

# Assessing the impact of replication on implicit association test effect by means of the extended logistic model for the assessment of change

EGIDIO ROBUSTO, FRANCESCA CRISTANTE, AND MICHELANGELO VIANELLO  
*University of Padua, Padua, Italy*

The present study consists of a longitudinal investigation in which implicit association measures were collected by means of the implicit association test (IAT), with the purpose of assessing the impact of replication on the test effect with the same participants. The extended logistic model for the assessment of change (ELMAC), a procedure that belongs to Rasch modeling, was applied. By means of the ELMAC application, the existence of latent traits characteristic of the impact on the IAT effect (due to replication through time) is shown; more precisely, the model time parameters show the decrement of the IAT effect through time, due to the decrease of the implicit preference in both conditions, compatible and incompatible, associated with sweet foods and salty foods. The results confirm that in the long run, the incompatible condition latency may approach the compatible condition latency, causing a significant attenuation of the IAT effect.

An *implicit construct* is typically defined as “the introspectively unidentified (or inaccurately identified) trace of past experience that mediates *R*,” where *R* refers to the responses that are assumed to be influenced by that construct (Greenwald & Banaji, 1995, p. 5). According to this definition, an implicit *attitude*, for instance, can be defined as the trace of past experience that mediates the relationship between the individual and his or her responses to all objects and situations with which the attitude is related. Most important, implicit cognition is different from its explicit counterpart because it could reveal associative information that people are either unwilling or unable to report. In other words, implicit cognition represents traces of past experiences that people might explicitly reject because they conflict with values or beliefs, or might avoid revealing because the expression could have negative social consequences (Nosek, Greenwald, & Banaji, 2006). It is even more likely that implicit cognition can reveal information that is not available to introspective access, even if people are motivated to retrieve and express it (see Wilson, Lindsey, & Schooler, 2000). Basically, all measures of implicit constructs are based on the assumption that the time required for simple mental processes is an important source of information. This observation can be traced back to Donders (1868/1969). More specific, latency-based implicit measures assume that the time needed by the participant to “move” from a concept stored in memory to another is a function of the association between those concepts (semantic distance). This information is interpreted as a measure of the association between constructs and, therefore, as a measure of an implicit construct.

In recent years, implicit cognition has often been studied by means of the implicit association test (IAT; Greenwald, McGhee, & Schwartz, 1998). The task requires the sorting of stimuli from four concepts (e.g., flowers, insects, good words, bad words), using just two response options (typically two keys of a computer keyboard), each of which is assigned to two of the four concepts. The assumption is that this sorting task should be easier when the two concepts that share a response option are strongly associated (e.g., when flowers and good words share the same key) than when they are weakly associated (e.g., when insects and good words share the same key).

The IAT consists of seven phases, which in turn are composed of blocks of trials (B#), some of which are just practice tasks. The critical blocks of the IAT involve the simultaneous sorting of stimuli representing four concepts with two response options. Items are provided serially at the center of the screen. In the example above, one critical phase (two blocks, B3 and B4) would provide items representing flowers and good words (e.g., names of flowers, such as *tulip* and *rose*, and good words, such as *wonderful* and *glorious*) that would receive one response (e.g., the “E” key), and items representing the concepts insects and bad (e.g., names of insects, and words such as *terrible* and *horrible*) receive the alternative response (e.g., the “I” key). In the second critical phase (two blocks, B6 and B7), items representing insects and good words are sorted with one response, and items representing flowers and bad words are sorted with the alternative response. Critical phases are typically composed of 60 trials each (20 trials in B3 and B6 and 40 trials in B4 and B7). The IAT effect

---

E. Robusto, egidio.robusto@unipd.it

---

is computed comparing the performance obtained in the first critical phase with that obtained in the second critical phase. For individuals with stronger associations between flowers and good words, the first sorting task should be much easier than the second. Likewise, respondents who possess stronger associations between insects and good words should find the second sorting task to be easier than the first. The association can be measured both by the speed of responding (faster responding indicates stronger associations) and by the frequency of errors (fewer errors indicates stronger associations). Examples of the IAT procedures can be found at [www.projectimplicit.com](http://www.projectimplicit.com), and a discussion of main psychometric issues concerning the measure can be found in Nosek et al. (2006).

The IAT has been used to measure many different constructs, including implicit attitudes, self-esteem, self-concept, stereotypes, and preferences, in different areas, such as social and cognitive psychology (Fazio & Olson, 2003; Greenwald & Nosek, 2001), clinical psychology (de Jong, Pasman, Kindt, & van den Hout, 2001; Teachman, Gregg, & Woody, 2001), developmental psychology (Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2004), neuroscience (Cunningham et al., 2004; Phelps et al., 2000; Richeson & Nussbaum, 2004), market research (Maison, Greenwald, & Bruin, 2001), and health psychology (Teachman, Gapinski, Brownell, Rawlins, & Jeyaram, 2003). This technique is widely used because it seems to have better psychometric characteristics than its forerunners. For example, Bosson, Swann, and Pennebaker (2000) observed a split-half internal consistency for the self-esteem IAT of  $r = .69$ , compared with  $r$ s of  $-.05$  to  $.28$  for other latency-based implicit self-esteem measures. Many researchers have been involved in issues concerning the processes underlying IAT effects (e.g., Mierke & Klauer, 2003; Rothermund & Wentura, 2004) and the effects of procedural features, such as types of stimuli (e.g., De Houwer, 2001; Steffens & Plewe, 2001), the relationship between implicit and explicit measures (e.g., Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, 2005), and temporal stability of implicit measures (e.g., Blair, 2002; Cunningham et al., 2001; Egloff, Schwerdtfeger, & Schmukle, 2005; Nosek et al., 2006).

As far as the latter is concerned, studies that are based on the assumption that part of the IAT effect can be attributed to the respondents' expertise can be distinguished into two main categories. In the first one, the comparison of different levels of expertise and an enormous number of Web studies are considered. Greenwald and Nosek (2001) investigated the dependence of the IAT effect on the number of IATs previously taken by means of correlational analyses. The results showed that the IAT effect decreases as the number of IATs previously taken increases. The authors observed a correlation of  $-.15$  between the IAT effect and the number of IATs previously taken. The same phenomenon is cited in Nosek, Banaji, and Greenwald (2002) and Greenwald, Nosek, and Banaji (2003). The second category of studies on the stability of the IAT effect is based on a test-retest approach. Almost all contributions involve only two within-participants obser-

vations. In the only four-wave published study, a disattenuated correlation of  $.68$  was observed (Cunningham et al., 2001). A recent review by Nosek et al. (2006) identified an average correlation of  $.59$  in test-retest analyses, and more recently, Lane, Banaji, Nosek, and Greenwald (2007) reported an average value of  $.50$  across 20 test-retest studies. Such correlations tend to decrease slightly as the number of days between the two observations increases. These studies provide a good amount of information on the relationship between a first and second IAT, but they say little (or nothing) about how these effects evolve in more than two successive occasions or why the effect decreases.

In the present longitudinal study, implicit association measures were collected at three successive occasions, with the same participants, by means of the IAT. A formal procedure that belongs to Rasch modeling, the extended logistic model for the assessment of change (ELMAC; Cristante & Robusto, 2007), was applied. The application of the ELMAC allowed us to estimate parameters at successive occasions that represented the impact of replication on the IAT effect in a period of time. Such parameters were compared in order to verify the existence of statistical differences between IAT replications as the number of days between two applications increased. Moreover, they indicated the tendency of implicit association effects to decrease or increase over time. Fit statistics calculated for each parameter allowed us to verify the validity of the model to the data. The goal of this study was twofold: It analyzed how the IAT effect decreases over subsequent presentations, and it provided evidence of the empirical validity of the ELMAC.

## METHOD

### Participants and Procedure

Fifty-eight psychology students at the University of Padua, Italy, participated in this study on a voluntary basis. The IAT in the version suggested by Greenwald et al. (2003) was administered on a personal computer (Inquisit software), presenting visual stimuli from sweet food and salty food categories, with the purpose of obtaining measures of preference between the two types of food. A first application of the IAT (Time 1) was followed by a second application (Time 2) after 1 week, and by a third application (Time 3) 1 month after the first. The set of stimuli comprised four categories: 16 pictures of sweet foods, 16 pictures of salty foods, 16 positive words (e.g., *marvelous*, *glorious*), and 16 negative words (e.g., *terrible*, *horrible*). Participants used separate computer keys to indicate whether each food stimulus was sweet or salty and used the same two keys to indicate whether each word was positive or negative. In half of the critical trials (sweet food stimuli and positive words, Blocks B3 and B4; see Figure 1), participants classified sweet food stimuli and positive words on one key and salty food stimuli and negative words on the other key. In the other half (the sweet food stimuli plus negative words, Blocks B6 and B7; see Figure 1), participants classified salty food stimuli and positive words on one key and sweet food and negative words on the other key. The order of the blocks was counterbalanced. Blocks 1, 2, and 5 were practice trials.

### Measures

Response latencies were calculated for the four blocks (B3, B4, B6, B7). In order to analyze the latency data, sweet food stimuli and salty food stimuli were treated separately. For each of the two

	Positive Words Left Key	Negative Words Right Key
Sweet Food	B3 (20 trials) B4 (40 trials)	B6 (20 trials) B7 (40 trials)
Salty Food	B6 (20 trials) B7 (40 trials)	B3 (20 trials) B4 (40 trials)

**Figure 1. Trial blocks and keys in the IAT tasks according to Greenwald, Nosek, and Banaji (2003). Participants were provided Blocks B1 and B2 for practice purposes. These blocks required the participants to classify sweet versus salty food (B1) and positive versus negative words (B2). Block B5 required the participants to classify negative versus positive words and was provided for practice purposes.**

categories, a participants × stimuli matrix (58 × 16) was considered. In the two matrices, each participant was associated to each stimulus by means of a ratio value. The numerator of this ratio was the individual response time to stimuli provided in the food stimulus–negative word condition, and the denominator was the response time to stimuli provided in the food stimulus–positive word condition. In this way, a single preference indicator was devised, which allowed us to measure the implicit preference for each stimulus—that is, a ratio value equal to 1 indicated a preference that was neither positive nor negative. A ratio value lower than 1 indicated a negative implicit preference—that is, a shorter latency characterized the negative condition words. A ratio value higher than 1 indicated a positive implicit preference—that is, a shorter latency characterized the positive condition words. Successively, for all the ratio values in the matrices, two percentile values ( $P_{33}$ ,  $P_{66}$ ) were calculated, and the ratio values were coded with 0 if the ratio was lower than  $P_{33}$  (low implicit preference), with 1 if the ratio fell between  $P_{33}$  and  $P_{66}$  (intermediate implicit preference), and with 2 if the ratio was higher than  $P_{66}$  (high implicit preference). In this way, all data were coded according to a three-level rating scale. In order to apply the ELMAC to the data, we constructed two new matrices—participants × time points (the three successive occasions in which the IAT was provided to the participants)—one for sweet food stimuli and the other for salty food stimuli. In each matrix, for each participant and for each time point, we calculated a value corresponding to the sum of the 16 rating-scale measures obtained by each participant. The sum value corresponded to a random variable, which could take values from 0 to 32, where low values indicated an association between food stimuli and negative words (and thus a negative implicit evaluation of the food stimuli) and high values indicated an association between food stimuli and positive words (and thus a positive implicit evaluation of the food stimuli). Separated implicit evaluations for sweet and salty foods were computed. Both matrices were analyzed by means of the ELMAC.

**The Model**

As explained by Cristante and Robusto (2007), the ELMAC is inscribed in a persons × time points matrix and is devised for the assessment of change—or, in other words, for the assessment of temporal stability or instability of a specific measure. Such an approach allows the definition of three parameters: a person parameter, a time parameter, and a dispersion parameter. The parameters are combined in the model, which is an interpretation of the extended logistic model (Andrich, 1982, 1985; Andrich & Luo, 2003).

In the ELMAC, the time parameter describes the intensity with which an event is experienced by persons through time and is located on the latent trait. The time parameter is estimated by taking into consideration a set of measures at successive time points. Such measures (in this study, the 16 stimuli coded on a three-level rating scale) are mapped into a single measure via a sum score (in this

study, the sum of the 16 rating-scale measures). According to this interpretation, at each time point, a 0 value indicates the lowest value (in this study, the minimum level of preference), and a value of  $m$  corresponds to the highest value (in this study, 32, the maximum level of preference).

The model is

$$P\{X_{vt} = x \mid \beta_v, \delta_t, \omega_t, m\} = \frac{\exp\left[x\left(\frac{m-x}{\omega_t}\right) + x(\beta_v - \delta_t)\right]}{\gamma_{vt}}, \quad (1)$$

where  $X_{vt} = x$  is the random sum variable, which in this study takes values equal to 0, 1, 2, . . . , 32;  $\beta_v$  is the person  $v$  parameter; and  $\delta_t$  is the time  $t$  parameter, which describes the intensity with which a particular event occurs at time  $t$ . In this study, it describes the level of IAT tasks’ replication impact.  $\omega_t$  is the dispersion parameter, and  $\gamma_{vt}$  is the normalizing factor.

In order to demonstrate the appropriateness of the application of the ELMAC in the context of an IAT longitudinal study, it is useful to underline the following aspects. A fundamental aspect is the fact that the ELMAC is based on the idea that one task given to the same person at different occasions in a period of time (e.g., at two successive time points) can be considered a pair of tasks. Any change occurring between testing occasions can be described as being a change of the task parameters, instead of describing change in terms of the person parameters. This central idea is also at the base of a broad class of applications of the linear logistic model (LLM) for measuring change by Fischer (1995). A comparison between the LLM for measuring change and the ELMAC shows that the matrices in which the models are applied are different. The matrix in which the model by Fischer is developed is participants × variables repeatedly observed during a period of time, whereas the ELMAC is inscribed in a matrix of participants × time points, where a polytomous single variable is introduced and observed at each time point. In this way, the ELMAC allows a very simple analysis in accordance with the properties of Rasch modeling (Cristante & Robusto, 2007).

A second, nontrivial aspect is the linear logistic property of the ELMAC. Within such a perspective (which characterizes the Rasch modeling approach), it is possible to construct a time variable, defined in this article as replication impact through time, that is measured according to the rules of fundamental measurement, overcoming in this way the limitations due to the participants’ responses’ qualitative scaling. Such result is new in the IAT literature and has been hypothesized by some authors (e.g., Embretson, 2006).

A third aspect that significantly qualifies a longitudinal study is that the ELMAC, due to its Rasch modeling properties, allows not only the ordering of the time parameters on the latent time variable, but also the verifying of the fit of each time point parameter. When the time parameters have a good fit, a constant change due to the effect of test replications is demonstrated. A time parameter’s bad fit is evidence of an interruption in the process of change (Fischer, 1995).

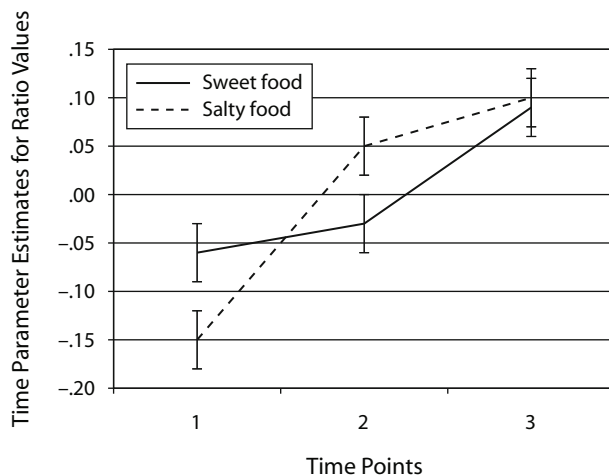
**Table 1**  
**Observed Scores, Time Parameter Estimates for Ratio Values (Negative Condition Latency, Positive Condition Latency), As Well As Standard Errors and Infit and Outfit Statistics**

Time Point	Sweet Food					Salty Food				
	Obs. Score	$\delta_t$	SE	Infit	Outfit	Obs. Score	$\delta_t$	SE	Infit	Outfit
$t_1$	988	-.06	.03	.86	.82	1072	-.15	.03	.96	.93
$t_2$	959	-.03	.03	1.08	1.08	888	.05	.03	1.12	1.07
$t_3$	854	.09	.03	1.07	1.09	846	.10	.03	.94	.92

Note—Reliabilities of separations (participants facet) are equal to .80 (sweet food) and .78 (salty food).

## RESULTS

Table 1 shows the  $\delta_t$  parameters estimated on the ratio values (negative word condition latency on positive word condition latency) for sweet food and for salty food, respectively, at each time point, as well as the standard errors and the corresponding infit and outfit statistics.<sup>1</sup> In order to interpret the results, it is useful to note that the order of the  $\delta_t$  parameter estimates is opposite the order of the corresponding sum values; that is, the lowest  $\delta_t$  value corresponds to the highest sum value, which in turn corresponds to the maximum positive level of preference. In Table 1, both for sweet foods and for salty foods, as the  $\delta_t$  parameters increase from Time 1 to Time 3, the corresponding sum values decrease through time, denoting a tendency toward a negative implicit preference. This result demonstrates that in the period of time considered, the ratios between latency measures for the positive word condition and those for the negative word condition change. In other words, the relationship between denominator and numerator (i.e., the IAT effect) is not stable. This finding is observed for both sweet foods and salty foods. Figure 2 shows that participants prefer salty foods at Time 1 and sweet foods at Time 2, and that they have basically no preference at Time 3. All the parameters also have satisfactory infit and outfit statistics; in other words, they are congruent with the definition of a *latent trait denominated* as replication impact on IAT effect through time.



**Figure 2.** Plot of the time parameter estimates for ratio values (negative condition latency, positive condition latency, y-axis) at Times 1, 2, and 3 (x-axis). Error bars represent two standard errors of the estimates.

In order to verify whether the change from one time point to another is significant, a standard score was calculated for each pair of  $\delta_t$  values, separately for sweet and salty foods. The results reported in Table 2 show that the preference for sweet foods significantly differs from Time 1 to Time 3 and from Time 2 to Time 3. The implicit preference for salty foods differs significantly from Time 1 to Time 2 and from Time 1 to Time 3. These results support the following remarks:

1. A constant change of the ratio values is observed both for sweet foods and for salty foods; that is, multiple presentations of the IAT influence its effects, and the implicit evaluation of both kinds of foods tends to decrease.

2. Such a change is represented by the increasing of the  $\delta_t$  parameters from Time 1 to Time 3, which corresponds to the decrement of the sum values. This decrement is due to the temporal instability of the relationship between the latencies of the two critical phases of the IAT (i.e., the positive word condition and the negative word condition).

3. A hypothesis is made that, if the ratio between the positive word condition latency and the negative word condition latency is changing through time (for both sweet and salty foods) due to the respondents' expertise, not only should the latency for the positive word condition decrease through time, but the latency for the negative word condition should diminish significantly as well.

The following analyses were planned with the purpose of verifying the aforementioned hypothesis: In the first analysis, in order to examine the tendency of the response latency to decrease through time, in both the negative and positive word conditions, two matrices were analyzed by means of the ELMAC, separately for sweet foods and salty foods. The matrices were participants  $\times$  time points—one for sweet foods in the two conditions (negative words and positive words) and another for salty foods in the two conditions. In each matrix, for each participant and for each time point (three time points for the negative word condition followed by the three time points for the positive word condition), a value that corresponded to the sum of the 16 stimuli rating-scale measures was calculated. For each stimulus, a rating-scale value was obtained by recoding the response latency measures on the basis of two percentiles ( $P_{33}$ ,  $P_{66}$ ) calculated on the data of a single matrix that included the two matrices, sweet and salty food. Measures below  $P_{33}$  were coded 0 (low response latency), measures between  $P_{33}$  and  $P_{66}$  were coded 1 (intermediate response latency), and measures above  $P_{66}$  were coded 2 (high response latency). The sum value corresponded to a random variable, which could take values from 0 (the minimum response latency) to 32 (the maximum response latency).

**Table 2**  
**Standard Scores of Differences Between Time Points for Ratio Value Parameters (Negative Condition Latency, Positive Condition Latency), As Well As Fit Probabilities**

Time Point Difference	Sweet Food	Salty Food
$t_1 - t_2$	-.71 <sup>††</sup>	-4.71 <sup>***</sup>
$t_1 - t_3$	-3.00 <sup>***</sup>	-5.89 <sup>***</sup>
$t_2 - t_3$	-2.40 <sup>†</sup>	-1.18 <sup>††</sup>

\*\*\* $p < .001$ . <sup>†</sup> $p = .008$ . <sup>††</sup>n.s.

Two separate ELMAC analyses were conducted—one for sweet foods and one for salty foods.

In the second analysis, in order to verify the differences between the time parameters obtained from the two analyses, a number of comparisons between  $\delta_t$  parameters were planned. First, the comparisons between time parameters of sweet and salty foods in the negative and positive conditions were performed. In both conditions, and for both sweet and salty foods, a response latency decrement through time was hypothesized. Successively, at each time point, comparisons of time parameters in the negative versus positive condition, separately for sweet and salty foods, were performed. Table 3 shows time parameters, standard errors, and fit indexes for the negative word and the positive word conditions, separately for sweet foods and salty foods, at each time point.

Similar to the previous results, and in order to interpret them, it is useful to remember that the order of the  $\delta_t$  parameter estimates is opposite that of the corresponding sum values; that is, the lowest  $\delta_t$  parameter estimate corresponds to the highest sum value, which in turn corresponds to the highest latency measure. In Table 3, for both sweet and salty foods, the  $\delta_t$  parameters increase from Time 1 to Time 3, in both the negative and positive conditions, showing not only that the response latency decreases from Time 1 to Time 3, but also that latencies of the negative

condition are always greater than those of the positive condition, at any time point considered. All the  $\delta_t$  parameters present satisfactory infit and outfit statistics, except for  $t_1(N)$  for salty foods, showing the existence of a latent trait denominated as replication impact on negative and positive association conditions through time. In order to learn whether the change between time points is significant, a standard score was calculated for each pair of  $\delta_t$  values, separately for sweet foods and salty foods (see Table 4).

Table 4 shows that the differences between time points are always significant, with two exceptions out of 12: the differences between Time 2 and Time 3 in the positive word condition. This means that the decrease of latency in the positive word condition ceases after the second replication, whereas in the negative word condition it persists until the last replication. This finding could be evidence of the fact that in the long run, the difference between latencies in the negative and positive word conditions becomes smaller and smaller, thus attenuating the IAT effect. Table 5 provides the differences between parameters of the negative word condition and those of the positive word condition, for both sweet and salty foods, at each time point.

The results show that the positive word condition latency measures are always significantly lower than the negative word condition latencies at each time point. In summary, a constant change in response latency was observed in both negative word and positive word conditions, for sweet foods and for salty foods (Table 3). Such change demonstrates a decrement of response latency in both conditions; that is, the respondents become faster with the replication of the IAT tasks. Such change is always significant, except in one case (between Time 2 and Time 3, concerning the positive word condition), for both sweet foods and salty foods (Table 4). All through the period of time considered, the response latency in the positive word condition is shorter than the response latency in the negative word condition, for both sweet foods and salty foods (Table 5).

**Table 3**  
**Observed Scores, Time Parameter Estimates for the Negative Word Condition (N) and the Positive Word Condition (P), As Well As Standard Errors and Infit and Outfit Statistics**

Time Points	Sweet Food					Salty Food				
	Obs. Score	$\delta_t$	SE	Infit	Outfit	Obs. Score	$\delta_t$	SE	Infit	Outfit
$t_1(N)$	1404	-.36	.03	1.20	1.08	1495	-.45	.03	1.48	1.36
$t_2(N)$	1167	-.17	.03	.96	1.00	1177	-.17	.03	1.00	.95
$t_3(N)$	1003	-.06	.03	1.00	.99	1054	-.08	.03	.86	.83
$t_1(P)$	809	.08	.03	1.00	1.06	894	.05	.03	.93	.91
$t_2(P)$	534	.28	.03	.89	.91	641	.25	.03	.89	1.00
$t_3(P)$	497	.31	.03	.93	1.18	583	.30	.03	1.16	1.17

Note—Reliabilities of separations (participants facet) are equal to .99 (both foods).

**Table 4**  
**Standard Scores of Differences Between Time Points, As Well As Fit Probabilities**

Time Point Difference	Negative Word Condition		Positive Word Condition	
	Sweet Food	Salty Food	Sweet Food	Salty Food
$t_1 - t_2$	-4.478 <sup>***</sup>	-6.600 <sup>***</sup>	-4.714 <sup>***</sup>	-4.714 <sup>***</sup>
$t_1 - t_3$	-7.071 <sup>***</sup>	-8.721 <sup>***</sup>	-5.421 <sup>***</sup>	-5.893 <sup>***</sup>
$t_2 - t_3$	-2.593 <sup>**</sup>	-2.121 <sup>**</sup>	-.707 <sup>†</sup>	-1.179 <sup>†</sup>

\*\* $p < .01$ . \*\*\* $p = .001$ . <sup>†</sup>n.s.

**Table 5**  
**Standard Scores of Differences Between the**  
**Negative Word Condition (*N*) and the**  
**Positive Word Condition (*P*), at Each Time Point**

Condition Difference	Sweet Food	Salty Food
$t_1(N) - t_1(P)$	-10.371	-11.785
$t_2(N) - t_2(P)$	-10.607	-9.899
$t_3(N) - t_3(P)$	-8.721	-8.957

Note—All  $ps < .001$ .

In other words, although the latency in the negative word condition diminishes through time, the latency in the positive word condition is always shorter, showing a constant preference for the positive word condition for both sweet foods and salty foods.

### REMARKS

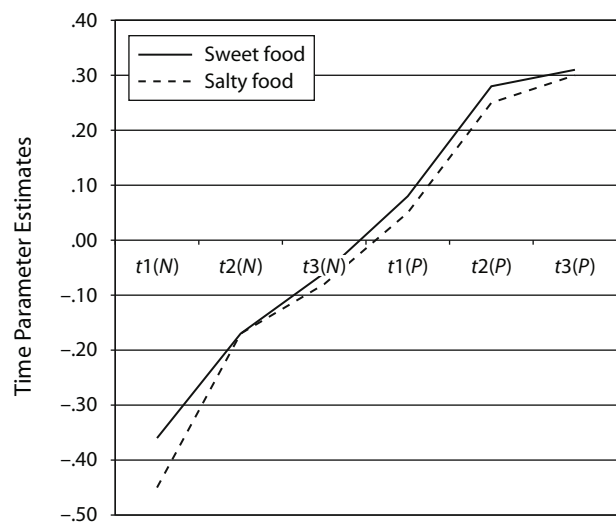
The main purpose of the present investigation was to address a recurring question regarding the stability of the IAT effect when its tasks are replicated on different occasions during a period of time with the same participants. In this perspective, examination of previous research showed that IAT effects decrease over time. This effect was known; Greenwald and Nosek's (2001) and many other studies confirmed a low test–retest reliability of the IAT. Lane et al. (2007) recently reported an average test–retest correlation of only .50 across 20 studies. Yet, these studies say little about how and why the effect decreases or what happens when measurement occasions exceed two. Such knowledge might be useful in adequately interpreting the IAT effects and might suggest how to improve the procedure's reliability. This phenomenon was investigated by means of the ELMAC. A latent trait model was chosen with the purpose of obtaining time parameter estimates independent of a particular population and located on a latent trait.

Fifty-eight university psychology students participated in a three-wave study in which the IAT (Greenwald et al., 2003) was provided a first time, then a week later, and then a month after the first time. A constant change of the ratio value between latencies of the negative and the positive words, for both sweet and salty foods, was observed. This result suggests that an impact of replication on the IAT effect takes place in the period of time considered, as expected on the basis of previous research on the topic. Exploring the data further, it was observed that such a change was represented by the decrement of the ratio values, showing a temporal instability of the relationship between latencies of the negative and positive word conditions. Then it was argued that if expertise is at the basis of this effect, decrements in both the positive and negative word conditions' latencies should be expected, due to the respondents' familiarity with the test procedure. The results support this prediction. Indeed, a significant decrease is observed for both negative and positive word conditions and for both sweet and salty foods. The data we obtained also revealed that the negative word condition latency decreases significantly through the period of

time considered, whereas the positive word condition latency reaches its minimum value at the second replication (Figure 3). In other words, it was observed that whereas participants constantly enhance their performance (i.e., get faster) in the negative word condition, their maximum performance in the positive word condition is reached at Time 2. As a result, latencies of the two critical conditions get closer over time, and the IAT effect decreases.

In this study, although a constant decrease of the negative word condition latency is observed, the response latency for the positive word condition, considered at each time point, is always shorter than the latency for the negative word condition, showing an association between foods and positive words. There may be two consequences of this difference in the evolution of performance that could be profitably investigated in future research. According to a first hypothesis, latencies of the negative word condition keep decreasing until they equal those of the positive word condition. In such a scenario, considering the trend observed in this study, the IAT effect should disappear in fewer than a dozen successive presentations. An alternative possibility is that after a smaller number of presentations, latencies of the negative word condition stop decreasing (due to a kind of floor effect of participants' performance). Consequences of this scenario are much more encouraging for the IAT, since if the effect holds over time, it might also be considered reliable, due to a stabilization of participants' performances at the tasks. Both consequences would be of great importance in the interpretation of the IAT's test–retest reliability.

Concerning the latent trait analyses conducted in this study by means of the ELMAC, the results support the definition of latent traits that describe the impact of the replication of IAT tasks over a period of time. These re-



**Figure 3.** Plot of the time parameter estimates for sweet food and salty food in both the negative and the positive word conditions. Note that  $t$  indicates any time point, and  $N$  and  $P$  indicate negative and positive conditions, respectively.

sults support the appropriateness of Rasch modeling with IAT measures (Embretson, 2006).

#### AUTHOR NOTE

Correspondence concerning this article should be addressed to E. Robusto, Dipartimento di Psicologia Generale, Università di Padova, Via Venezia, 8, 35131 Padova, Italy (e-mail: egidio.robusto@unipd.it).

#### REFERENCES

- ANDRICH, D. (1982). An extension of the Rasch model for ratings providing both location and dispersion parameters. *Psychometrika*, *47*, 105-113.
- ANDRICH, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. B. Tuma (Ed.), *Sociological methodology* (pp. 33-80). San Francisco: Jossey-Bass.
- ANDRICH, D., & LUO, G. (2003). Conditional pairwise estimation in the Rasch model for ordered response categories using principal components. *Journal of Applied Measurement*, *4*, 205-221.
- BARON, A. S., & BANAJI, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science*, *17*, 53-58.
- BLAIR, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality & Social Psychology Review*, *6*, 242-261.
- BOND, T. G., & FOX, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Erlbaum.
- 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 & Social Psychology*, *79*, 631-643.
- CRISTANTE, F., & ROBUSTO, E. (2007). Assessing change with the extended logistic model. *British Journal of Mathematical & Statistical Psychology*, *60*, 367-375.
- CUNNINGHAM, W. A., JOHNSON, M. K., RAYE, C. L., GATENBY, J. C., GORE, J. C., & BANAJI, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychological Science*, *15*, 806-813.
- CUNNINGHAM, W. A., PREACHER, K. J., & BANAJI, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, *12*, 163-170.
- DE HOUWER, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*, 443-451.
- DE JONG, P. J., PASMAN, W., KINDT, M., & VAN DEN HOUT, M. A. (2001). A reaction time paradigm to assess (implicit) complaint-specific dysfunctional beliefs. *Behaviour Research & Therapy*, *39*, 101-113.
- DONDERS, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412-431. (Original work published 1868)
- DUNHAM, Y., BARON, A., & BANAJI, M. (2004, June). *Developmental social psychology: Outlining a new approach to the study of prejudice in children*. Paper presented at the 34th Annual Meeting of the Jean Piaget Society, Toronto.
- EGLOFF, B., SCHWERTFEGER, A., & SCHMUKLE, S. C. (2005). Temporal stability of the Implicit Association Test—Anxiety. *Journal of Personality Assessment*, *84*, 82-88.
- EMBRETSON, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *American Psychologist*, *61*, 50-55.
- FAZIO, R. H., & OLSON, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297-327.
- FISCHER, G. H. (1995). Linear logistic models for change. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 157-180). New York: Springer.
- GREENWALD, A. G., & BANAJI, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4-27.
- GREENWALD, A. G., MCGHEE, D. E., & SCHWARTZ, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality & Social Psychology*, *74*, 1464-1480.
- GREENWALD, A. G., & NOSEK, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift für Experimentelle Psychologie*, *48*, 85-93.
- GREENWALD, A. G., NOSEK, B. A., & BANAJI, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality & Social Psychology*, *85*, 197-216.
- 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 & Social Psychology Bulletin*, *31*, 1369-1385.
- 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: Guilford.
- MAISON, D., GREENWALD, A. G., & BRUIN, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin*, *32*, 61-69.
- MIERKE, J., & KLAUER, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality & Social Psychology*, *85*, 1180-1192.
- NOSEK, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*, 565-584.
- NOSEK, B. A., BANAJI, M. R., & GREENWALD, A. G. (2002). E-research: Ethics, security, design, and control in psychological research on the Internet. *Journal of Social Issues*, *58*, 161-176.
- NOSEK, B., GREENWALD, A., & BANAJI, M. (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: Psychology Press.
- PHELPS, E. A., O'CONNOR, K. J., CUNNINGHAM, W. A., FUNAYAMA, E. S., GATENBY, J. C., GORE, J. C., & BANAJI, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience*, *12*, 729-738.
- RICHESON, J. A., & NUSSBAUM, R. J. (2004). The impact of multiculturalism versus color-blindness on racial bias. *Journal of Experimental Social Psychology*, *40*, 417-423.
- ROTHERMUND, K., & WENTURA, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*, 139-165.
- 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.
- TEACHMAN, B. A., GAPINSKI, K. D., BROWNELL, K. D., RAWLINS, M., & JEYARAM, S. (2003). Demonstrations of implicit anti-fat bias: The impact of providing causal information and evoking empathy. *Health Psychology*, *22*, 68-78.
- TEACHMAN, B. A., GREGG, A. P., & WOODY, S. R. (2001). Implicit associations for fear-relevant stimuli among individuals with snake and spider fears. *Journal of Abnormal Psychology*, *110*, 226-235.
- WILSON, T. D., LINDSEY, S., & SCHOOLER, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101-126.

#### NOTE

1. The infit and outfit statistics determine how well any set of empirical data meets the requirements of the model. The infit statistic places more emphasis on unexpected responses near the item (or person) measure, whereas the outfit statistic places more emphasis on unexpected responses far from an item (or person) measure. More precisely, the outfit statistic is based on the squared residuals for each person on each item, whereas the infit measure is an information-weighted sum. The infit and the outfit statistics have a scale form with an expected value of +1 and a range from 0 to positive infinite. An infit and outfit statistic in the range 0.70–1.30 indicates a satisfactory fit of the empirical data to the model (Bond & Fox, 2001).

(Manuscript received May 9, 2008;  
revision accepted for publication June 13, 2008.)