuscript, Nijmegen, The Netherlands: Raboud University.
Wright, B. D. (1996). *Reliability and separation. Rasch measurement transactions, 9*, 472.
Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement, 29*, 23–48.

Received March 20, 2010
Revision received December 2, 2010
Accepted December 3, 2010
Published online May 18, 2011

Pasquale Anselmi

Department of Applied Psychology
University of Padua
Padua 35131
Italy
E-mail pasquale.anselmi@unipd.it