Social Cognition, Vol. 30, No. 5, 2012, pp. 610–630

ASSESSING IMPLICIT ATTITUDES: WHAT CAN BE LEARNED FROM SIMULATIONS?

Boon-Kiat Quek and Andrew Ortony
Northwestern University and Agency for Science, Technology and Research, Singapore

Empirical studies involving the Implicit Association Test (IAT) have often revealed a general implicit preference for European-American as opposed to African-American stimuli. While it has been pointed out that this does not establish the existence of an implicit negative attitude toward the less preferred target concept, the existence (or absence) of such attitudes is empirically difficult to ascertain. We describe a computational model of performance on the Race-IAT through which the influence of attitudinal positivity or negativity on expected IAT performance is explored using simulations in which the strengths and nature of target-attribute associations in memory are manipulated. Results indicate that IAT effects readily emerge from different patterns of implicit associations without any need for absolutely positive or absolutely negative implicit attitudes. Pitting a simulation of the standard IAT against a simulated Sorting Paired Features Task demonstrates an advantage for the latter in distinguishing each of the implied target-attribute associations.

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) is a computer-administered test designed to assess automatic associations between concepts, where such associations are assumed to underlie implicit attitudes[1] to-

1. While some (e.g., Gawronski, Hofmann, & Wilbur, 2006) have cautioned against claiming that implicit attitudes are inaccessible to consciousness, it is presumably true that what makes such attitudes implicit as opposed to explicit is that people who hold them do not know that they hold them. To be sure, people might sometimes be able to abductively infer that they hold a particular implicit attitude by observing their own behavior, but that is a different point; they still do not have direct access to that attitude. So, we maintain that such attitudes cannot be reliably revealed by explicit measures such as self-reports.

The authors wish to thank Miguel Brendl, Gregory Dam, Mesmin Destin, and William Revelle for their valuable comments and feedback. Boon-Kiat Quek is supported by a postdoctoral fellowship awarded by the Agency for Science, Technology and Research (A*STAR), Singapore.

Correspondence concerning this article should be addressed to Boon-Kiat Quek, Institute of High Performance Computing, Agency for Science, Technology and Research, 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore; E-mail: quekbk@ihpc.a-star.edu.sg.

ward or beliefs about attitude objects (Greenwald & Banaji, 1995; Greenwald et al., 2002). The test measures subjects' response times on a number of classification tasks which involve pressing a *left* or *right* key on a keyboard to classify sequentially presented stimuli (words or pictures) into one of two categories, each comprising a target concept (e.g., *flower* vs. *insect*) paired with an attribute concept (e.g., *pleasant* vs. *unpleasant*). The IAT relies on the assumption that association strengths between an attitude object and its attributes reflect the strength of the implicit attitude toward that object (Neumann & Seibt, 2001). Thus, a person with a negative attitude toward insects is taken to have a strong association between the concept of *insect* and some general evaluative attribute for negativity. The strengths of these associations are held to be largely responsible for people's response times on the classification tasks, with longer latencies for incompatible disjunctive category pairs (e.g., "*insect or pleasant*," "*flower or unpleasant*") than for compatible pairs (e.g., "*flower or pleasant*," "*insect or unpleasant*"). The mean difference in response times between trials for incompatible and compatible pairings is known as the *IAT effect*.

Of various applications of the IAT, the Race-IAT is particularly controversial because empirical results seem to suggest a prevalent implicit prejudice in favor of European-Americans over African-Americans (Greenwald et al., 1998), given that longer response latencies were consistently observed in trials involving the category pairs "*European-American or unpleasant*" and "*African-American or pleasant*" than for trials involving "*European-American or pleasant*" and "*African-American or unpleasant*" (Frantz, Cuddy, Burnett, Ray, & Hart, 2004; Greenwald et al., 1998; Karpinski & Hilton, 2001; Nosek, Banaji, & Greenwald, 2002). While it might be tempting to interpret this as evidence for an implicit negative attitude toward the less preferred target concept, it is certainly not the only explanation (Brendl, Markman, & Messner, 2001; Greenwald et al., 1998). For example, it might well be that both target concepts are positively evaluated, but with one more positive than the other, or conversely, that both are negatively evaluated, but with one more negative than the other. However, it is unclear how one would test this empirically (if at all) because there is no direct way of accessing either the qualitative (i.e., valence) or quantitative (i.e., strength) aspects of implicit attitudes (apart from using some implicit measure, which begs the question), and no practical way of systematically manipulating them in experiments. An experimenter could not, for instance, arrange for certain subjects to possess only positive (as opposed to only negative) evaluations for a pre-existing attitude object, nor control the degree to which subjects might endorse a certain attitude toward that object (e.g., preferring one target *twice as much as* another). It is possible that one might be able to manipulate the valence and strength of implicit attitudes with classical conditioning (Olson & Fazio, 2001; Olson & Fazio, 2002) by arranging for novel associations between previously unknown attitude objects to be formed through frequent pairing with other strongly valenced stimuli. However, we would still be unable to manipulate pre-existing associations such as implicit racial attitudes on the Race-IAT.

On the other hand, if a computational model can be shown to replicate IAT effects, it might make a suitable platform for conducting virtual experiments to explore a range of issues relating to IAT performance. We have developed such a model (Quek & Ortony, 2011)—a model that incorporates representations of concepts and their associative relationships in memory in an associative network. By altering the strengths, nature (excitatory or inhibitory), and directions of associations between concept representations, the model supports the sort of attitudinal manipulations necessary for conducting simulated experiments.

While the computational model was initially developed to explore different mechanisms underlying human performance on the IAT, our focus in this article is on its utility for performing virtual experiments. Whereas one of our goals is to show that IAT effects can emerge from various other configurations of associations, and not just those that might be taken to imply the presence of negative attitudes, we are also interested in exploring various other factors that might be contributing to performance on the IAT (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; Mierke & Klauer, 2003). Such factors include differences in stimulus familiarity (Dasgupta, McGhee, Greenwald, & Banaji, 2000; Ottaway, Hayden, & Oakes, 2001), salience asymmetries between target concepts (Rothermund & Wentura, 2004), extra-personal knowledge about prevailing cultural norms (Karpinski & Hilton, 2001), cognitive fluency (McFarland & Crouch, 2002), and the presence of stereotype threat (Frantz et al., 2004). In addition, subjects have been known to employ various strategies (which they might discover spontaneously as they go through the IAT) that facilitate their responses, such as focusing on stimuli that are instances of just the two concepts in the left combined category, while keeping in mind a mental note to relegate all other non-focal stimuli to the *right* response (Sriram & Greenwald, 2009). These factors and strategies would need to be taken into account in the case of an actual experiment involving human subjects, whereas in the case of a simulation, they could be isolated from the main influence of the automatic associations being studied. In light of the above, we first set out to simulate the influence of different configurations of associations, and only then to extend the model to explore the effects of some of the other proposed factors and strategies, as well as other sources of method-specific variance on the IAT.

MODELING AND SIMULATION OVERVIEW

The computational model employs a localist connectionist network (see, e.g., Page, 2000) that emulates the multiple processing pathways beginning from the initial perception of a visual stimulus (in our case, a word or image) to the automatic activation of associated concepts in memory, and from these concepts to a motor response (i.e., a *left* or *right* key-press). Nodes in this network represent semantically meaningful entities (e.g., input stimuli, target concepts and their associated attributes, action tendencies, and motor actions), while connections between nodes encode the strength (magnitude), nature (excitatory or inhibitory), and direction of the associations. The network can be represented formally as a graph $G = \{V, E\}$ with nodes V and connections or edges, E. For simplicity, each node v_i is defined as a tuple comprising a semantically meaningful label or meme $name_i$ representing a particular concept, and a quantity x_i representing its activation level:

$$v_i = \langle name_i\, , x_i \rangle;\ v_i \in V;\ x_i \in [-1, 1]. \tag{1}$$

Similarly, connections or edges between nodes are defined as:

$$\varepsilon_{i,j} = \langle v_i\, , v_j\, , w_{i,j} \rangle;\ \varepsilon_{i,j} \in E\ ;\ v_i\, , v_j \in V\ ;\ w_{i,j} \in [-1, 1], \tag{2}$$

where $w_{i,j}$ is the connection weight or strength from node v_i to v_j. Unless otherwise stated, weights are initialized to either 0.5 or -0.5, depending on whether the associations are meant to be excitatory or inhibitory, prior to the application of a random perturbation factor (see below). Information is propagated through the network over successive iterations via a propagation rule that is defined as:

$$x_i(k+1) = (1-\delta)x_i(k) + \alpha \sum_{\varepsilon_{j,i} \in E} x_j(k) \cdot w_{j,i}(k)\,; \delta \in [0,1], \alpha \in [0,1]\,, \tag{3}$$

where α is the gain (set to 0.2) and δ is a decay parameter (set to 0.001) that reduces activation over time. We selected parameter values in these ranges to ensure that activation levels do not saturate prematurely and yet have sufficient momentum to accumulate. In this way, activation spreads to v_i from each neighbor v_j at a rate proportional to $w_{j,i}$ (which can be thought of as the *conductance* of the connection) in each time step, and in the absence of input, activation levels of inactive nodes will gradually recede back to zero. In the course of our work, we have experimented with various alternative forms of equation (3), and found them all to work comparably (provided activation increases monotonically with the summation term), so the choice of the eventually deployed algorithm is less critical than its topology.

SIMULATING VIRTUAL SUBJECTS

Each virtual subject model has three components: an *Associative Network*, a *Task Mapper* and a *Response Generator* (see Figure 1), all of which are represented in the network architecture defined above. The *Associative Network* comprises nodes that represent target and attribute concepts, and edges representing associations between them. The *Task Mapper* dynamically transmits activation levels of concepts from the *Associative Network* to the *Response Generator*, which generates one of the two motor responses (i.e., the simulated activity of actually pressing a key).

Each virtual subject is represented by a network of the described topology, but with the initial connection weights randomly perturbed with Gaussian noise of ~$N(0, 0.1^2)$ so as to produce a varied sample of simulated subjects. With an initial weight $w_{j,i}$ of ±0.5, these perturbations result in a majority (95%) of connections across subject populations having weights in the [0.3, 0.7] or [-0.7, -0.3] range. These random perturbations were introduced to ensure that the data generated from groups of simulated subjects were not dependent on specific values of associative strengths.

MAPPING STIMULI TO CONCEPTS

The standard input stimuli for the Race-IAT are names or pictures of (presumed-to-be unfamiliar) European-American or African-American individuals, and words belonging to the semantic fields *pleasant* or *good* (e.g., *happy, wonderful, joy*), and *unpleasant* or *bad* (e.g., *evil, horrible, hurt*). The nodes representing stimulus pictures or words have a directed connection to nodes representing their corresponding target *concepts*—AFRICAN-AMERICAN (AA) and EUROPEAN-AMERICAN (EA), and generic evaluative attributes for positivity (POS) and negativity (NEG). So, for instance, the presentation of a *European-American* stimulus activates a concept node

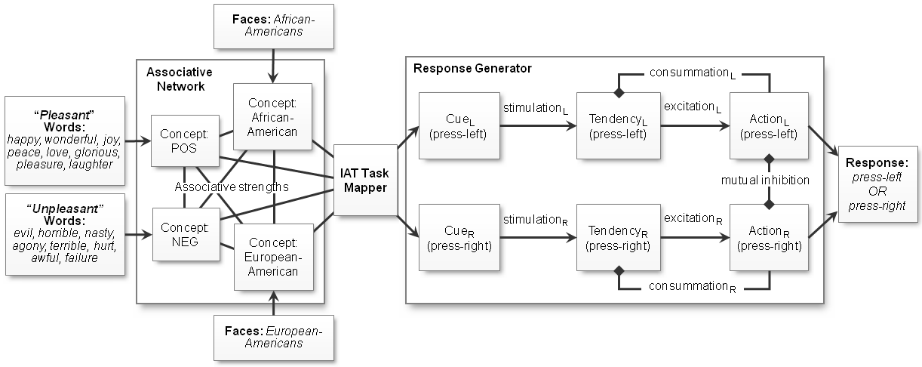

FIGURE 1. Schematic of computational model for simulating virtual subjects' performance on the Implicit Association Test.

corresponding to a European-American individual, which in turn activates the concept EA, and the presentation of the word *wonderful* activates the concept node WONDERFUL, which in turn increases the activation level of POS. In all cases, the path lengths from input nodes to EA, AA, POS or NEG are the same.

ASSOCIATIONS BETWEEN TARGET AND ATTRIBUTE CONCEPTS

The implicit associations between target and attribute concepts are represented as connections[2] between all four target-attribute pairs, namely EA↔POS, EA↔NEG, AA↔POS, and AA↔NEG. By manipulating these connections, considerable freedom in the configuration of associations is afforded (as outlined in Table 1 for the simulations in the rest of this article). For instance, a positive attitude toward AA could be represented as an excitatory AA↔POS association, or an inhibitory AA↔NEG association, or both, such that activation of the AA concept would increase the activation of POS, but reduce the activation of NEG.

MAPPING CONCEPTS TO RESPONSE CUES FOR DIFFERENT TASKS

The *Task Mapper* functions as a network switchboard that dynamically transmits accumulated activation from the target concepts and their evaluative attributes to the nodes cue_L and cue_R—nodes that indicate that a response is required. For example, if the current task requires a *left* key-press for "*African-American* or *pleasant*," and a *right* key-press for "*European-American* or *unpleasant*," the *Task Mapper* would establish connections from the concept nodes AA and POS to cue_L and from EA and NEG to cue_R. The assigned connections remain active throughout each block of trials, but are reconfigured for each of the subsequent task blocks, as shown in Table 2. In this way, the dynamically formed connections represent

2. For simplicity, these associations are assumed to be bi-directional, and implemented as a pair of directed edges in opposite directions (e.g., EA→POS is implemented as two connections, EA→POS and POS→EA).

TABLE 1. Configurations of Associations for Experimental and Control Conditions

Conditions	Strength of association between target and attribute concepts			
	EA↔POS	EA↔NEG	AA↔POS	AA↔NEG
Simulation 1				
a. Like EA, Dislike AA	+.5	-.5	-.5	+.5
b. Like AA, Dislike EA	-.5	+.5	+.5	-.5
c. Like/Dislike Both	+.5	+.5	+.5	+.5
d. Like/Dislike Neither	0	0	0	0
Simulation 2				
e. Like EA more than AA	+.8	0	+.2	0
f. Like AA more than EA	+.2	0	+.8	0
g. Dislike EA more than AA	0	+.8	0	+.2
h. Dislike AA more than EA	0	+.2	0	+.8

Note. EA: European-American; AA: African-American; POS: Positivity; NEG: Negativity. Numbers shown are initial values of weights that are further perturbed with Gaussian noise for each virtual subject so as to generate a sample of unique networks representing different individuals. Configurations (c) and (d) are control conditions.

subjects' awareness of and adherence to the instructions for each task block and the temporary association of concept categories to the *left* and *right* response cues. To some extent, this reconfiguration process is an emulation of task-set switching (Klauer & Mierke, 2005), and is compatible with the suggestion that dynamically switching connections from one task block to another causes a temporary association (i.e., *acquired meaning*, see De Houwer, 2001) of positive or negative valence to a *left* or *right* response.

RESPONSE GENERATION

The *Response Generator* part of the network is an implementation of the cue-tendency-action (CTA) model (Revelle, 1986), which is itself a re-parameterization of Atkinson and Birch's (1970) dynamics of action theory. The CTA model captures the dynamic interactions between conflicting action tendencies and incompatible actions, as well as the inertial properties of action tendencies and actions themselves. Using the CTA model as a template, two response generation pathways for the *left* and *right* key-presses are created. Activation of cue_L and cue_R spreads to nodes representing their respective action-tendencies, $tendency_L$ and $tendency_R$, which provides excitation to $action_L$ and $action_R$, the motor responses of pressing the *left* or *right* key. The reduction in action-tendencies that results from the successful completion of the corresponding actions is a form of negative feedback (i.e., *consummation*, in Figure 1). In the present context, this captures the fact that pressing a key satisfies the need to produce a response, which, together with the mutual inhibition between the two competing actions (i.e., pressing *left* vs. *right*) ensures that only one of the two actions will be performed. Thus, in effect, CTA plays the role of an action arbitration and selection mechanism.

TABLE 2. Left and Right Categories in Each Task Block on the Simulated IAT and Their Corresponding Activated Concepts to Response Cues Mappings

	IAT Task Block				
Task Mapping	**ITCD**	**AAD**	**ICT**	**RTCD**	**RCT**
Categories					
Left	AA	Unpleasant	AA OR Unpleasant	EA	EA OR Unpleasant
Right	EA	Pleasant	EA OR Pleasant	AA	AA OR Pleasant
Associations					
POS→Cue_L	—	—	—	—	—
POS→Cue_R	—	+.5	+.5	—	+.5
NEG→Cue_L	—	+.5	+.5	—	+.5
NEG→Cue_R	—	—	—	—	—
EA→Cue_L	—	—	—	+.5	+.5
EA→Cue_R	+.5	—	+.5	—	—
AA→Cue_L	+.5	—	+.5	—	—
AA→Cue_R	—	—	—	+.5	+.5

Note. ITCD: Initial Target Concept Discrimination; AAD: Associated Attribute Discrimination; ICT: Initial Combined Task; RTCD: Reversed Target Concept Discrimination; RCT: Reversed Combined Task; EA: European-American; AA: African-American; POS: Positivity; NEG: Negativity; Cue_L: Left response cue; Cue_R: Right response cue; Dashes: No connections. Weights shown are initial values, prior to random perturbations.

PERFORMING THE TASK

The interactions between the above mechanisms and representations result in a competition between all propagation pathways originating from stimulus input nodes and terminating in the response nodes. The first output node to reach a threshold[3] is taken as the winning motor response. For example, suppose the task requires a *left* response for *European-American* stimuli or *pleasant* words and a *right* response for *African-American* stimuli or *unpleasant* words. When a *European-American* stimulus is presented to the model, the concept EA is activated, and activation is transmitted to cue_L in the *Response Generator*. However, if the simulated subject was initially configured with a stronger association between EA and NEG, activation would also be transmitted to cue_R, leading to competition with cue_L. This competition between the *left* and *right* response cues reduces the degree to which activation accumulates in the *left* response node, $action_L$, and consequently more time is required for it to reach its threshold. This slowdown is expected to have a significant impact on the IAT effect being measured.

3. The response threshold is defined as the activation level that a response node must reach in order for an action to be executed. For simplicity, it is set to the maximum activation level of 1.0 for the simulations described.

SIMULATION 1: THE RACE-IAT

To the extent that IAT effects can be explained in terms of different patterns of associations between concepts in memory, if we configure the network to represent a particular pattern of associations, the model ought to be able to reproduce the corresponding behaviors that are observed in experiments involving human subjects. As a proof of concept for the computational model, the purpose of the first simulation was to demonstrate that by using the propagation algorithm running over the network described above, IAT effects can indeed arise merely from differences in automatic associations.

PROCEDURE

Twenty-five simulated subjects, each with its own unique associative network, were assigned to four groups, differing only in their association configurations. The four configurations were (a) "Like EA, Dislike AA"—excitatory EA↔POS and AA↔NEG associations, inhibitory EA↔NEG and AA↔POS associations; (b) "Like AA, Dislike EA"—excitatory AA↔POS and EA↔NEG associations, inhibitory AA↔NEG and EA↔POS associations; (c) "Like/Dislike Both"—equal EA↔POS, EA↔NEG, AA↔POS and AA↔NEG associations; and (d) "Like/Dislike Neither"—no associations, that is, connection weights with a mean of zero, between target concepts and evaluative attributes. Configurations (a) and (b) were experimental conditions while (c) and (d) were control conditions. The association configurations for all conditions are summarized in the upper panel of Table 1. Weights in each configuration were perturbed with Gaussian noise, together with all other connections in the network.

Each virtual subject was put through five standard IAT tasks (see Table 2), namely, the Initial Target Concept Discrimination task (ITCD), Associated Attribute Discrimination task (AAD), Initial Combined Task (ICT), Reversed Target Concept Discrimination task (RTCD), and the Reversed Combined Task (RCT). On each trial, each simulated subject was presented with a simulated verbal or pictorial stimulus and the number of iterations taken to produce a response was recorded. The number of iterations was then transformed into a simulated response time (in milliseconds) using a linear scaling factor of 24 ms per time step. This scaling factor was selected so as to yield mean response times in the region of 600 ms (which is comparable to those reported by Greenwald et al., 1998). In this way, mean response times across all subjects and conditions were scaled to the same order of magnitude as those reported in experiments with human subjects. The IAT effect was then computed as the difference in mean response times between the ICT and RCT blocks.

RESULTS

Mean response times for all task blocks in the different configurations and magnitudes of IAT effects are shown in Figure 2. As expected, IAT effects were observed in the experimental groups (a) and (b). The effect in (a) was a preferential

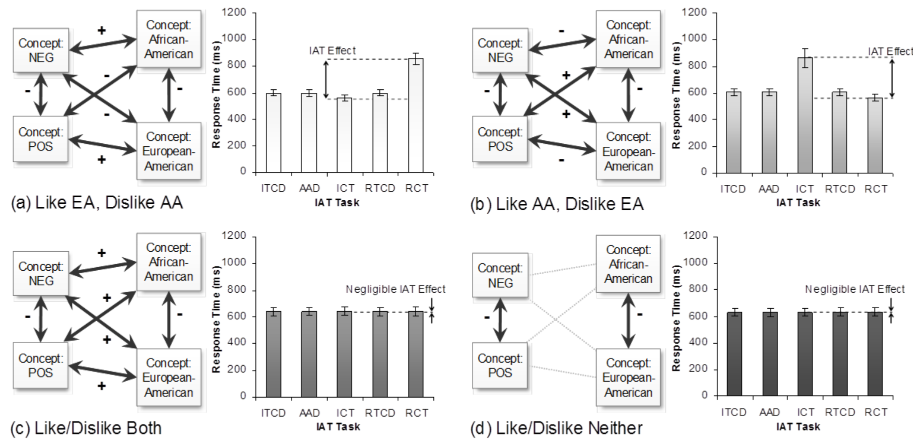

FIGURE 2. Response times across task blocks for association configurations (a) to (d) in Simulation 1. Left panels: association configurations. Right panels: mean response times for each task block, and IAT effects (i.e., difference in response times of ICT and RCT blocks). Error bars: standard deviations.

evaluation of *European-American* stimuli, $t(24) = 46.4$, $p < .001$, and in (b) of *African-American* stimuli, $t(24) = -29.1$, $p < .001$. No significant effects were found in the control groups (c), $t(24) = -1.22$, $p = 0.234$, and (d), $t(24) = 0.832$, $p = 0.413$. Summary statistics for all four conditions are shown in column I of Table 3. Just as in actual experiments with human subjects (e.g., Greenwald et al., 1998; Klauer, Voss, Schmitz, & Teige-Mocigemba, 2007), response latencies were longer for combined tasks in which target-attribute pairings were incompatible with the association configurations, while latencies for tasks in which the pairings were compatible with the association configurations were similar to those on non-combined task blocks (i.e., ITCD, AAD, and RTCD). To ensure replicability, the simulation was conducted ten times, and a consistent pattern of results was observed across the ten runs (Table 3, column II).

DISCUSSION

Simulation 1 demonstrates that the computational model is indeed capable of simulating the emergence of IAT effects. Furthermore, considering that the manipulations in this simulation were made only in terms of the differences in association configurations across experimental conditions, this simulation demonstrates the feasibility that a computational model of this form can be used to conduct virtual experiments that require manipulating the magnitude, nature (excitatory or inhibitory), and direction of such associations in memory. As we had discussed, this is generally difficult to achieve with human subjects in actual experiments—and certainly for the Race-IAT (Joy-Gaba & Nosek, 2010)—even though it is sometimes possible to establish new associations between attitude objects and strongly va-

TABLE 3. Summary of Results for Simulations 1 and 2

	I. IAT effect (ms)			II. Mean IAT effect (ms) across 10 replications		
Conditions (N = 25 each)	M	SD	**95% CI**	M_{means}	SD_{means}	**95% CI**
Simulation 1						
a. Like EA, Dislike AA	297	32.0	[283, 310]	286	9.53	[280, 292]
b. Like AA, Dislike EA	-298	51.2	[-319, -277]	-286	11.7	[-293, -278]
c. Like/Dislike Both	-3.84	15.7	[-10.3, 2.64]	-0.357	2.63	[-1.99, 1.27]
d. Like/Dislike Neither	2.43	14.6	[-3.60, 8.46]	0.501	3.19	[-1.47, 2.48]
Simulation 2						
e. Like EA more than AA	72.4	25.4	[61.9, 82.9]	64.9	5.79	[61.3, 68.5]
f. Like AA more than EA	-65.5	27.4	[-76.8, -54.2]	-65.7	3.10	[-67.6, -63.8]
g. Dislike EA more than AA	-66.5	27.7	[-78.1, -55.1]	-64.5	5.83	[-68.1, -60.9]
h. Dislike AA more than EA	72.4	34.1	[58.3, 86.5]	63.8	4.04	[61.3, 66.3]

Note. (I) Summary statistics of the IAT effect obtained in each condition; (II) Means and distribution of IAT effects across ten replications of both simulations. EA: European-American; AA: African-American; CI: confidence interval.

lenced stimuli through classical conditioning (Olson & Fazio, 2001; Olson & Fazio, 2002).

SIMULATION 2: POSITIVITY AND NEGATIVITY OF IMPLICIT ATTITUDES

In the second simulated experiment, we examine the possibility that IAT effects could emerge as a result of various other association configurations in addition to those belonging to experimental conditions (a) and (b) in the first simulation. In particular, we sought to determine whether IAT effects could be obtained in virtual subjects configured with either only positive or only negative attitudes toward both target concepts.

PROCEDURE

This simulation was conducted in the same manner as the first, but instead of the configurations in (a) to (d), four different configuration groups were tested (see the lower panel of Table 1). These are defined and labeled as: (e) "Like EA more than AA"—stronger EA↔POS associations than AA↔POS, no associations with NEG; (f) "Like AA more than EA"—stronger AA↔POS associations than EA↔POS, no

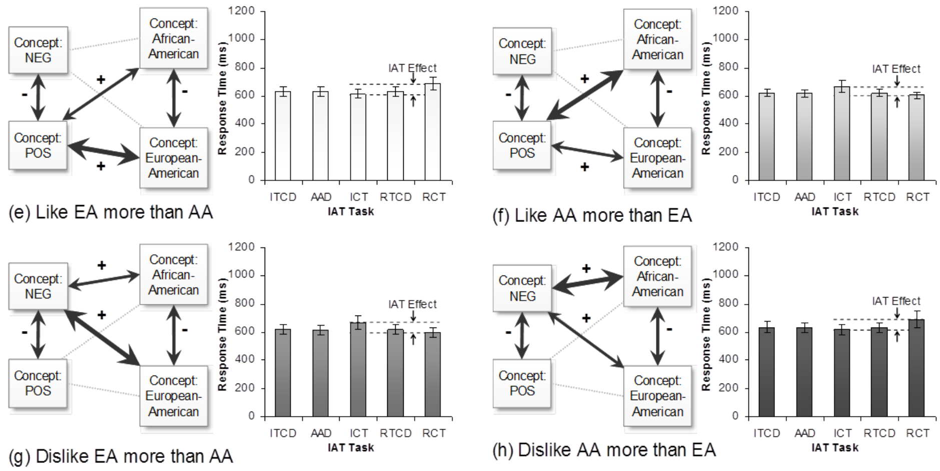

FIGURE 3. Response times across task blocks for association configurations (e) to (h) in Simulation 2. Left panels: association configurations. Right panels: mean response times for each task block, and IAT effects (i.e., difference in response times of ICT and RCT blocks). Arrow thickness: relative associative strengths. Error bars: standard deviations.

associations with NEG; (g) "Dislike EA more than AA"—stronger EA↔NEG associations than AA↔NEG, no associations with POS; and (h) "Dislike AA more than EA"—stronger AA↔NEG associations than EA↔NEG, no associations with POS. In these four configurations, stronger associations were initialized with weights of 0.8, whereas weaker associations were initialized with weights of 0.2. These weights were perturbed with Gaussian noise, together with the other connections in the network, as described earlier. As in Simulation 1, 25 simulated subjects were assigned to each configuration group, and put through all five standard IAT tasks (Table 2). Response latencies and IAT effects were determined in the same manner as in the previous simulation.

RESULTS

Response times for all task blocks and magnitudes of the IAT effect in the four configurations are shown in Figure 3. All four groups showed significant IAT effects. In (e) a preferential evaluation of *European-American* stimuli was observed, $t(24) = 14.3$, $p < .001$, and likewise, for (h), $t(24) = 10.6$, $p < .001$. In (f) a preferential evaluation of *African-American* stimuli was observed, $t(24) = -12.0$, $p < .001$, and likewise for (g), $t(24) = -12.0$, $p < .001$. Summary statistics for all four conditions are shown in column I of Table 3. No significant differences were found between (e) and (h), $t(44) = -0.004$, $p = 0.997$, or between (f) and (g), $t(48) = 0.134$, $p = 0.894$. Just as in Simulation 1, the simulation was conducted ten times to ensure replicability, and a consistent pattern of results was observed across all replications (Table 3, column II).

DISCUSSION

Consistent with the suggestions of Brendl et al. (2001) and others, the fact that IAT effects were observed in the absence of any associations between target concepts and negative attributes in conditions (e) and (f) suggests that the presence of a negative implicit attitude (if such attitudes are taken to be associations between concepts in memory) is not a prerequisite for an observed IAT effect. Similarly, results of conditions (g) and (h) indicate that it is not necessary for there to be a positive implicit attitude toward either of the target concepts for there to be an IAT effect. Furthermore, despite their different configurations, condition (e), in which EA and AA were not connected to NEG, and condition (h), in which EA and AA were not connected to POS, had similar patterns of response latencies and IAT effects that could be interpreted as a preference for *European-American* stimuli. Thus, it is not possible to differentiate between these two association configurations on the basis of their performance on the simulated IAT. Similarly, both (f) and (g) displayed comparable IAT effects that could be interpreted as a preference for *African-American* stimuli, despite having different association configurations. Therefore, the fact that an IAT effect is observed in an individual (simulated or otherwise) offers at best a partial glimpse into the characteristics of the underlying associations between the concepts in question.

VARIANTS OF THE IAT

The results of Simulations 1 and 2 indicate that, because of the ambiguity with respect to the source of the IAT effect in terms of the association configurations, the test might not be the best way of assessing implicit attitudes. In our case, we have generated IAT effects in the absence of associations of AA with NEG, or EA with POS, which is consistent with the idea that an IAT effect in favor of European-American might just as easily be the result of, in our terms, stronger EA↔POS than AA↔POS associations, stronger AA↔NEG than EA↔NEG associations, the absence of EA↔NEG or AA↔POS associations, or any combinations of these (Brendl et al., 2001; Karpinski & Steinman, 2006). The standard IAT is unable to disambiguate between these different alternatives, as illustrated in Figure 4(a). This raises the question of whether variants of the IAT, such as the Brief IAT (BIAT; Sriram & Greenwald, 2009), the Single Category IAT (SC-IAT; Karpinski & Steinman, 2006), or the Sorting Paired Features Task (SPF; Bar-Anan, Nosek, & Vianello, 2009), might be better able to distinguish between the different candidate pairs of automatic associations.

The BIAT is an abridged variant of the IAT that has only two task blocks. Instead of two combined categories, subjects are instructed in the first task block to keep in mind various exemplars from two focal categories (e.g., a short list of *pleasant* words, and an instruction to pay attention to only pictures of *African-American* individuals), and to press a *focal* key for any stimulus that matches either of the focal categories, or a *non-focal* key for all other stimuli (e.g., non-*pleasant* words or pictures of *European-American* individuals). In the second block, the two tar-

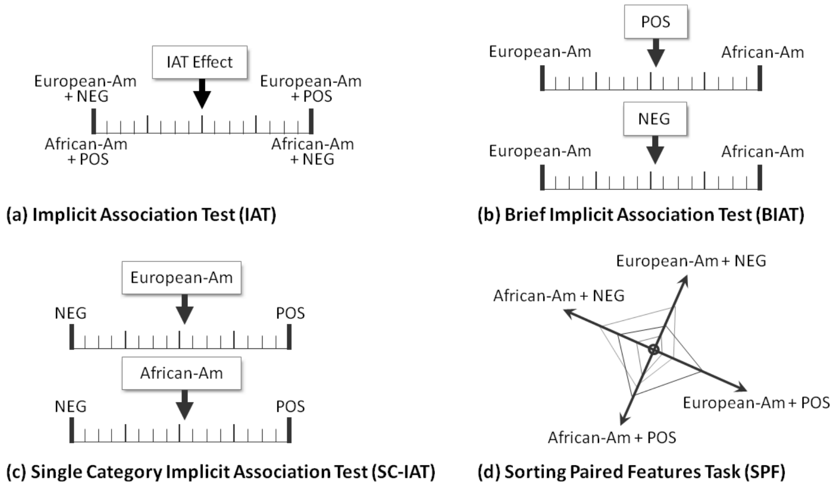

FIGURE 4. Conceptual comparison between (a) the Implicit Association Test, (b) the Brief Implicit Association Test, (c) the Single Category Implicit Association Test, and (d) the Sorting Paired Features Task.

get concepts are swapped over (for more details, see Sriram & Greenwald, 2009). Conceptually, the BIAT is equivalent to determining the relative location of POS along the EA–AA continuum, or in other words, which of EA or AA is more closely associated with POS, as illustrated in Figure 4(b). The same logic applies when *unpleasant* is the focal attribute in a second BIAT with the same target concepts.

The SC-IAT is similar to the BIAT, but it does away with the *focal* versus *non-focal* distinction and assesses only a single target concept at a time, as shown in Figure 4(c). For example, subjects classify input stimuli into the categories "*European-American* or *pleasant*" or "*unpleasant*" in one task block, and "*pleasant*" or "*European-American* or *unpleasant*" in a different block. Consequently, the comparison is not between the two target concepts per se, but between the two evaluative attributes. Conceptually, this is equivalent to determining the location of EA on the POS–NEG continuum, independently of AA, and vice versa, as shown in Figure 4(c). In other words, while BIAT is relative vis-à-vis the target concepts, in that it seeks to determine whether POS or NEG is (independently) more closely associated with EA as compared to AA, the SC-IAT is relative vis-à-vis attribute concepts, in that it seeks to determine whether EA or AA is (independently) more closely associated with POS as compared to NEG. Thus, in principle, comparing the scores of two SC-IATs (with EA and AA as their respective target concepts) or two BIATs (with POS as a focal attribute in the first and NEG in the second) should enable all four pairs of associations to be identified.

The considerably more complex SPF task (Bar-Anan et al., 2009) requires subjects to sort pairs of stimuli, each comprising a target concept exemplar and an attribute concept word (e.g., the word *wonderful* and a picture of a person are displayed on screen simultaneously), into one of four combined conjunctive categories. Each combined category, located in one of the four corners of the screen, corresponds to one of the four possible target-attribute pairings (i.e., "*European-American + pleasant*," "*European-American + unpleasant*," "*African-American + pleasant*," and "*Af-*

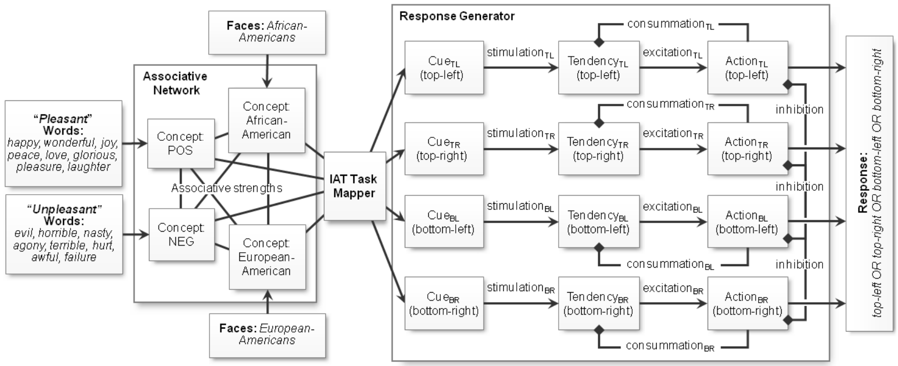

FIGURE 5. Modified network model for simulating the Sorting Paired Features Task.

rican-American + unpleasant"). The mean response latencies for sorting the paired stimuli into each combined category are expected to be inversely proportional to the strengths of the corresponding implied implicit associations (i.e., EA↔POS, EA↔NEG, AA↔POS, and AA↔NEG).

The apparent advantage of the SPF over the other IAT variants discussed above lies in its purported ability to identify or tease apart each of the four automatic associations in question, as illustrated in Figure 4(d), although the strength of each inferred associations can only be interpreted in relation to the other three (Bar-Anan et al., 2009). This possibility, and the fact that we currently know much less about the SPF than we do about other implicit measures, seemed to merit further study. For this reason, in the following section, the computational model is modified and expanded to simulate human performance on the SPF. Our aim is to evaluate the possibility that this variant of the IAT can indeed identify and measure each of the different pairs of automatic associations, independently of the others.

SIMULATION 3: THE SORTING PAIRED FEATURES TASK

In order to simulate the Sorting Paired Features Task, it is necessary to extend the network model employed in Simulations 1 and 2 with two additional (and competing) CTA response pathways (to make a total of four), and to configure the task mappings for each of the four responses (i.e., *top-left*, *top-right*, *bottom-left*, and *bottom-right*). The modified network model is shown in Figure 5, while the updated task mappings for each of the four task blocks on the SPF and their assignment to different corners of the screen are shown in Table 4.

PROCEDURE

The simulation was conducted in the same manner as Simulations 1 and 2. Twenty-five simulated subjects were assigned to each of the eight configuration groups (a) to (h) defined earlier (Table 1). Each simulated subject was put through all four

TABLE 4. Categories in Each Task Block on the Simulated SPF and their Corresponding Activated Concepts to Response Cues Mappings

	SPF Task Block			
Task Mapping	**1**	**2**	**3**	**4**
Categories				
Top Left	EA + Pleasant	EA + Unpleasant	AA + Unpleasant	AA + Pleasant
Top Right	EA + Unpleasant	AA + Unpleasant	AA + Pleasant	EA + Pleasant
Bottom Left	AA + Pleasant	EA + Pleasant	EA + Unpleasant	AA + Unpleasant
Bottom Right	AA + Unpleasant	AA + Pleasant	EA + Pleasant	EA + Unpleasant
Associations				
POS→Cue_{TL}	+.5	—	—	+.5
POS→Cue_{TR}	—	—	+.5	+.5
POS→Cue_{BL}	+.5	+.5	—	—
POS→Cue_{BR}	—	+.5	+.5	—
NEG→Cue_{TL}	—	+.5	+.5	—
NEG→Cue_{TR}	+.5	+.5	—	—
NEG→Cue_{BL}	—	—	+.5	+.5
NEG→Cue_{BR}	+.5	—	—	+.5
EA→Cue_{TL}	+.5	+.5	—	—
EA→Cue_{TR}	+.5	—	—	+.5
EA→Cue_{BL}	—	+.5	+.5	—
EA→Cue_{BR}	—	—	+.5	+.5
AA→Cue_{TL}	—	—	+.5	+.5
AA→Cue_{TR}	—	+.5	+.5	—
AA→Cue_{BL}	+.5	—	—	+.5
AA→Cue_{BR}	+.5	+.5	—	—

Note. EA: European-American; AA: African-American; POS: Positive valence; NEG: Negative valence; Cue_{TL}: Top-Left response cue; Cue_{TR}: Top-Right response cue; Cue_{BL}: Bottom-Left response cue; Cue_{BR}: Bottom-Right response cue; Dashes: No connections. Weights shown are initial values, prior to random perturbations.

task blocks on a Race-SPF test. These four task blocks exist only for the purpose of cycling the four combined categories through all four corners of the screen to control for screen placement effects. To emulate the presentation of each stimulus pair (comprising a *pleasant* or *unpleasant* word, together with a picture of a European-American or African-American individual), the input nodes corresponding to each of the conjuncts of the stimulus pair were activated simultaneously, instead of just a single input node in the case of the simulated IAT. The first response node to reach its activation threshold was identified as the winning response. Mean response latencies for each of the four possible target-attribute concept pairings were determined in the same manner as in the previous simulations, but with a larger scaling factor (80 ms per time step) so as to transform the number of iterations taken (by each network to produce a response) into response times that are of the same order of magnitude as those reported in Bar-Anan et al. (2009). Once this was done, no further computations were necessary, as there is no analog of the IAT effect for the SPF that needs to be evaluated; the mean response latencies for

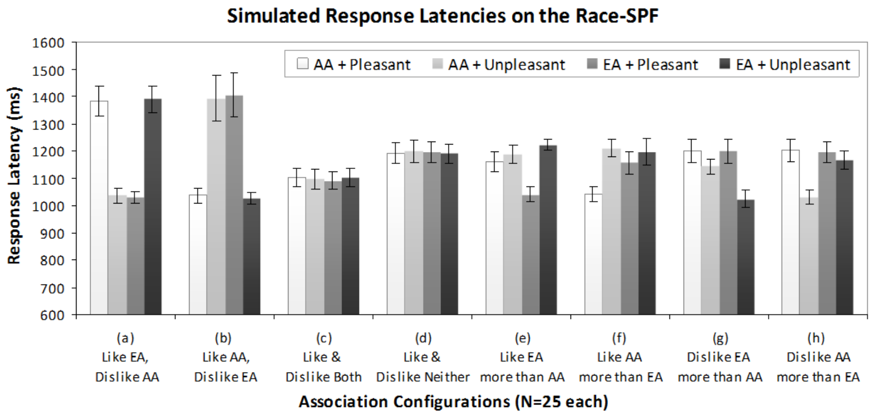

FIGURE 6. Simulated response latencies across the four target-attribute combined categories on the Race-SPF for association configurations (a) to (h) defined in Table 1. Error bars: standard deviations.

the four combined categories are themselves deemed to reflect the strengths of the corresponding automatic associations between each target-attribute concept pair.

RESULTS

Mean response times for the four combined categories across all task blocks, broken down according to each of the association configurations are shown in Figure 6 and summarized in Table 5. In configuration (a), consistent with the results of Simulation 1, response latencies were faster for the combined categories of "*European-American + pleasant*" and "*African-American + unpleasant*," and slower for those of "*African-American + pleasant*" and "*European-American + unpleasant*." The converse was true for configuration (b). Both control groups (c) and (d) showed similar response latencies across all task blocks, but in the case of (c), the mean latency was faster than that of (d), which is arguably due to a facilitating effect from the equal and positive associations (random perturbations notwithstanding) between all target-attribute concept pairs in (c).

In configuration (e), the fastest responses were for the pair "*European-American + pleasant*," which was expected from the pre-configured associative strength of 0.8 for EA↔POS. The next quickest responses were for the pair "*African-American + pleasant*," which was expected from the pre-configured associative strength of 0.2 for AA↔POS. The remaining two combined categories "*European-American + unpleasant*" and "*African-American + unpleasant*" had response latencies that were similar to those obtained in control configuration (d), which was expected from the fact that their corresponding associations (i.e., EA↔NEG and AA↔NEG) were pre-configured with weights of zero. A pattern similar to (e) was observed for (f), (g), and (h), in each of which the fastest responses occurred for the target-attribute concept pair that had the strongest pre-configured associative strength of 0.8 (as opposed to 0.2).

TABLE 5. Simulated SPF Response Latencies Across Target-Attribute Combined Categories for Each Association Configuration in Simulation 3

	Response latency (ms) by combined category							
	AA + Pleasant		AA + Unpleasant		EA + Pleasant		EA + Unpleasant	
Conditions (*N* = 25 each)	*M*	*SD*	*SD*	*SD*	*SD*	*SD*	*M*	*SD*
Simulation 1								
a. Like EA, Dislike AA	1383	54.4	1036	27.5	1031	21.8	1390	49.8
b. Like AA, Dislike EA	1037	26.4	1393	83.7	1404	81.3	1025	22.9
c. Like/Dislike Both	1102	35.3	1096	34.5	1091	31.7	1101	34.3
d. Like/Dislike Neither	1192	38.1	1198	41.4	1196	38.8	1190	37.1
Simulation 2								
e. Like EA more than AA	1160	35.1	1187	34.9	1040	27.6	1222	18.8
f. Like AA more than EA	1041	28.6	1210	31.9	1157	40.6	1198	50.8
g. Dislike EA more than AA	1201	42.4	1143	28.5	1200	44.7	1023	30.3
h. Dislike AA more than EA	1203	39.7	1031	24.6	1197	38.1	1166	32.4

Note. EA: European-American; AA: African-American

DISCUSSION

Simulation 3 demonstrates the capacity of the SPF to tease apart the four associations corresponding to the target-attribute concept pairings, especially in the case of association configurations (e), (f), (g), and (h). In Simulation 2, despite having different configurations, both (e) and (h) had similar IAT scores (Table 3) indicating a preference for EA stimuli. Thus, while the IAT was unable to differentiate between them, the SPF could. The same is true of conditions (f) and (g), both having similar IAT effects in favor of the same target concept despite being initialized with different configurations, but which can be disambiguated by the SPF.

We can now ask whether the SPF can provide an assessment of the strength of each association. It would seem from the above simulation that sorting each target-attribute pair by ascending order of response times would produce a ranking that matches the descending order of their initially configured associative strengths. For instance, sorting response latencies in condition (g) produces the order EA-NEG, AA-NEG, EA-POS, AA-POS, which matches their initially configured associative strengths of 0.8, 0.2, 0, 0, in that order. However, one could not infer from this order that EA↔POS and AA↔POS were initially configured with associative strengths of zero. In order to do this, the simulation results for control condition (d), in which all target-attribute associations were initialized to zero, have to be used as a baseline. When we do this, the associations of target-attribute combined categories whose response latencies do not differ significantly from their counterparts in (d) can be taken as zero. In the case of configuration (g), applying this method suggests that EA↔POS and AA↔POS can be inferred as having insignificant associative strengths. However, such a comparison is possible only when a set of response latencies such as (d) is available as a baseline. While relatively easy to achieve in simulation, it would not be feasible to obtain such a baseline from an actual SPF involving human subjects, as it would entail setting pre-existing target-attribute associations to zero for a group of control subjects. One could arguably conduct an SPF with two novel attitude objects and then utilize the resulting response latency scores as a surrogate baseline against which Race-SPF response latencies could be compared. Unfortunately this makes for a poor comparison because the target concepts across the two SPF tests are obviously different. Nevertheless, this might suffice as an interim solution. For more general purposes that only require the identification of the relative strengths of each of the four target-attribute associations (as opposed to identifying them in absolute terms), our simulation results suggest that the SPF does indeed have greater resolving power than the IAT.

GENERAL DISCUSSION

The results of our simulations suggest that the standard IAT cannot differentiate between IAT effects induced by a positive attitude toward one target concept and a negative attitude toward the other, or between positive attitudes for both target concepts (but with one more positive than the other), or between negative attitudes for both (with one more negative than the other). Thus it seems that an unambiguous interpretation of IAT effects is not possible. IAT variants such as the Single Category IAT and the Sorting Paired Features Task go some way toward addressing this problem, even though these measures assess only relative associative strengths. For instance, Simulation 3 suggests that the SPF is capable of distinguishing IAT effects caused by, for example, strong EA↔POS associations, from those induced by strong AA↔NEG associations. Similarly, one might expect the SC-IAT to be able to determine whether a target concept is more closely associated with POS or with NEG, independently of other target concepts.

It might be objected that a weakness of our model is the seemingly arbitrary values that were initially assigned to the connection weights. One response to this criticism would be to note that despite the Gaussian perturbations randomly applied to these values, simulated subjects in all of the experimental groups in Simulations 1 and 2 showed IAT effects, suggesting that the initial weights are probably not critical. This might appear counterintuitive, but in fact it is consistent with data from the real world indicating that thousands of people from all walks of life (which can be thought of as an analog of randomly distributed weights) routinely display significant IAT effects (Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Nosek et al., 2002). Obviously, it is not possible to assign initial values to connection weights on the basis of empirical data about actual implicit association strengths, because such data are not available. If they were, there would be no need for implicit measures in the first place!

Another criticism is that in modeling virtual subjects, we have not emphasized a distinction between automatic and controlled processes (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977), despite the recognized importance of this distinction (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005; Mierke & Klauer, 2001). Nor have we attempted to deal with issues related to the role of selective attention (Fazio, 2001; Roskos-Ewoldsen & Fazio, 1992; Rothermund & Wentura, 2004), or cognitive load (Snyder-Tapp & Dale, 2009). Effects resulting from other causes (apart from association configurations), such as asymmetries in target and attribute salience (Rothermund & Wentura, 2001; Rothermund & Wentura, 2004) or familiarity (Ottaway et al., 2001), the influence of non-focal features other than valence (De Houwer, 2001), and task-set switching overheads (Klauer & Mierke, 2005), remain to be explored. Arguably, task instructions might themselves be regarded as input stimuli, since subjects need to attend to issues such as the location of category labels in the upper corners of the display, which effectively make the IAT a multi-input task. Addressing these issues in our model would involve augmenting the network with nodes representing IAT tasks that would dynamically alter connection weights from attitude objects to the response cues, instead of assigning these weights before each task.

At this moment, our simulation of the Sorting Paired Features Task amounts to little more than proof of concept. Because of its complexity, there are several reasons why the SPF would warrant a separate and more detailed treatment from the simpler IAT variants. First, all four possible target-attribute pairings are in focus, which renders the use of spontaneous strategies that rely on the heuristic of mentally assigning certain concepts to focal or non-focal groups (Rothermund & Wentura, 2001; Sriram & Greenwald, 2009) unfeasible. Second, in the IAT, because the combined categories are disjunctive, classification can be simply achieved on the basis of only a partial match. However, because in the SPF the combined categories are conjunctive rather than disjunctive, a complete match with both categories is mandatory, and thus modeling performance will be more involved due to the need to accommodate different processing strategies (e.g., hierarchical vs. sequential searches).

The computational model we have described is generic in that it can be reconfigured for any type or number of input stimuli by the simple insertion of new input nodes or the relabeling of existing ones. The same can be done to configure the model for different target concepts and attributes to simulate versions of the IAT for other domains (e.g., age, sexual preferences, self-esteem, identity, and well-being). In addition, just as in the case of the SPF, extending the model to simulate the Single Category IAT or the Brief IAT would require slight modifications to accommodate the new task mappings (in the case of SC-IAT) or the mapping of non-focal stimuli to the alternative response (in the case of BIAT). Extending the model along these lines might offer additional insight into the inner workings of implicit social cognition and how they might influence behavior and decision making.

In conclusion, whereas one cannot realistically expect to build simulations that take into account all possible strategic alternatives in the processing of implicit attitudes (or any other complex cognitive tasks for that matter), we think that the sort of simulations we have described have the potential to provide useful insights into the kinds of processes that are possible or even likely to be implicated. We believe that such simulations are a useful way of concretizing theoretical models—in the present case, the role of automatic associations in implicit attitudes. We also think that they can provide data of a kind that might sometimes be difficult to collect in laboratory studies. In the present case, this was demonstrated by our simulations of IAT effects arising from particular configurations of implicit associations, the existence of which would be difficult to find or confirm in human subjects.

REFERENCES

Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. New York: Wiley.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The Sorting Paired Features (SPF) Task: A measure of association strengths. *Experimental Psychology*, *56*, 329-343. doi:10.1027/1618-3169.56.5.329

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test (IAT). *Journal of Personality and Social Psychology, 81*, 760-773. doi:10.1037/0022-3514.81.5.760

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*, 469-487. doi:10.1037/0022-3514.89.4.469

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for White Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*, 316-328. doi:10.1006/jesp.1999.1418

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test (IAT). *Journal of Experimental Social Psychology, 37,* 443-451. doi:10.1006/jesp.2000.1464

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*, 347-368. doi:10.1037/a0014211

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, *15,* 115-141. doi:10.1080/02699930125908

Frantz, C. M., Cuddy, A. J. C., Burnett, M., Ray, H., & Hart, A. (2004). A threat in the computer: The Race Implicit Association Test (Race-IAT) as a stereotype threat experience. *Personality and Psychology Bulletin*, *30*, 1611-1624. doi:10.1177/0146167204266650

Gawronski, B., Hofmann, W., & Wilbur, C. J. (2006). Are "implicit" attitudes unconscious? *Consciousness and Cognition*, *15*, 485-499. doi:10.1016/j.concog.2005.11.007

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem and stereotypes. *Psychological Review, 102,* 4-27. doi:10.1037/0033-295X.102.1.4

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109,* 3-25. doi:10.1037/0033-295X.109.1.3

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test (IAT). *Journal of Personality and Social Psychology, 74,* 1464-1480. doi:10.1037/0022-3514.74.6.1464

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test (IAT): III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17-41. doi:10.1037/a0015575

Joy-Gaba, J. A., & Nosek, B. A. (2010). The surprisingly limited malleability of implicit racial evaluations. *Social Psychology*, *41*, 137-146. doi:10.1027/1864-9335/a000020

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test (IAT). *Journal of Personality and Social Psychology, 81,* 774-788. doi:10.1037/0022-3514.81.5.774

Karpinski, A., & Steinman, R. B. (2006). The single category Implicit Association Test (IAT) as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*, 16-32. doi:10.1037/0022-3514.91.1.16

Klauer, K. C., & Mierke, J. (2005). Task-set inertia, attitude accessibility and compatibility-order effects: New evidence for a task-set switching account of the Implicit Association Test (IAT) effect. *Personality and Social Psychology Bulletin, 31*, 208-217. doi:10.1177/0146167204271416

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test (IAT): A diffusion-model analysis. *Journal of*

Personality and Social Psychology, 93, 353-368. doi:10.1037/0022-3514.93.3.353

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test (IAT). *Social Cognition*, *20*, 483-510. doi:10.1521/soco.20.6.483.22977

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie, 48*, 107-122. doi:10.1026/0949-3946.48.2.107

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test (IAT). *Journal of Personality and Social Psychology*, *85*, 1180-1192. doi:10.1037/0022-3514.85.6.1180

Neumann, R., & Seibt, B. (2001). The structure of prejudice: Associative strength as a determinant of stereotype endorsement. *European Journal of Social Psychology*, *31*, 609-620. doi:10.1002/ejsp.69

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*, 101-115. doi:10.1037/1089-2699.6.1.101

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science*, *12*, 413-417. doi:10.1111/1467-9280.00376

Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition*, *20*, 89-103. doi:10.1521/soco.20.2.89.20992

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the Implicit Association Test (IAT). *Social Cognition, 19,* 97-144. doi:10.1521/soco.19.2.97.20706

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23*, 443-512. doi:10.1017/S0140525X00003356

Quek, B. K., & Ortony, A. (2011). Modeling underlying mechanisms of the Implicit Association Test (IAT). In L. Carlson, C. Hoelscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1330-1335). Austin, TX: Cognitive Science Society.

Revelle, W. (1986). Motivation and efficiency of cognitive performance. In D. R. Brown & J. Veroff (Eds.), *Frontiers of motivational psychology: Essays in honor of J. W. Atkinson* (pp. 105-131). Berlin: Springer.

Roskos-Ewoldsen, D. R., & Fazio, R. H. (1992). On the orienting value of attitudes: Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality and Social Psychology, 63,* 198-211. doi:10.1037/0022-3514.63.2.198

Rothermund, K., & Wentura, D. (2001). Figure-ground-asymmetries in the Implicit Association Test (IAT). *Zeitschrift für Experimentelle Psychologie, 48*, 94-106. doi:10.1026/0949-3946.48.2.94

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology, 133*, 139-165. doi:10.1037/0096-3445.133.2.139

Schneider, W., & R. M. Shiffrin. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1-66. doi:10.1037/0033-295X.84.1.1

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190. doi:10.1037/0033-295X.84.2.127

Snyder-Tapp, K., & Dale, R. (2009). Dynamic competition and the cognitive bottleneck. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (pp. 3052-3057). Austin, TX: Cognitive Science Society.

Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test (BIAT). *Experimental Psychology*, *56*, 283-294. doi:10.1027/1618-3169.56.4.283