Modeling the Influence of Cognitive Fluency and Stereotype Threat on the Processing of Implicit Attitudes

Boon-Kiat Quek (boonkiat.quek@northwestern.edu)
Andrew Ortony (ortony@northwestern.edu)
Department of Psychology, Northwestern University, Evanston, IL 60208, USA, and
Institute of High Performance Computing, Agency for Science, Technology and Research, Singapore

Abstract

Studies reveal that the processing of implicit attitudes could be affected by individual differences in cognitive fluency, as well as by the presence of stereotype threat induced when subjects were primed with negative prejudices about their own social group. Using a previously proposed computational model of human performance on the Implicit Association Test, we examine possible processing mechanisms in which cognitive fluency and stereotype threat could influence the processing of implicit attitudes. Our goal is to extend the model to provide a cohesive and computationally plausible account for these effects; this is achieved by manipulating several model parameters that are analogous to human cognitive ability (in terms of processing speed and information retention ability) and shifts in confidence criteria for decision-making.

Keywords: Implicit attitudes; cognitive ability; simulation; localist-connectionist networks.

Introduction

Implicit attitudes are generally assumed to underlie people's thoughs, actions, choices and behavior (Greenwald & Banaji, 1995). Understanding how such attitudes are processed might therefore provide some insight about why people behave in the way they do. Some ways in which such processes could be investigated include affective priming (Fazio, Sanbonmatsu, Powell & Kardes, 1986) and the Implicit Association Test (IAT; Greenwald, McGhee & Schwartz, 1998). The IAT was designed to assess automatic associations between concepts in memory. It relies on a simple two-choice response time paradigm which measures the time taken by subjects to classify sequentially presented input stimuli (words or images) into one of two composite categories, each comprising a target concept (e.g., *flower*, *insect*) paired with an attribute concept (e.g., *pleasant*, *unpleasant*). Response latencies are expectedly shorter when targets are paired with *compatible* attributes (e.g., "*flower* or *pleasant*", "*insect* or *unpleasant*"), and longer when paired with *incompatible* attributes (e.g., "*flower* or *unpleasant*", "*insect* or *pleasant*"). The difference in mean response times between *compatible* and *incompatible* categories is known as the *IAT effect*, and is taken as the relative preference for one target over another.

Despite its wide application, many issues concerning the construct validity of the IAT have been raised (e.g., De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009; Mierke & Klauer, 2003). Apart from automatic associations, performance on the IAT seems to also depend on various other factors, such as stimulus familiarity (Ottaway, Hayden & Oakes, 2001), concept saliency (Rothermund & Wentura, 2004), and extra-personal knowledge about prevailing cultural or societal norms (Karpinski & Hilton, 2001). Furthermore, several anomalous effects have also been observed. In a recent review, De Houwer, Teige-Mocigemba, Spruyt and Moors (2009) suggested that the processing of implicit attitudes could be influenced by differences in cognitive ability, citing McFarland and Crouch (2002) who observed significant correlations between response latencies and magnitudes of IAT effects, and Hummert, Garstka, O'Brien, Greenwald and Mellott (2002) who observed that IAT effects tended to increase with age. Given that processing speed is an important aspect of cognitive ability (Hunt, 1983) and declines with age (Salthouse, 1996), we will expect subjects with lower cognitive abilities (especially with age-induced decline) to exhibit longer response latencies across all tasks on the IAT. Why this is associated with larger IAT effects, however, remains to be determined.

Another intriguing aspect of performance on the IAT is the possible role of stereotype threat. In a number of Race-IATs, Frantz, Cuddy, Burnett, Ray & Hart (2004) consistently observed that White subjects exhibited stronger pro-White IAT effects on the Race-IAT when they were instructed beforehand that the test might expose their racial prejudices, as compared to other White subjects in control groups who were not similarly informed. Frantz et al. suggested that being told beforehand of the actual purpose of the Race-IAT would present a stereotype threat experience (Steele & Aronson, 1995) to the informed subjects, where knowledge of the test's purpose might induce anxiety over the risk of confirming negative stereotypes about the racial attitudes that people in their social group are often presumed to endorse (e.g., being pro-White or anti-Black). Thus, we would expect subjects informed of the test's purpose to have a greater interest in positive self-presentation and hence stronger motivation to respond in a more egalitarian manner (Frantz et al., 2004). Ironically, attempts to avoid the negative stereotype appeared to interfere with performance on the Race-IAT, producing a stronger pro-White IAT effect instead of reducing it. However, no suggestions were provided to explain how such task interference might have taken place, nor the manner in which strategies for coping with the stereotype threat experience might have backfired.

Both the *cognitive fluency effect* and the *stereotype threat effect* are noteworthy because they have important implications for our understanding of the nature of information processing that underlie performance on the IAT. In this paper, we examine some of these implications, and propose

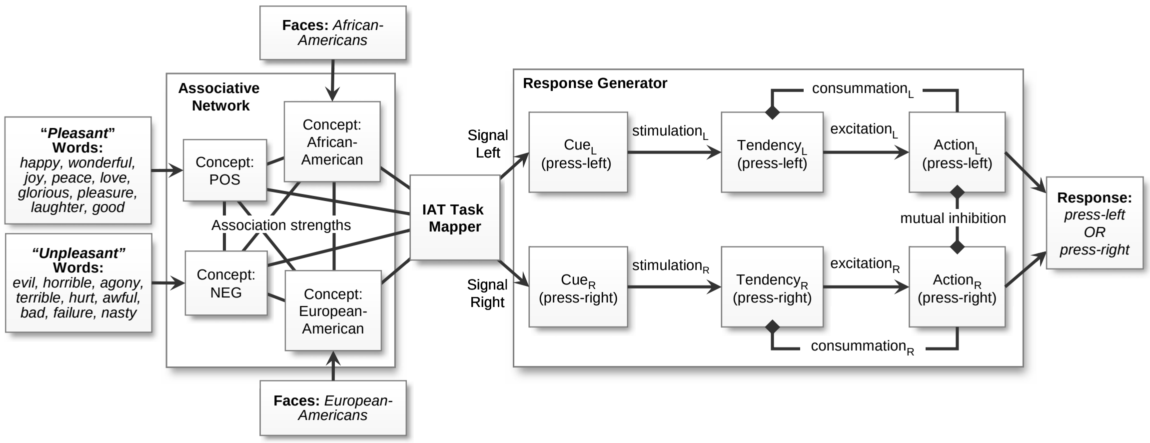

Figure 1. Network model for simulating IAT performance (Quek & Ortony, 2011)

a cohesive information processing account of these two effects, by means of a previously proposed computational model of implicit task performance on the IAT (Quek & Ortony, 2011). Our approach is to replicate the observed effects in simulations by manipulating model parameters that provide analogs for human cognitive ability and confidence criteria for decision-making. The first simulation allows us to explore how differences in cognitive fluency affect response latencies, as well as allowing us to explain the causes of the larger IAT effects. In the second simulations we examine plausible mechanisms behind how certain actions taken by subjects to cope with the stereotype threat would ironically exacerbate IAT effects instead of reducing them. This, as we discuss, has implications for the effortful control or influence over performance on the IAT.

Model Overview

In this section, we provide a brief overview of the proposed computational model; more details can be found in Quek & Ortony (2011). The model employs a spreading activation algorithm over a localist-connectionist network (e.g., Page, 2000) to emulate multiple processing pathways from the visual perception of a stimulus (i.e., a word or image) to the automatic activation of associated concepts in memory and motor responses. Nodes in the network represent concepts while edges or connections represent associations between them. Propagation of activation through the network is governed by the following rule:

$$x_i(k+1) = (1-\delta)x_i(k) + \alpha \sum_{\varepsilon_{j,i} \in E} x_j(k) \cdot w_{j,i}(k)\ , \qquad (1)$$

where x_i is the activation level of a node v_i , $w_{j,i}$ is the weight of the connection $\varepsilon_{j,i}$ from a node v_j to v_i , E is the set of all edges, α is the propagation gain and δ is a decay parameter that reduces activation over time. In each time step k, activation spreads to v_i from each neighbor v_j at a rate proportional to the weight $w_{j,i}$ of the connection $\varepsilon_{j,i}$ between them.

Virtual subjects are each represented by a network of the topology described in Figure 1. The *Associative Network* contains nodes representing the target concepts AFRICAN-AMERICAN (AA) and EUROPEAN-AMERICAN (EA), generalized concepts for positivity (POS) and negativity (NEG), input stimuli such as words belonging to the semantic fields *pleasant* and *unpleasant* (e.g., *happy, wonderful, joy, evil, horrible, hurt*), and pictures of *European-American* and *African-American* individuals. Connections between these concepts, for instance, EA↔POS, EA↔NEG, AA↔POS, and AA↔NEG represent implicit associations between them. As an example, positive attitudes towards AA can be represented as excitatory AA↔POS or inhibitory AA↔NEG associations, or both, such that activation of AA will excite POS but inhibit NEG.

The *Task Mapper* dynamically transmits activation accumulated from target concepts and evaluative attributes to nodes cue_L and cue_R indicating that a left or right key-press is required. If the current task requires a right response for "*European-American* or *pleasant*", the *Task Mapper* routes both POS and EA to cue_R. These connections remain active throughout the task block but are reconfigured prior to each subsequent task block (see Quek & Ortony, 2011, Figure 2).

The *Response Generator* is a network-based instantiation of the cue-tendency-action model (CTA; Revelle, 1986) of the dynamic interactions between conflicting tendencies and competing actions. Using CTA as a template, two response-generating pathways (for the left and right key-presses) are instantiated. Activated cues stimulate tendency nodes that in turn excite the left and right motor response nodes. When either $action_L$ or $action_R$ exceeds a response threshold x_{thres} (set to 1.0 by default), it is taken as the winning action.

The interactions between the above representations occur in the form of excitations and inhibitions between all input stimuli to motor response propagation pathways. For example, in a task block requiring a left key-press for "*European-American* or *pleasant*" stimuli and a right response for "*African-American* or *unpleasant*" stimuli, a picture of a European-American individual would activate EA, and activation will be transmitted to cue_L. However, if the network is configured with a strong EA↔NEG connection, activation will also be transmitted to cue_R, competing with cue_L. This reduces the rate that activation accumulates in the left response node, and thus a longer time is required for it to reach the response threshold.

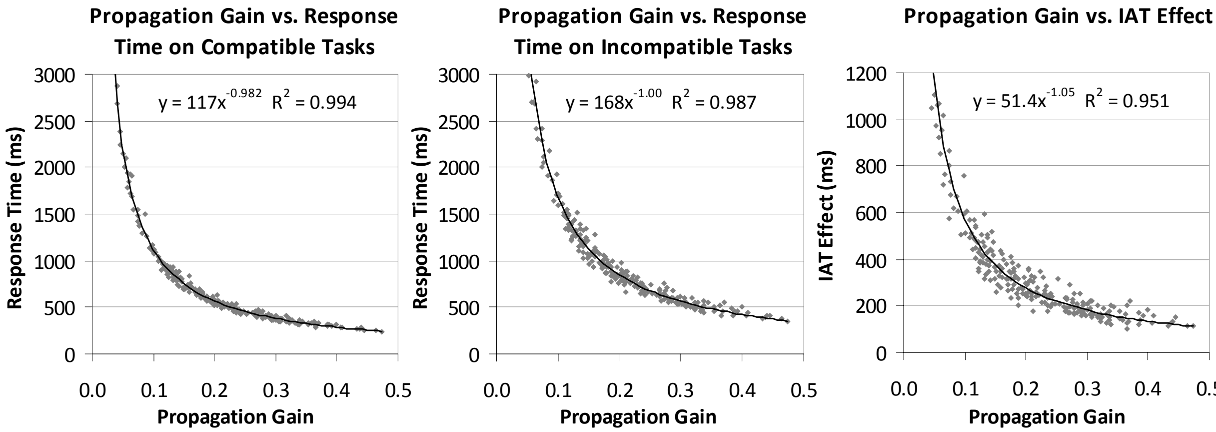

Figure 2. Distributions of response latencies on compatible and incompatible task blocks and corresponding IAT effects with variations in propagation gain α for 250 virtual subjects configured with a relative preference for one target concept over the other.

Simulating the IAT

When simulating the IAT, each virtual subject's network is initialized with a set of associative strength configurations, and put through all task blocks. On each trial, the virtual subject is presented with a verbal or pictorial stimulus and the input node corresponding to the stimulus is set with an activation of 1.0. The number of iterations taken to produce a response (i.e., when the activation level of either the left or right response node reaches its threshold) is recorded. This quantity is transformed by a scaling factor into mean response times (in milliseconds) of approximately the same magnitudes as those observed in human subjects (e.g., Greenwald et al., 1998; Klauer, Voss, Schmitz & Teige-Mocigemba, 2007). The IAT effect is then taken to be the difference between mean response times in the two combined task blocks.

Modeling Cognitive Fluency Effects

Cognitive ability and intelligence are often considered to be closely related to information processing speed (e.g., Lansman, Donalson, Hunt & Yantis, 1982; Hunt, 1983). In addition, the ability to retain information during the execution of cognitive operations is another important aspect (Salthouse, 1996). An inability to retain products from earlier processing operations due to information decay or displacement would impair problem solving performance especially when processing speed is slow. Thus, information needed for later processing stages might be partially lost by the time it is needed, in which case additional time would be required to reprocess it.

To the extent that our computational model could emulate information processing on the IAT, it should be capable of replicating both the increase in response latency and the corresponding increase in IAT effect that arises with a reduction in cognitive ability. This can be achieved by manipulating both the parameters for propagation gain α and propagation decay δ in Equation (1), which governs the rate at which activation is propagated through the network, and the rate at which activation is reduced or lost in the absence of excitatory inputs, respectively. Both of these parameters can be considered as the model's analog for the aforementioned aspects of general cognitive ability that relate to information processing and retention.

To examine the effect of variations in the propagation gain α on response latencies and IAT scores, we generated instances of the network model for a population (N = 250) of virtual subjects using the associative strength configuration in which EA↔POS and AA↔NEG were set to 0.5 (i.e., excitatory), while AA↔POS and EA↔NEG were set to -0.5 (i.e., inhibitory), representing individuals with positive attitudes towards EA and negative attitudes towards AA. As shown in Quek & Ortony (2011), this configuration produces an IAT effect in favor of EA. As in earlier simulations, weights in the network were randomly perturbed with Gaussian noise of $\sim N(0, 0.1^2)$, to ensure inter- and intra-subject variability. Each virtual subject's network was then configured with a random value of α (within a reasonable range), and put through all five standard IAT tasks.

The distributions of response latencies on both compatible and incompatible task blocks and the corresponding IAT effects in Figure 2 reveal a distinct inverse relationship between the propagation gain α and the response latencies, as well as the magnitude of the simulated IAT effects. While smaller values of α resulted in longer response latencies on both the compatible and incompatible task blocks, and stronger IAT effects, the converse is true for larger values of α. From the different coefficients of curve-fitting indicated in the figure, we can infer that the increase in magnitude of the IAT effect with a reduction in α is due to a divergence between the response latencies of the two task blocks.

We repeated the above simulation by varying the propagation decay parameter δ while keeping the propagation gain α at its default value. Plots of response latencies and IAT effects with respect to the decay parameter δ are shown in Figure 3. The scatter plots reveal a direct relationship between δ and the response latencies, and magnitudes of the IAT effects. Faster rates of information decay, as implied by higher values of δ not only resulted in longer response times

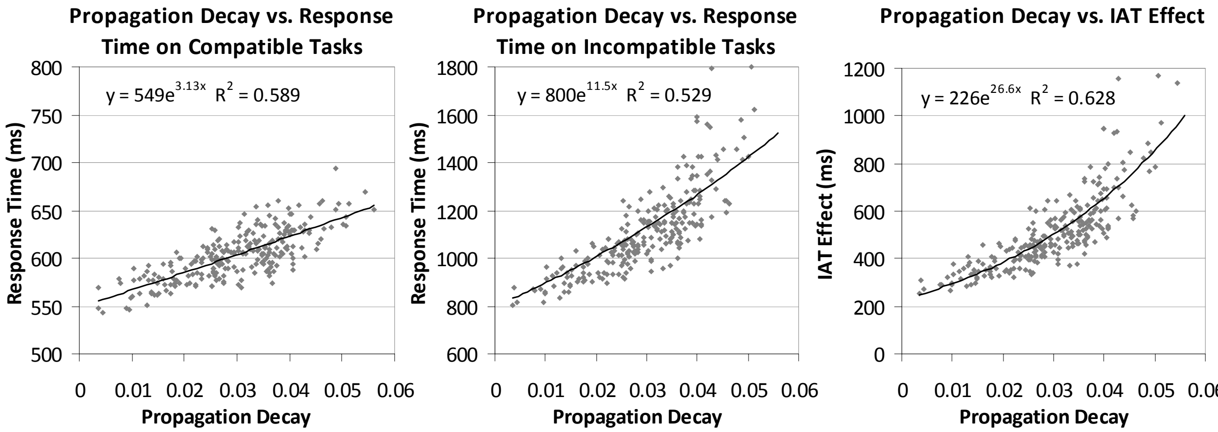

Figure 3. Distributions of response latencies on compatible and incompatible task blocks and corresponding IAT effects with variations in the propagation decay δ for 250 virtual subjects configured with a relative preference for one target concept over the other.

on both the compatible and incompatible task blocks, but also produced stronger IAT effects. The converse is true for smaller values of δ which, if taken to imply better information retention ability, results in better overall performance on the IAT. As in the case of propagation gain, the larger IAT effects with reduction in information retention is due to a divergence between the response latencies of the two task blocks, as inferred from their curve-fitted coefficients.

The above simulations demonstrate the efficacy of the network model in replicating cognitive fluency effects on the IAT, namely that a lower propagation gain or higher decay rate (both indicative of lower cognitive ability) in the model would lead to higher response latencies on both compatible and incompatible tasks, and stronger IAT effects. Since both response latencies and IAT effect magnitudes vary inversely with propagation gain, the first simulation shows consistency with the observed correlation between response latencies and IAT effect magnitudes reported in McFarland and Crouch (2002). Higher values of the decay parameter δ, corresponding to a higher rate at which activation in nodes are leaked or lost over time, appear to impair virtual subjects' efficiency and performance on the tests—an observation consistent with Salthouse (1996).

Modeling Stereotype Threat Effects

The anxiety resulting from stereotype threat could potentially interfere with performance in various ways, for instance, via both automatic and strategic processes (Beck & Clark, 1997), and causing a diversion of resources from the task, heightened self-consciousness, or over-cautiousness (Steele & Aronson, 1995). Diversion of resources from task-relevant to threat-relevant information processing translates to a reduction in the overall information processing throughput on the task. A slowdown in information processing might even be effortful (Gazzaley, Cooney, McEvoy, Knight & D'Esposito, 2005). Using our model as an exploratory framework, this would be analogous to a reduction in the propagation gain α of the network model, or suppression (i.e., negative bias) of activation on all nodes in the network, or both. In the case of the former, a reduction in propagation gain would lead to an increase in the amount of time required for every unit increase in a node's activation. In the latter, the negative bias due to the suppression would need to be countered before activation can be accumulated to a level that approaches the threshold for a key-press response. The effect is the same as directly increasing the response activation threshold itself.

This brings us to the second possible coping mechanism, namely the deliberate act of exercising greater caution in the completion of the tasks, wherein subjects might require of themselves a higher degree of confidence before committing to a response. This explanation is similar in spirit to Brendl, Markman, and Messner's (2001) suggestion that subjects might increase their response activation thresholds in response to an increase in the perceived difficulty of the task (i.e., with tasks in the incompatible task blocks being more difficult or demanding than those in other task blocks). In addition, the increase in confidence criteria is consistent with Beck and Clark's (1997) proposal that anxiety activates reflective modes of thinking. In our terms, the process of exercising additional caution or adopting more stringent response criteria is analogous to an upward shift in x_{thres}, which is the level of activation that $action_L$ or $action_R$ must reach before a motor action is performed.

We have already shown in the previous section that a decrease in propagation gain would be accompanied by longer response latencies as well as more pronounced IAT effects (see Figure 2). Without having to repeat this simulation, the same reasoning in explaining the relationship between cognitive fluency and performance on the IAT can be applied here to account for the increase in IAT effects due to slowdowns in information processing (as a result of task interference). Such reductions in the rate of information processing (or propagation gain, as shown in Figure 2) would have resulted in stronger IAT effects, confirming the observations in Frantz et al. (2004).

Our next and remaining task is to examine the impact that increasing the response threshold x_{thres} would have on IAT performance. So far, x_{thres} has been set to the maximum pos-

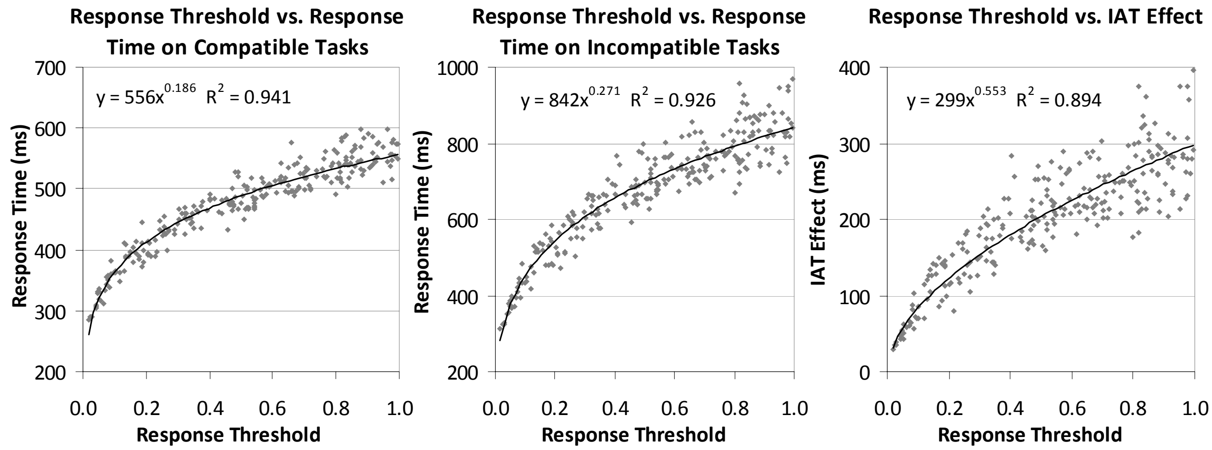

Figure 4. Distributions of response latencies on compatible and incompatible task blocks and corresponding IAT effects with variations in response threshold x_{thres} for 250 virtual subjects configured with a relative preference for one target concept over the other.

sible activation level of 1.0, but in this simulation, its value is varied within a reasonable range to determine how it affects response latencies on the combined tasks and the corresponding IAT scores. We begin by generating a population of 250 virtual subjects and initializing them with the same associative strengths as in the previous simulation, namely with EA↔POS and AA↔NEG set to 0.5 while AA↔POS and EA↔NEG were set to -0.5. Connections in each virtual subject's network were then randomly perturbed with Gaussian noise, before being put through all five standard tasks on the simulated Race-IAT.

Simulated response latencies on both the compatible and incompatible task blocks, and corresponding IAT effects are shown in Figure 4. We observe that higher response thresholds resulted in longer response latencies on both combined task blocks, as well as more pronounced IAT effects. In addition, the increase in response threshold is accompanied by higher variance in both response latencies and IAT effect. Thus, to the extent that an increase in the response activation threshold can be interpreted as the coping strategy of exercising caution or raising the confidence criterion, such a strategy could be responsible for a portion of the increase in IAT effect, as predicted by the simulation.

Discussion

The fact that cognitive ability has a moderating effect on IAT performance has important implications for the validity of the IAT as a measure of the automatic associations and not some other construct. Suppose two subjects have IAT effect scores in the same direction but with different magnitudes. The fact that one of them has an IAT effect with a larger magnitude than the other would not necessarily imply that he or she definitely endorses a stronger positive implicit attitude towards the favored target concept, or a stronger negative implicit attitude towards the less preferred target concept, since the larger IAT effect could be due to differences (e.g., a reduction) in cognitive ability. In the case of this simulation, the associative strength configurations in all 250 subjects were the same (apart from their superposed random perturbations), yet they had a wide distribution of response latencies and IAT effects through just the manipulation of the processing gain and decay rate. For this reason, other means of calculating IAT effects, such as standardizing IAT scores with the "improved scoring algorithm" (Greenwald, Nosek & Banaji, 2003) should be utilized if a reasonable between-subject comparison is desired, although there is continued debate on this matter (cf. Blanton, Jaccard, Gonzales & Christie, 2006; Nosek & Sriram, 2006).

The computational model that we have employed provides some insight about reactions or strategies that might be adopted by subjects during the IAT to cope with stereotype threat. The idea that subjects actually engage in the effortful control of information processing (through either enhancement or suppression) is supported by empirical evidence from fMRI and EEG studies (Gazzaley, Cooney, McEvoy, Knight & D'Esposito, 2005). Furthermore, the ability to exercise greater caution in terms of focusing attentional resources and raising the response threshold or confidence criterion (Treisman & Faulkner, 1984; Petrusic & Baranski, 2009) is equally plausible, especially if the presence of stereotype threat increases the perceived difficulty of the tasks (cf. Brendl, Markman & Messner, 2001). The remaining research question lies in clarifying the nature of these mechanisms through experiment involving human subjects, especially with the advent of neuro-imaging techniques (e.g., Gazzaley et al., 2005; Stanley, Phelps & Banaji, 2008).

Conclusion

Using a computational model of performance on the IAT, we examined the influence that cognitive fluency, and strategies for coping with stereotype threat could have on the processing of implicit attitudes. By varying several critical model parameters that are analogous to human cognitive ability (such as processing speed and information retention ability), the model accounts for the correlation between longer response latencies and IAT effects that arises with lower cognitive ability. Furthermore, a reduction in information processing, and the adoption of a more conservative response criterion (which was modeled as shifts in response thresholds) was found capable of reproducing the exacerbated IAT effects that were empirically observed when stereotype threat was present.

Acknowledgments

We wish to thank the anonymous reviewers for their valuable feedback and suggestions. B.-K. Quek is supported by a postdoctoral fellowship from the Agency for Science, Technology and Research (A*STAR), Singapore.

References

Beck, A. T., Clark, D. A. (1997). An information processing model of anxiety: Automatic and strategic processes. *Behavior Research and Therapy*, *35*, 49–58.

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

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

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

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *69*, 229–238.

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

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience*, *17*, 507–517.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem and stereotypes. *Psychological Review, 102,* 4–27.

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

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197–216.

Hummert, M. L., Garstka, T. A., O'Brien, L. T., Greenwald, A. G., & Mellott, D. S. (2002). Using the Implicit Association Test to measure age differences in implicit social cognitions. *Psychology and Aging*, *17*, 482-495.

Hunt, E. (1983). On the nature of intelligence. *Science*, *New Series*, *219*, 141 - 146.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81,* 774–788.

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

Lansman, M., Donaldson, G., Hunt, E., Yantis, S. (1982). Ability factors and cognitive processes. *Intelligence*, *6*, 347–386.

McFarland, S. G., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition*, *20*, 483–510.

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*, 1180–1192.

Nosek, B. A., & Sriram, N. (2007). Faulty assumptions: A comment on Blanton, Jaccard, Gonzales and Christie (2006). *Journal of Experimental Social Psychology*, *43*, 393–398.

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

Page, M. (2000). Connectionist modeling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23*, 443–512.

Petrusic, W. M., & Baranski, J. V. (2009). Probability assessment with response times and confidence in perception and knowledge. *Acta Psychologica*, *130*, 103–114.

Quek, B.-K., & Ortony, A. (2011). Modeling underlying mechanisms of the Implicit Association Test. 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*. Berlin: Springer.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test: Dissociating salience from associations. *Journal of Experimental Psychology, 133*, 139–165.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Stanley, D., Phelps, E., & Banaji, M. (2008). The neural basis of implicit attitudes. *Current Directions in Psychological Science*, *17*, 164–170.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African-Americans. *Journal of Personality and Social Psychology*, *69*, 797-811.

Treisman, M., & Faulkner, A. (1984). The setting and maintenance of criteria representing levels of confidence. *Journal of Experimental Psychology*, *10*, 119–139.