

A discrimination–association model for decomposing component processes of the Implicit Association Test

Luca Stefanutti · Egidio Robusto · Michelangelo Vianello · Pasquale Anselmi

Published online: 13 December 2012
© Psychonomic Society, Inc. 2012

Abstract A formal model is proposed that decomposes the Implicit Association Test (IAT) effect into three process components: stimuli discrimination, automatic association, and termination criterion. Both response accuracy and reaction time are considered. Four independent and parallel Poisson processes, one for each of the four label categories of the IAT, are assumed. The model parameters are the rate at which information accrues on the counter of each process and the amount of information that is needed before a response is given. The aim of this study is to present the model and an illustrative application in which the process components of a Coca–Pepsi IAT are decomposed.

Keywords Poisson race model · Poisson process · Implicit Association Test · Implicit measure

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is a computerized two-choice discrimination task in which stimuli have to be categorized as belonging to object categories (e.g., *flowers* and *insects*) or attribute categories (e.g., *good* and *bad*) by pressing, as quickly and accurately as possible, one of two response keys. The IAT consists of seven blocks. Three practice blocks involve the categorization of stimuli that represent either the object categories or the attribute categories. Four critical blocks involve the simultaneous categorization of stimuli representing the four categories, with two response mappings. In one mapping, the categories *flowers* and *good* share a response key, and the categories *insects* and *bad*

share the other. In the other mapping, the categories *flowers* and *bad* share a response key, and the categories *insects* and *good* share the other. The mapping that leads to faster and more accurate responses is called *compatible* mapping, whereas the other is called *incompatible* mapping. The difference in performance between the two kinds of mapping is known as the IAT effect.

A few models for the analysis of process components underlying the IAT effect are present in the literature. The quad model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) is a multinomial model that disentangles components concerning the automatic activation of associations, the ability to discriminate stimuli, the ability to overcome automatically activated associations, and the influence of general guessing biases. Since the model accounts only for response accuracy, it uses a very small part of the information provided by an IAT. With regard to this limitation, the four parameters of the quad model account for only 24 % of the variance of the *D* scores (Greenwald, Nosek, & Banaji, 2003) computed on the same data (Conrey et al., 2005).

More recently, a diffusion model (DM) analysis has been proposed (Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007), which disentangles components concerning the rate of information accumulation (*drift*), the speed–accuracy setting (*threshold*), and the nondecision components. The analysis assumes serial processing of the stimuli, and it simultaneously accounts for response accuracy and reaction time. Both drifts and thresholds vary across blocks of trials. The DM provides useful information in the analysis of IAT data. However, it also presents some shortcomings. At the estimation level, only IAT procedures including many more trials than usual can be analyzed, and different DMs have to be estimated for the critical blocks. Hence, parameter values estimated for different blocks cannot be compared apart from the other parameters to obtain an individual measure of automatic association. For instance, drift rates of different blocks cannot be compared

L. Stefanutti (✉) · E. Robusto · M. Vianello · P. Anselmi
FISPPA Department, University of Padua,
via Venezia 8,
35131 Padua, Italy
e-mail: luca.stefanutti@unipd.it

without taking into account thresholds as well. At the theoretical level, the DM assumes that the individual serially processes each stimulus, although the empirical evidence accumulated using many different visual search tasks, including multiple-target searches (van der Heijden, 1975) and short words as stimuli, seems to support most parallel, rather than serial, processing of stimuli (Andrews, 1992; Evans, Horowitz, & Wolfe, 2011; Thornton & Gilden, 2007; Townsend, 1990).

As a result, there is a need for a formal model that simultaneously allows the separation of automatic and controlled processes, accounts for both response times (RTs) and errors, and provides researchers with a quantitative and detailed comparison of individual performance at the critical tasks. This study presents a model that overcomes current limitations. The model has been specifically derived to fit the characteristics of the IAT and the needs of the researchers using it. The discrimination–association model (DAM) decomposes the IAT effect into three process components: stimuli discrimination, automatic association, and termination criterion (i.e., task difficulty or individual cautiousness). Unlike the DM analysis of Klauer et al. (2007), in this model, the automatic association component depends on the stimulus categories, but it does not vary across blocks.

The aim of this study is to present the DAM and its validation. In the next section, the DAM and its mathematical specification are introduced. Next, the results of an empirical application of the model are provided. Finally, advantages of the analysis of IAT data derived from using the DAM, instead of other models, are explored.

Overview of the model

In this section, the development of a stochastic model of the accuracy and RTs of an individual respondent to an IAT is described. As such, compatible and incompatible blocks are defined within each participant and can, therefore, vary across participants.

We shall start by considering a collection O of *object* stimuli and a collection A of *attribute* stimuli. The task of an individual is to classify the objects in O into the two categories $[a]$ and $[b]$ (e.g., *flowers* and *insects*) and the elements in A into the two categories $[+]$ and $[-]$ (e.g., *good* and *bad*). The assumption behind the model is that every stimulus in $S = O \cup A$ potentially contains—albeit in a variable quantity—evidence for each of the four categories $[a]$, $[b]$, $[+]$, $[-]$ and that processing of every single stimulus in S occurs in parallel. In particular, evidence in favor of each of the four categories is accumulated on separate, parallel, and independent stochastic processes (called *counters*) that are engaged in a competition for the emission of the observable response.

According to this idea, the existence of four separate and independent counters, denoted as $X_a(t)$, $X_b(t)$, $X_+(t)$, and $X_-(t)$, is assumed. Once a participant is presented with a particular stimulus s , each of these four counters starts accruing selective information about a certain characteristic of it. More precisely, for $j \in \{a, b, +, -\}$, counter $X_j(t)$ accrues information about membership of s in category $[j]$. A specific assumption concerning the nature of these counters is that each of them behaves as a Poisson process. Technically, this implies that, in every single process, (1) interarrival times (i.e., time intervals between consecutive units of information) are independent and identically distributed and (2) their distribution is exponential with rate λ (in this respect, see, e.g., Townsend & Ashby, 1983).

Stimulus discrimination and automatic association

A key property of the model is that the rate at which information is accumulated on each of the four processes depends only on the process and the actual category of the presented stimulus, and on nothing else. Under this assumption, the rate of every single process is not affected by the particular block (practice, compatible, or incompatible) in which the stimulus appears. Thus, in the most general formulation of the model, there will be a different rate for each pair that can be formed by taking one of the four processes $X_a(t)$, $X_b(t)$, $X_+(t)$, $X_-(t)$ and one of the four categories $[a]$, $[b]$, $[+]$, $[-]$, as is shown in Table 1.

According to this formulation, for $i, j \in \{a, b, +, -\}$, the parameter λ_{ij} is the average amount of information that process $X_i(t)$ accumulates in the time unit, when a stimulus of category $[j]$ is presented. For example, if the presented stimulus belongs to category $[a]$, the rate of process $X_a(t)$ will be as large as λ_{aa} , while that of process $X_+(t)$ will be λ_{+a} .

The 4×4 matrix consisting of the 16 rates listed in Table 1 is naturally decomposed into four submatrices of four rates each. The 2×2 submatrix containing the four rates λ_{aa} , λ_{ab} , λ_{ba} , λ_{bb} (the upper left submatrix of the table) is involved in the discrimination between categories $[a]$ and $[b]$. The better the discrimination, the smaller will be the two

Table 1 Accumulation rates of each of the four Poisson processes vary with stimulus category

Processes	Stimulus categories			
	$[a]$	$[b]$	$[+]$	$[-]$
$X_a(t)$	λ_{aa}	λ_{ab}	λ_{a+}	λ_{a-}
$X_b(t)$	λ_{ba}	λ_{bb}	λ_{b+}	λ_{b-}
$X_+(t)$	λ_{+a}	λ_{+b}	λ_{++}	λ_{+-}
$X_-(t)$	λ_{-a}	λ_{-b}	λ_{-+}	λ_{--}

rates λ_{ab} and λ_{ba} . A perfect discrimination would indeed imply $\lambda_{ab} = \lambda_{ba} = 0$. Similarly, the lower right submatrix is involved in the discrimination between [+] and [-]. In fact, the most interesting part of Table 1 is represented by the lower left and upper right 2×2 submatrices, because they are involved in an association between object and attribute categories. For instance, with regard to the lower left subtable, the parameter λ_{+a} is the rate at which information concerning membership in category [+] is accumulated when the presented stimulus belongs, indeed, to category [a]. To give another example, λ_{-b} is the rate at which evidence about membership in category [-] is accrued when the presented stimulus actually belongs to category [b].

Overall, the particular set of values taken by the four parameters $\lambda_{+a}, \lambda_{-a}, \lambda_{+b}, \lambda_{-b}$ can be regarded as an *association pattern*. In a practical application of the model, this pattern will typically change from one individual to another, emphasizing individual differences in both a quantitative and qualitative fashion. Obviously, the four rates $\lambda_{a+}, \lambda_{b+}, \lambda_{a-}, \lambda_{b-}$ can also be regarded as an association pattern (for a comparison between the two submatrices, see the following paragraph). To give some examples, the case $\lambda_{+a} > 0, \lambda_{-a} = 0$ can be regarded as a perfect association between categories [a] and [+]. Similarly, the condition $\lambda_{-b} > 0, \lambda_{+b} = 0$ represents a perfect association between [b] and [-]. Obviously, many other possibilities might arise in an empirical setting.

It is clear that both the lower left and upper right submatrices are involved in an association between objects and attributes, although with different meanings. The lower left submatrix is related to what will be referred to as an *object-driven association* in the sequel. An association of this type arises when the presented stimulus is an object. The opposite situation occurs when the stimulus is an attribute, a condition in which the association is called *attribute driven*. This is precisely what happens in the upper right submatrix of Table 1. Obviously, it is a working hypothesis that the two association types (object driven and attribute driven) are different for the same individual. Acceptance or rejection of this specific assumption can occur only on an empirical basis and, in any case, is not critical, because it is always possible to introduce equality constraints of the form $\lambda_{ij} = \lambda_{ji}$ for all the association rates.

To summarize, the rate parameters presented in Table 1 can be grouped into four separate categories: (1) object discrimination rates (upper left submatrix), (2) attribute discrimination rates (lower right submatrix), (3) object-driven association rates (lower left submatrix), and (4) attribute-driven association rates (upper right submatrix).

Process superposition and the IAT effect

Not much has been stated so far about the termination of the accumulation process and how each process is involved in

producing an observable response. Figure 1 is a pictorial representation of how the four processes are related to both the stimulus and observable response on a single trial of the compatible (left-hand diagram) and incompatible (right-hand diagram) blocks of an IAT.

The left-hand diagram refers to a trial in which categories [a] and [+] are mapped to the left key, while [b] and [-] are mapped to the right key. Arrows departing from the stimulus and reaching each of the four processes represent information accumulation, which depends on the rates of the four processes.

Concerning the relationship between each of the processes and each of the two response categories (*left* or *right*), it can be seen that both processes $X_a(t)$ and $X_+(t)$ produce the same observable response—namely, *left*—whereas the response is *right* for both $X_b(t)$ and $X_-(t)$. That is, the four processes always operate in opposite pairs, or stated another way, there is a race between pairs of processes. The pair that first accumulates the required amount of information produces the observable response. In the standard two-process Poisson race model, this amount of information is often called the *termination criterion*.

According to a well-known property of Poisson processes, the superposition $X_1(t) + X_2(t)$ of two independent Poisson processes, having rates λ_1 and λ_2 , is itself a Poisson process with a rate $\lambda_1 + \lambda_2$. Therefore, the superposition $X_a(t) + X_+(t)$ is a Poisson process whose rate is $\lambda_{a+} + \lambda_{+a}$, where $i \in \{a, b, +, -\}$ is the category of the presented stimulus. Similar considerations follow for the superposition $X_b(t) + X_-(t)$, whose rate will be $\lambda_{b-} + \lambda_{-b}$.

It is a precise assumption of the proposed model that in the compatible blocks of the IAT, a race takes place between the two compound processes $X_{a+}(t) = X_a(t) + X_+(t)$ and $X_{b-}(t) = X_b(t) + X_-(t)$ which have the same termination criterion (Fig. 2). The process that first meets the criterion will produce the observable response. With the introduction of this assumption, the model can be recast as the standard two-process Poisson race model described in Townsend and Ashby (1983).

The right-hand diagram of Fig. 1 refers to a trial of the incompatible block. Here, the categories [a] and [-] are

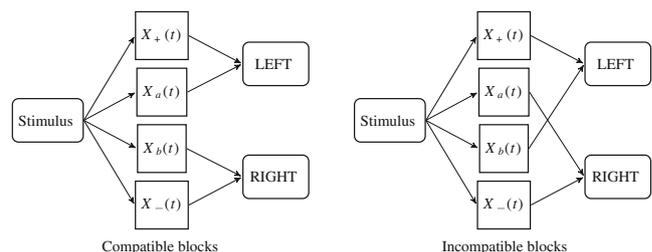


Fig. 1 How processes are connected to the stimulus and the observable response in the compatible (left-hand diagram) and incompatible (right-hand diagram) blocks

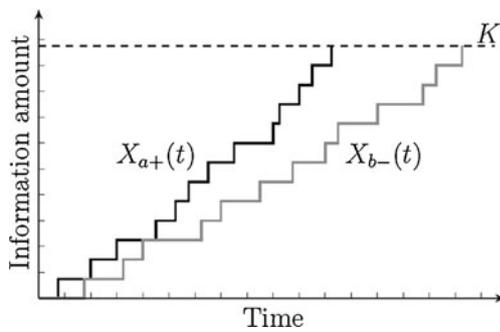


Fig. 2 A race between two-compound Poisson processes. Process $X_{a+}(t)$ has a higher rate than process $X_{b-}(t)$. In this case, the race is won by process $X_{a+}(t)$, which accumulates the required information in the shortest time

mapped to the right key, while $[b]$ and $[+]$ are mapped to the left key. When compared with the trial of the compatible block, the only difference is that the labels $[a]$ and $[b]$ swapped positions on the screen. As is shown in the diagram, the corresponding arrows are inverted accordingly so that the race, in this block, is between $X_{a-}(t) = X_a(t) + X_-(t)$ and $X_{b+}(t) = X_b(t) + X_+(t)$.

This inversion plays a fundamental role in the model and represents the basic mechanism of an important component of the IAT effect. Let us suppose that the presented stimulus belongs to category $[a]$. Then, $X_{a+}(t)$ is the process that provides the correct response (or simply, it is the *correct process*) in the compatible block, whereas the correct process in the incompatible blocks is $X_{a-}(t)$. By assuming identical termination criteria in the two blocks, if the condition $\lambda_{+a} > \lambda_{-a}$ holds true, the expected RT for the correct response will be longer in the incompatible blocks, when compared with the compatible blocks. This happens because the rate of the correct process in the incompatible blocks ($\lambda_{aa} + \lambda_{-a}$) is less than that of the correct process in the compatible blocks ($\lambda_{aa} + \lambda_{+a}$).

Additional components of the IAT effect: the termination criteria

In general, there is no reason for assuming identical termination criteria in the two critical blocks, and thus, they can be left free to vary across blocks (practice, compatible, or incompatible). Every single termination criterion is denoted by K_b , where b specifies the blocks (P = practice, C = compatible, I = incompatible). There are no additional restrictions on this set of three termination criteria. However, some specific hypotheses on how they are related to one another can be formulated. To begin with, let us suppose that a stimulus of category $[a]$ appears on the screen. If the condition $\lambda_{+a} = \lambda_{-a}$ holds true, the two correct processes $X_{a+}(t)$ (compatible blocks) and $X_{a-}(t)$ (incompatible blocks) will have identical rates. Nonetheless, an IAT effect will still be

observed whenever the condition $K_C < K_I$ is satisfied. This will happen because of the larger amount of information required by the correct process in the incompatible blocks, which, under the equality condition stated above, would have a longer expected RT. The positive difference $K_I - K_C$ is thus an additional component of the IAT effect. In general, when the practice blocks are considered as well, it seems reasonable to expect the following inequalities across the blocks:

$$K_P < K_C < K_I.$$

This inequality allows one to interpret the termination criteria as either task difficulty or individual cautiousness.

Separation between discrimination and association and model identifiability

Perhaps the most appealing feature of the DAM is that it is aimed at separating association from discrimination in the critical blocks. This separation is meaningful if the discrimination and association rate parameters can be uniquely determined. This is an aspect that cannot be taken for granted. In this respect, let us suppose that a stimulus of category $[a]$ is presented to a participant. Then the rates of the correct and wrong responses in the compatible blocks are $\lambda_{aa} + \lambda_{+a}$ and $\lambda_{ba} + \lambda_{-a}$, respectively, whereas in the incompatible blocks, we have $\lambda_{aa} + \lambda_{-a}$ and $\lambda_{ba} + \lambda_{+a}$. With no other restrictions on this set of four parameters, let us define an alternative parameter set $\{\lambda'_{aa}, \lambda'_{ba}, \lambda'_{+a}, \lambda'_{-a}\}$, such that $\lambda'_{aa} = \lambda_{aa} + c$, $\lambda'_{ba} = \lambda_{ba} + c$, $\lambda'_{+a} = \lambda_{+a} - c$, and $\lambda'_{-a} = \lambda_{-a} - c$ for some arbitrary nonnegative constant $c \leq \min\{\lambda_{-a}, \lambda_{+a}\}$. Then it can be easily seen that the following four equalities hold true:

$$\begin{aligned} \lambda'_{aa} + \lambda'_{+a} &= \lambda_{aa} + \lambda_{+a}, & \lambda'_{ba} + \lambda'_{-a} &= \lambda_{ba} + \lambda_{-a}, \\ \lambda'_{aa} + \lambda'_{-a} &= \lambda_{aa} + \lambda_{-a}, & \lambda'_{ba} + \lambda'_{+a} &= \lambda_{ba} + \lambda_{+a}, \end{aligned}$$

signifying that this new set of parameters will make exactly the same predictions as the original one. In other words, there exists a linear dependence among the four parameters at hand, and this is a clear sign of the fact that, with no further restrictions, the model would not be identifiable. To have a unique separation between discrimination and association parameters, we first observe that identification is restored whenever at least one of the four parameters at hand is not free anymore to vary within the critical blocks. Bearing this in mind, we note that quite reasonable assumptions can indeed be introduced that can solve the separation problem. They establish a special link between the critical and practice blocks. First, we observe that the average amount of information that a stimulus of category $[a]$ provides in the time unit concerning “being a stimulus of category $[a]$ ” should be

essentially the same in all blocks, including the practice blocks. This assumption implies that the correct discrimination rate λ_{aa} will be equal in both practice and critical blocks. Following an analogous reasoning, a similar assumption can be formulated with respect to the wrong discrimination rate λ_{ba} . These are key assumptions of the model; it is clear, in fact, that discrimination concerns stimuli and not blocks. Evidently, the plausibility of these assumptions can be tested empirically. Moreover, any of the two assumptions can solve the identification problem because, on the basis of the type of equality constraints it introduces, the (either correct or wrong) discrimination parameters of all blocks are uniquely determined in the practice block.

Each of the two assumptions introduced earlier might be criticized because, indeed, the type of discrimination task involved in the critical blocks is different from that involved in the practice blocks. However, it is important to realize that assuming equal discrimination rates across practice and critical blocks by no means implies that the discrimination process is the same in all such blocks. In fact, eventual differences between the two types of processes can still be captured by the different termination criteria K_B , K_C , K_I , and if the inequality $K_P < K_C < K_I$ proves to be empirically valid, the difference $K_C - K_P$ would provide a measure of the relative difficulty of discrimination in critical blocks, when compared with discrimination in the practice blocks.

An empirical application

Method

Participants A total of 199 psychology students at the University of Padua participated in the study on a voluntary basis. Their mean age was 23.66 years ($SD = 2.85$), and 122 were female.

Materials and procedure The participants were tested individually in a laboratory. First, they were presented with a Coca–Pepsi IAT, according to the structure represented in Table 2.

Table 2 Structure of the Coca–Pepsi IAT

Block	No. of trials	Left labels	Right labels
1 (Practice)	32	Coca Cola	Pepsi Cola
2 (Practice)	20	bad	Good
3 (Critical)	20	Coca Cola–bad	Pepsi Cola–good
4 (Critical)	36	Coca Cola–bad	Pepsi Cola–good
5 (Practice)	20	Pepsi Cola	Coca Cola
6 (Critical)	20	Pepsi Cola–bad	Coca Cola–good
7 (Critical)	36	Pepsi Cola–bad	Coca Cola–good

The Coca–Pepsi IAT used the category labels *Coca Cola*, *Pepsi Cola*, *good*, and *bad*. Ten color brand pictures were used to represent the object categories *Coca Cola* and *Pepsi Cola*, and 16 words were used to represent the attribute categories *good* (glory, good, happiness, joy, laughing, love, peace, pleasure) and *bad* (annoying, bad, evil, failure, hate, horrible, pain, terrible). The stimuli were presented in the center of the computer screen in an alternating fashion, and the participants were asked to categorize them by pressing, as quickly and accurately as possible, the response key “Q” or “P,” respectively.

Then participants were presented with a demographic questionnaire and with two dichotomous questions that asked them to indicate the tastiest cola and the most attractive brand, between Coca Cola and Pepsi Cola. At the end, the experimenter invited the participants to choose between a free can of Coca Cola or Pepsi Cola. The two cans were disposed on a table. The experimenter registered the choice of the participants after they had left the laboratory.

Analysis procedure Four vectors were obtained for each participant. The vectors specify the kind of block (*practice*, *Pepsi–bad/Coca–good*, *Coca–bad/Pepsi–good*), the category to which a stimulus belongs (*C–Coca Cola*, *P–Pepsi Cola*, *G–good*, *B–bad*), the accuracy of a response (1 = correct, 0 = incorrect), and the latency of the response in milliseconds. The length of the vectors was equal to the number of trials ($n = 184$). The four vectors represent the input of a MATLAB function for computing maximum likelihood estimates of the parameters of the DAM for every single participant. A software for estimating and testing the DAM is available upon request to the first author. Maximum likelihood estimation is sensitive to outliers and contaminants in the distributions of RTs. For this reason, following Klauer et al. (2007), trials whose latencies were outliers in the RT distribution were discarded according to Tukey’s criterion (see, e.g., Hoaglin, Mosteller, & Tukey, 1983). This led to the deletion of 6.92 % of all the responses.

The method for minimizing the negative log-likelihood function of the model was the BFGS optimization procedure (Broyden, 1970; Fletcher, 1970; Goldfarb, 1970; Shanno, 1970) with finite difference gradients. Similar to many other optimization methods, the BFGS is not guaranteed to converge to the global minimum of the negative log-likelihood function. The problem of local minima was tackled by applying the optimization procedure 100 times to the same observed data, each time starting from a different point of the parameter space. Among all different solutions, the one with the largest likelihood was retained. This reestimation procedure was applied to the data set of every single participant.

Another important issue that had to be solved in estimating parameters was the empirical identifiability of

the model. Even when a model is identifiable in theory, the data at hand might not be adequately informative for computing reliable estimates of one or more parameters in the model. As far as IAT data are concerned, this is especially true for the wrong discrimination parameters of the DAM. In fact, the discrimination task is usually easy, and the data of some participants might not contain any wrong responses in some blocks of the IAT (typically, the practice blocks). An observed frequency of such responses, greater than zero, is a necessary condition for estimating the wrong discrimination rates.

Irrespective of the missing information in the data, a formal test of the empirical identifiability of the model involves the computation of the Hessian matrix of the model's negative log-likelihood function at the point at which the latter attains its minimum (Gradshteyn & Ryzhik, 2000). If this matrix happens to be positive definite, the model is empirically identifiable. If not, one or more parameters are not identifiable. In this case, the optimization procedure still provides some estimate for those parameters, but the value obtained is meaningless and should thus be discarded. We note in passing that positive definiteness of the Hessian matrix also ensures that we are really on a minimum and not, for instance, on a saddle point or in a flat valley of the log-likelihood surface, far away from the minimum.

Empirical identification was tested for every single participant. In those cases where the Hessian matrix was not positive definite, the nonidentifiable parameters were not considered in the subsequent analyses. The procedure for detecting nonidentifiable parameters was as follows. The rank r of the Hessian matrix was computed. If r was less than the number m of the free parameters of the model, this was a symptom of a linear dependence among some of the rows (or, equivalently, columns) of the Hessian, indicating possible linear dependencies among the model's parameters. It is possible to detect parameters that are involved in linear dependencies by computing, for example, the reduced row echelon form of the Hessian matrix (see, e.g., Stefanutti, Heller, Anselmi, & Robusto, 2012, for a similar approach in a different context).

The model's goodness of fit was tested by applying the method described in the [Appendix](#).

Friedman's test and Wilcoxon's test for dependent sample were used for comparing the difficulty of the three kinds of blocks across the participants. We expected the practice blocks to be the easiest. In addition, given the implicit preference for Coca Cola relative to Pepsi Cola commonly observed in the literature (see, e.g., Maison, Greenwald, & Bruin, 2004; Sriram & Greenwald, 2009), we expected the *Pepsi-bad/Coca-good* blocks to be easier than the *Coca-bad/Pepsi-good* blocks. Friedman's

test was used for comparing the rates concerning correct and incorrect discrimination of each stimulus category. We expected the former to be greater than the latter.

Separate contrast measures, *DISC*, *ASSO*, and *DIFF*, were computed for the discrimination rates, association rates, and termination criteria, respectively. A *DISC* was computed for each stimulus category by taking the difference between the rate concerning the correct discrimination of the category with that concerning the incorrect discrimination (e.g., $DISC_C = \lambda_{CC} - \lambda_{PC}$). Positive values of the *DISC*s indicate that, on average, the stimuli provide more information, in the time unit, about their own category than about the contrasted one. An *ASSO* was computed for each stimulus category by contrasting its association rates. Positive values of $ASSO_C$ indicate that, on average, the *Coca Cola* stimuli provide more information, in the time unit, about the category *good* than about *bad* ($ASSO_C = \lambda_{GC} - \lambda_{BC}$). In other words, the *Coca Cola-good* association is stronger than the *Coca Cola-bad* association. Similarly, positive values of $ASSO_G$ indicate that the *good-Coca Cola* association is stronger than the *good-Pepsi Cola* association ($ASSO_G = \lambda_{CG} - \lambda_{PG}$). Interpretation of $ASSO_P$ and $ASSO_B$ is opposite to that of $ASSO_C$ and $ASSO_G$ ($ASSO_P = \lambda_{BP} - \lambda_{GP}$; $ASSO_B = \lambda_{PB} - \lambda_{CB}$). A *DIFF* was computed by contrasting the termination criteria of the critical blocks ($DIFF = K_{Coca-Bad/Pepsi-Good} - K_{Pepsi-Bad/Coca-Good}$). A positive value to *DIFF* indicates that the *Pepsi-bad/Coca-good* blocks are easier than the *Coca-bad/Pepsi-good* blocks. A regression analysis was run, in which the nine contrast measures were the independent variables and the *D* score was the dependent variable. The *D* is the most common measure of the IAT effect size. It involves dividing the difference in average response latency between the compatible and incompatible blocks by the standard deviation of latencies for all the critical blocks (for details, see *D2* in the work by Greenwald et al., 2003). *D* expresses the implicit preference for one cola relative to the other. We expected it to be predicted by *DIFF* and the contrast measures concerning the association rates *ASSOs*. Since *D* does not express the difference between correct and incorrect discrimination, it should not be predicted by the contrast measures *DISCs*.

The nine contrast measures were also the independent variables of three logistic regressions, in which taste preference, brand attractiveness, and choice of the cola were alternatively the dependent variable (0 denotes *Pepsi*, 1 denotes *Coca Cola*). We expected the three variables to be predicted by *DIFF* and the contrast measures *ASSOs*, but not by the *DISCs*. Three logistic regressions were also run in which the *D* score was the independent variable.

Results

The analysis presented in this section is restricted to 161 out of 199 participants whose model was accepted¹ (χ^2 p -value > .05), and 2.37 % of the λ parameters were removed from the analysis because they happened to be non-identifiable. All of them concerned only incorrect discrimination (i.e., λ_{PC} , λ_{CP} , λ_{BG} , and λ_{GB}). In the IAT, it is not unusual to observe participants who make only a few errors. When this happens, there might be no sufficient information for stable estimates of the λ parameters concerning incorrect discrimination. Out of the 161 participants, 38 presented at least one nonidentifiable parameter.

Table 3 provides the summary statistics of parameter estimates. The first subscript of the λ parameters refers to the process, whereas the second refers to the stimulus category. The letters *C*, *P*, *G*, and *B* denote *Coca Cola*, *Pepsi Cola*, *good*, and *bad*, respectively. The subscript of the *K* parameters denotes the kind of block (*practice*, *Pepsi–bad/Coca–good*, *Coca–bad/Pepsi–good*). Friedman’s test for dependent samples showed that the blocks were of different difficulty (mean rank = 1.00, 2.37, and 2.63 for *practice*, *Pepsi–bad/Coca–good*, and *Coca–bad/Pepsi–good* blocks, respectively), $\chi^2(2) = 246.72$, $p < .001$. Wilcoxon’s test highlighted that the *Pepsi–bad/Coca–good* blocks were more difficult than the *practice* blocks, $Z = -11.01$, $p < .001$, and less difficult than the *Coca–bad/Pepsi–good* blocks, $Z = -3.77$, $p < .001$. These results confirm our expectations that the practice blocks were the easiest and that the *Pepsi–bad/Coca–good* blocks were easier than the *Coca–bad/Pepsi–good* blocks.

The *D* score was computed for each participant. A positive *D* means that the stimuli are categorized faster in the *Pepsi–bad/Coca–good* blocks than in the *Coca–bad/Pepsi–good* blocks, indicating an implicit preference for Coca Cola relative to Pepsi Cola. Such a preference, although very small, was observed on the overall sample ($M = 0.07$, $SE = 0.03$), $t(160) = 2.35$, $p < .05$.

For each stimulus category, Wilcoxon’s test showed that the rate concerning the correct discrimination was greater than that concerning the incorrect discrimination ($\lambda_{CC} = 7.49$, $\lambda_{PC} = 3.00$, $Z = 10.37$, $p < .001$; $\lambda_{PP} = 7.21$, $\lambda_{CP} = 3.22$, $Z = 10.55$, $p < .001$; $\lambda_{GG} = 5.76$, $\lambda_{BG} = 1.99$, $Z = 10.27$, $p < .001$; $\lambda_{BB} = 5.48$, $\lambda_{GB} = 1.84$, $Z = 10.70$, $p < .001$; the values are the mean ranks across participants). This result confirms our expectation that the correct discrimination is

¹ We ran an analysis where the two assumptions concerning equality of the discrimination rates across the blocks were dropped. This resulted in a more general model with 24 λ parameters and 3*K* parameters. This model fit 22 additional participants, suggesting that, for them, equality constraints were violated. However, as was discussed in presenting the DAM, this alternative model is not capable of separating association rates from discrimination rates.

Table 3 Summary statistics of parameter estimates

Parameter	Min	Max	<i>M</i>	<i>SD</i>	Mean rank
Object discrimination rates					
λ_{CC}	3.16	69.03	21.72	11.79	7.49
λ_{CP}	<0.01	59.31	11.72	9.11	3.22
λ_{PC}	<0.01	46.26	10.69	8.12	3.00
λ_{PP}	2.61	71.10	21.26	11.53	7.21
Attribute discrimination rates					
λ_{GG}	3.00	63.14	18.66	10.51	5.76
λ_{GB}	<0.01	42.42	7.77	8.38	1.84
λ_{BG}	<0.01	42.93	7.95	7.68	1.99
λ_{BB}	3.15	62.26	18.20	10.40	5.48
Object-driven association rates					
λ_{GC}	<0.01	15.87	2.65	3.07	4.18
λ_{GP}	<0.01	13.16	2.72	2.93	4.13
λ_{BC}	<0.01	14.20	2.65	2.90	4.07
λ_{BP}	<0.01	17.08	2.44	2.90	3.80
Attribute-driven association rates					
λ_{CG}	<0.01	16.17	3.48	3.40	4.99
λ_{CB}	<0.01	14.63	3.81	3.54	5.25
λ_{PG}	<0.01	16.84	3.36	3.47	4.61
λ_{PB}	<0.01	21.25	3.56	3.70	4.97
Termination criteria					
$K_{Practice}$	4.58	81.04	22.35	11.77	1.00
$K_{Pepsi-Bad/Coca-Good}$	7.51	85.96	28.04	13.21	2.37
$K_{Coca-Bad/Pepsi-Good}$	9.04	93.33	29.00	14.19	2.63

The first subscript of the λ parameters refers to the process, whereas the second refers to the stimulus category. The letters *C*, *P*, *G*, and *B* denote *Coca Cola*, *Pepsi Cola*, *good*, and *bad*, respectively. The subscript of the *K* parameters denotes the kind of block

faster than the incorrect one. No significant differences were found among the mean ranks of the association rates across participants.

Let us now consider the results of the regression analyses. The analyses were performed on 123 participants for whom all parameters were identifiable. It is worth recalling that a positive value of *DIFF* indicates that the *Pepsi–bad/Coca–good* blocks are easier than the *Coca–bad/Pepsi–good* blocks. Positive values of the *DISCs* indicate that, on average, the stimuli provide more information, in the time unit, about their own category than about the contrasted one. Positive values of *ASSO_C* indicate that the *Coca Cola–good* association is stronger than the *Coca Cola–bad* association, and positive values of *ASSO_G* indicate that the *good–Coca Cola* association is stronger than the *good–Pepsi Cola* association. Interpretation of *ASSO_P* and *ASSO_B* is opposite to that of *ASSO_C* and *ASSO_G*. Confirming our expectation, the *D* score was predicted by *DIFF* and by all the four contrast measures concerning the association rates *ASSO_C*, *ASSO_P*, *ASSO_G*, and *ASSO_B* (Table 4). These five measures altogether

Table 4 Regression of *D* score on the nine contrast measures

Contrast measure	β	<i>t</i>	<i>p</i>
$DISC_C = \lambda_{CC} - \lambda_{PC}$	-0.14	-1.78	.08
$DISC_P = \lambda_{PP} - \lambda_{CP}$	0.08	1.14	.26
$DISC_G = \lambda_{GG} - \lambda_{BG}$	-0.05	-0.57	.57
$DISC_B = \lambda_{BB} - \lambda_{GB}$	-0.06	-0.78	.44
$ASSO_C = \lambda_{GC} - \lambda_{BC}$	0.39	6.00	<.001
$ASSO_P = \lambda_{BP} - \lambda_{GP}$	0.32	5.35	<.001
$ASSO_G = \lambda_{CG} - \lambda_{PG}$	0.35	4.93	<.001
$ASSO_B = \lambda_{PB} - \lambda_{CB}$	0.14	2.04	<.05
$DIFF = K_{Coca-Bad/Pepsi-Good} - K_{Pepsi-Bad/Coca-Good}$	1.01	14.45	<.001

$R^2 = .66$; $AdjR^2 = .65$, for the five significant contrast measures

accounted for 65 % of the variance of *D* ($R^2 = .66$; $AdjR^2 = .65$), and none of the measures showed multicollinearity (tolerance ≥ 0.40).

The choice of the cola was predicted by *DIFF* and *ASSO_P*, and the taste preference was predicted by *DIFF* (Table 5). None of the contrast measures predicted the brand attractiveness. These results confirm our expectations only partially, indicating that *DIFF* is an important predictor of the choice of the cola and taste preference. All three variables were predicted by the *D* score ($\beta = 2.26, p < .001$ for choice of the cola; $\beta = 2.37, p < .001$ for taste preference; $\beta = 1.87, p < .01$ for brand attractiveness).

Case studies Figure 3 provides four fictitious cases of how association rates might highlight automatic associations that differ in nature and meaning. An “object-driven association” indicating an implicit preference for Coca Cola relative to Pepsi Cola is observed when $\lambda_{GC} > \lambda_{BC}$ and $\lambda_{BP} > \lambda_{GP}$ (Fig. 3a). This means that the *Coca-good* association is stronger than the *Coca-bad* association and the *Pepsi-bad*

association is stronger than the *Pepsi-good* association. An “attribute-driven association” indicating an implicit preference for the same cola is observed when $\lambda_{CG} > \lambda_{PG}$ and $\lambda_{PB} > \lambda_{CB}$ (Fig. 3b). An “object-driven association” and an “attribute-driven association” indicating an implicit preference for Pepsi Cola relative to Coca Cola are observed when $\lambda_{GP} > \lambda_{BP}$ and $\lambda_{BC} > \lambda_{GC}$ (Fig. 3c) and when $\lambda_{PG} > \lambda_{CG}$ and $\lambda_{CB} > \lambda_{PB}$ (Fig. 3d), respectively.

Figure 4 depicts the standardized association rates (obtained by dividing each association rate by its standard error) of 7 participants in our study. They are exemplary cases of the profiles of automatic association that have been observed in the present application. Unbroken arrows represent significant rates. The *D* scores of these participants are also provided. It is worth recalling that a positive *D* indicates an implicit preference for Coca Cola relative to Pepsi Cola, whereas a negative *D* indicates the opposite.

An object-driven and an attribute-driven association indicating a preference for Coca Cola relative to Pepsi Cola are observed in participants *a* and *b*, respectively (Fig. 4a, b). Exemplars of object-driven and attribute-driven associations indicating a preference for the other cola could not be observed in the present sample. An association is observed in participant *c*, which can be regarded as a partial attribute-driven association because only the *bad* stimuli contribute to the measure (Fig. 4c). Interestingly, the *D* scores of participants *b* and *c* are similar in size but different in meaning. Participant *c* associates only the *bad* stimuli with Pepsi Cola, whereas participant *b* also associates the *good* stimuli with Coca Cola. The implicit preference for Coca Cola relative to Pepsi Cola observed in participant *c* expresses only the attribution of negative features to the Pepsi Cola, whereas that observed in participant *b* also exhibits the attribution of positive features to the Coca Cola. Other exemplars in which the *D* scores are similar in size but different in meaning could be observed by comparing participants *d* and

Table 5 Regression of cola choice, taste preference, and brand attractiveness on the nine contrast measures

Contrast measure	Cola choice			Taste preference			Brand attractiveness		
	β	Wald	<i>p</i>	β	Wald	<i>p</i>	β	Wald	<i>p</i>
$DISC_C = \lambda_{CC} - \lambda_{PC}$.00	.00	.95	-.06	.72	.40	.05	.35	.56
$DISC_P = \lambda_{PP} - \lambda_{CP}$.01	.04	.83	-.02	.05	.82	.05	.38	.54
$DISC_G = \lambda_{GG} - \lambda_{BG}$.02	.10	.75	.10	1.50	.22	-.00	.00	.99
$DISC_B = \lambda_{BB} - \lambda_{GB}$	-.03	.33	.57	.00	.00	.99	-.02	.16	.69
$ASSO_C = \lambda_{GC} - \lambda_{BC}$.03	.09	.76	.07	.44	.51	.13	1.20	.27
$ASSO_P = \lambda_{BP} - \lambda_{GP}$.24	4.51	.03	.10	.62	.43	.15	1.02	.31
$ASSO_G = \lambda_{CG} - \lambda_{PG}$.16	1.89	.17	.05	.14	.71	.18	1.37	.24
$ASSO_B = \lambda_{PB} - \lambda_{CB}$	-.07	.38	.54	.19	1.90	.17	.16	1.14	.29
<i>DIFF</i>	.21	3.95	.05	.28	5.20	.02	.12	1.00	.32

Degrees of freedom of the Wald statistic are equal to 1. $DIFF = K_{Coca-Bad/Pepsi-Good} - K_{Pepsi-Bad/Coca-Good}$

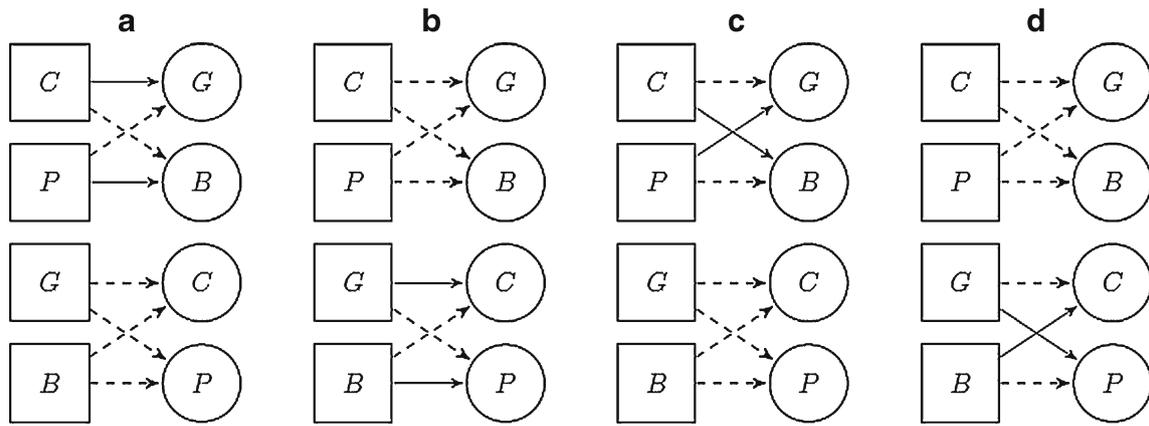


Fig. 3 Fictitious cases of automatic associations. Squares represent stimuli, circles represent processes, and arrows represent association rates. *C* = Coca Cola, *P* = Pepsi Cola, *G* = good, *B* = bad. Unbroken arrows represent the strongest association rates

e (Fig. 4d, e) and participants *f* and *g* (Fig. 4f, g). The implicit preference for Coca Cola relative to Pepsi Cola observed in

participant *d* results from associating the *Coca Cola* stimuli with *good*, whereas that observed in participant *e* results from

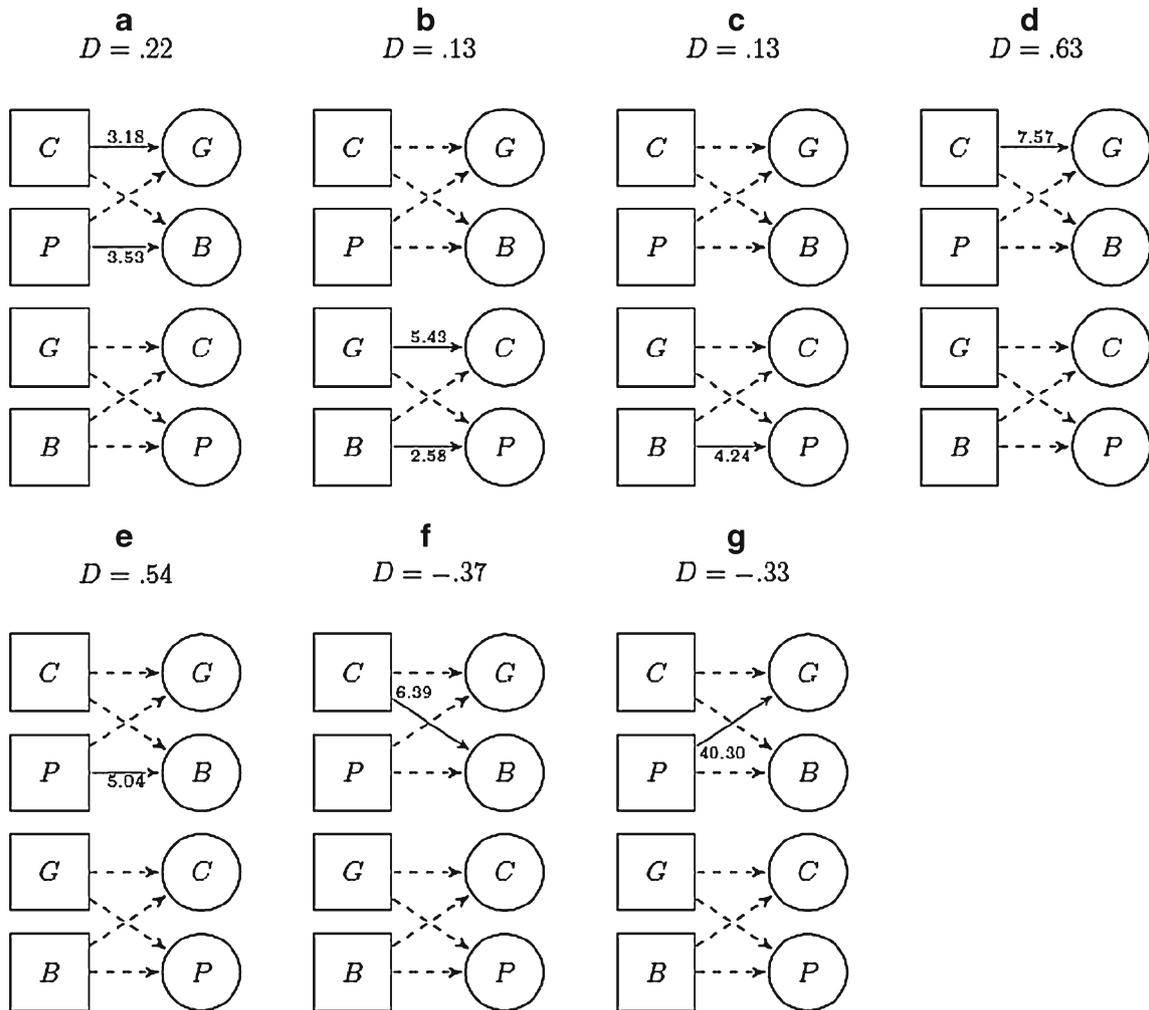


Fig. 4 Standardized association rates and *D* scores of 7 participants in the study. Squares represent stimuli, circles represent processes, and arrows represent association rates. *C* = Coca Cola, *P* = Pepsi Cola, *G* =

good, *B* = bad. Unbroken arrows represent significant association rates ($p < .05$; Bonferroni correction)

associating the *Pepsi Cola* stimuli with *bad*. The implicit preference for Pepsi Cola relative to Coca Cola observed in participant *f* results from associating the *Coca Cola* stimuli with *bad*, whereas that observed in participant *g* results from associating the *Pepsi Cola* stimuli with *good*.

Discussion

A formal model has been presented that decomposes the IAT effect into three process components: stimuli discrimination, automatic association, and termination criterion. The model has been applied to the responses provided to a Coca–Pepsi IAT.

Stimuli discrimination provides information about the functioning of the stimuli. The discrimination rates enable the researchers to explore whether the stimuli that have been chosen to represent a category can be easily recognized and correctly categorized. In our application, the pictures representing the categories *Coca Cola* and *Pepsi Cola* and the words representing the categories *good* and *bad* were highly recognizable.

Automatic association provides information about the association between objects and attributes. The association rates enable the researchers to distinguish between implicit measures that have the same direction but different meanings. In our study, we have been able to distinguish participants whose implicit preference for a certain cola resulted from a positive evaluation of that cola from those whose implicit preference resulted from a negative evaluation of the contrasting cola. This has a strong potential in some research fields. In the investigation of implicit race attitudes, for example, researchers would be allowed to distinguish individuals with an actual implicit prejudice (e.g., *Black–bad* association in a white participant) from those who merely hold an implicit preference for their own group (e.g., *White–good* association in a white participant). Since these parameters express the average amount of information that the object (attribute) stimuli provide about the attribute (object) categories in the time unit, they have been denoted as *association rates*. However, it is worth noting that, depending on the application context, they might reflect effects that are not based on associations, but on a recoding of the object and attribute categories on the basis of shared features (e.g., salience, valence, shape, color; Greenwald, Nosek, Banaji, & Klauer, 2005; Rothermund & Wentura, 2004).

Termination criteria represent the amount of information that is needed before a response is given. They can be interpreted as either task difficulty or individual cautiousness.

The model enables a fine-grained analysis of the IAT effect. It is interesting to note that in our Coca–Pepsi IAT, the *D* score mostly reflected the different difficulty in categorizing the stimuli in the critical blocks and the different strength of object–attribute associations. The different difficulty of the

categorization tasks also predicted taste preference and the choice of a free can of cola.

There are some connections between the DAM and the DM. Since the DAM is an extension of the Poisson race model, these connections are essentially the same as those described by Van Zandt, Colonius, and Proctor (2000). Both of the models disentangle components of the IAT effect concerning the rate at which information about the stimulus is accumulated (drift in the DM, stimuli discrimination and automatic association in the DAM) and the amount of information that must be accumulated before a response is given (threshold in the DM, termination criterion in the DAM). However, there are several differences between the two models that might enable the DAM to be a good alternative to the DM. First, in the presented model, only termination criteria vary across blocks, whereas in the DM analysis, both drifts and thresholds vary across blocks. This might confound their relative contribution when compatible and incompatible blocks are compared. The DAM allows estimating accumulation rates separately for objects and attributes by using the number of trials of a typical IAT. On the contrary, a DM analysis, separated for objects and attributes, requires a number of trials, which is almost 4 times greater than usual (see Klauer et al., 2007, Study 1). In the DM, information accumulation of contrasting information is perfectly negatively correlated, whereas, in the DAM, it is assumed to be independent. These two approaches are based on the assumptions of serial and parallel processing of the stimuli, respectively. Finally, the DAM derives from a model (the Poisson race model) that is formally simpler than the DM, and therefore, it is easier to investigate its mathematical properties.

Limits and future research It is not atypical to observe participants who make no or few errors in one or more of the IAT blocks. When this happens, there might be no sufficient information for obtaining stable estimates of all model parameters. For example, in the DAM, with no wrong responses to the stimuli of a specific category, it would be impossible to estimate the λ parameter concerning their incorrect discrimination. Similarly, no or few wrong responses represent a limit for the quad model and the DM. An advantage of the DAM is that it considers separate processes for correct and incorrect responses. Hence, even if the estimates of the parameters concerning wrong responses are not reliable, this would not be the case for those concerning correct responses.

The aim of the empirical application presented in this study was to show how the DAM can be applied to the analysis of IAT data and what information it can provide about the response process. In this first application of the model, it was decided to present all participants with the IAT blocks in the same order for comparability reasons. The

Pepsi–bad/Coca–good blocks always preceded the *Coca–bad/Pepsi–good* blocks. The fixed sequence of the critical blocks might have influenced the size of the IAT effect (see, e.g., Nosek, Greenwald, & Banaji, 2005). However, it should be considered that the aim of this study was to illustrate the usefulness and practical potential of the DAM, and not to draw conclusions about implicit preferences toward one cola over the other.

As far as the interpretation of the parameters of the DAM is concerned, it is worth underlining that they still need to be validated. As was already noted, the empirical application that we presented in this article is merely a demonstration of the potential usefulness of the DAM. For instance, the results we presented do not provide any clear evidence that the lambda parameters can be interpreted as association strengths. There are many other plausible interpretations of the association parameters of the DAM, and an exhaustive set of studies including manipulations, test–retest investigations, and known groups is needed to establish their validity.

One further limit is that the regression analyses presented in this study were performed on 123 participants, those for whom all parameters were identifiable. Future research will be devoted to developing procedures for reducing the proportion of nonidentifiable parameters.

Also, the model in its current specification does not allow the study of the effects of single stimuli but estimates only the effects of stimulus categories. Previous research has described that stimuli influence the IAT effect (see, e.g., Bluemke & Friese, 2006; Gast & Rothermund, 2010; Govan & Williams, 2004; Steffens & Plewe, 2001). A possibility for taking the effects of single stimuli into account might consist of extending the model into a hierarchical framework (see, e.g., Vandekerckhove, Tuerlinckx, & Lee, 2011).

The aim of future research will also be to improve the model fit. Response latencies of objects and attributes in the different blocks, as well as the amount of information concerning incorrect responses, will be considered. A possibility for enlarging the number of fitting models might consist of modeling intertrial variability to account for the differences between the stimuli of a certain category and the effects of passing time on responses to the trials (e.g., training to the task, fatigue). Other possibilities might be the incorporation of nondecision components (e.g., encoding stimuli, motor response, distractions) within the model, and allowing the termination criteria to vary between objects and attributes.

Conclusions

In this article, a formal model has been proposed that decomposes the IAT effect into three process components: stimuli discrimination, automatic association, and termination

criterion. Discrimination regards the amount of information that an object (attribute) stimulus provides about the object (attribute) categories. Association regards the amount of information that an object (attribute) stimulus provides about the attribute (object) categories. The DAM has been tailored to fit the IAT, and it provides a number of desirable features. By means of an illustrative application, we explained why and how, once fully validated, the model might help to shed light on the meaning of the IAT effect, providing, for instance, estimates of many unique associations that are typically hidden behind every IAT score.

Appendix

Method for testing the goodness of fit of the DAM

The method used for testing the goodness-of-fit of the DAM is based on the procedure described in the study by Klauer et al. (2007). Let T be a continuous random variable, whose realizations are RTs for both correct and wrong responses, with those of the wrong responses multiplied by -1 . The method consists of providing a discrete approximation to both the expected and observed frequency distributions of T , for which a chi-square distributed goodness-of-fit statistic can be constructed.

To this aim, the real line is partitioned into a number of k intervals, where k is chosen according to Moore (1986). Then, the observed data points $t_i, i = 1, \dots, n$ (realizations of T in a finite sample of size n) are ordered from the smallest to the largest (so that $t_i < t_{i+1}$), and the empirical cumulative distribution of such data points is considered. The interval $[0, 1]$ of this cumulative distribution is then partitioned into a number of k intervals, each of which has upper bound $\lambda_i = i/k$, and the largest number $n_i \leq n\lambda_i + 1$ is considered for each λ_i . Denoting the quantile of order n_i as t_{n_i} , the following statistic is defined:

$$Y^2(\theta) = n \sum_{i=1}^k \{ [F(t_{n_i}|\theta) - F(t_{n_{i-1}}|\theta)] - p_i \}^2 / p_i$$

where $F(t|\theta)$ is the cumulative probability of T predicted by the model with parameters θ and $P_i = \lambda_i - \lambda_{i-1} = 1/k$. If $\hat{\theta}$ is an estimator that minimizes $Y^2(\theta)$, then the statistic $Y^2(\hat{\theta})$ is asymptotically chi-squared distributed, with $k - q - 1$ degrees of freedom, where q is the number of free parameters.

For a single participant, the model predicts a different $F(t|\theta)$ for each of the three kinds of blocks (practice, compatible, incompatible) of the IAT. Therefore, three different values of $Y^2(\theta)$ were computed, one for each block type. As an overall goodness-of-fit statistic of the model for a single participant, the sum $Y_P^2(\theta) + Y_C^2(\theta) + Y_I^2(\theta)$ was computed, where P , C , and I denote practice, compatible, and incompatible blocks,

respectively. For the general property of the sum of chi-square distributed random variables, this sum is also asymptotically chi-square distributed with $k_P + k_C + k_I - m - 3$ degrees of freedom, where k_P , k_C , k_I are the (possibly different) number of intervals in the three block types and m is the total number of free parameters in the model.

References

- Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(2), 234–254.
- Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, 42, 163–176.
- Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications*, 6, 76–90.
- Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, 89(4), 469–487.
- Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). When categories collide: Accumulation of information about multiple categories in rapid scene perception. *Psychological Science*, 22, 739–746.
- Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13, 317–322.
- Gast, A., & Rothermund, K. (2010). When old and frail is not the same: Dissociating category and stimulus effects in four implicit attitude measurement methods. *The Quarterly Journal of Experimental Psychology*, 63(3), 479–498.
- Goldfarb, D. (1970). A family of variable metric updates derived by variational means. *Mathematics of Computing*, 24, 23–26.
- 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.
- Gradshteyn, I. S., & Ryzhik, I. M. (2000). *Table of integrals, series and products* (6th ed.). New York, NY: Academic Press.
- 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(6), 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(2), 197–216.
- Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the IAT: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General*, 134, 420–425.
- Hoaglin, D. C., Mosteller, F., & Tukey, J. W. (Eds.). (1983). *Understanding robust and exploratory data analysis*. New York, NY: John Wiley & Sons.
- 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(3), 353–368.
- Maison, D., Greenwald, A. G., & Bruin, R. (2004). Predictive validity of the Implicit Association Test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, 14(4), 405–415.
- Moore, D. S. (1986). Tests of chi-squared type. In R. B. D'Agostino & M. A. Stephens (Eds.), *Goodness-of-fit techniques* (pp. 63–95). New York: Marcel Dekker.
- 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(2), 166–180.
- Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, 133, 139–165.
- Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computing*, 24, 647–656.
- Sriram, N., & Greenwald, A. G. (2009). The brief Implicit Association Test. *Experimental Psychology*, 56(4), 283–294.
- Stefanutti, L., Heller, J., Anselmi, P., & Robusto, E., (2012). Assessing the local identifiability of probabilistic knowledge structures. *Behavior Research Methods*. doi:10.3758/s13428-012-0187-z
- Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie*, 48(2), 123–134.
- Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, 114(1), 71–103.
- Townsend, J. T. (1990). Serial vs parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science*, 1, 46–54.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. New York, NY: Cambridge University Press.
- van der Heijden, A. H. C. (1975). Some evidence for a limited capacity parallel self-terminating process in simple visual search tasks. *Acta Psychologica*, 39, 21–41.
- Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two reaction time models applied to perceptual matching. *Psychonomic Bulletin and Review*, 7(2), 208–256.
- Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, 16(1), 44–62.