Running Heading: ONTOLOGICAL TEST OF THE IAT

An ontological test of the IAT: Self-activation can increase predictive validity

Marco Perugini[1]

Rick O'Gorman[2]

Andrew Prestwich[1]

[1]Department of Psychology, University of Essex (UK)

[2]Department of Psychology, University of Kent (UK)

KEYWORDS: IAT, self-activation, predictive validity

An ontological test of the IAT: Self-activation can increase predictive validity

Abstract

Extensive research has been conducted demonstrating the predictive validity and reliability of the IAT for a broad array of behaviors and contexts. However, less work has been done examining its underlying construct validity. This contribution focuses on examining whether a core theoretical foundation of the IAT paradigm is valid, specifically, whether the IAT effect draws on the Social Knowledge Structure. We present four studies within different domains that show that the IAT does indeed appear to draw on the SKS. The data show that activation of the self before the categorization task enhances the predictive validity of the IAT, as one would expect if the IAT reflects the SKS. We discuss theoretical reasons for these findings, with emphasis also on underlying statistical/psychometric issues.

Implicit measures have achieved a prominent status in psychological research in the last few years. The Implicit Association Test (IAT) represents the most popular of these measures. Since the original paper in which the IAT was introduced (Greenwald, McGhee, & Schwartz, 1998), dozens of studies have applied the paradigm to an impressively diverse array of issues (for a review, see Poehlman, Uhlmann, Greenwald, & Banaji, 2006). There have been relatively fewer attempts investigating the mechanisms underlying the IAT. Moreover, most of the research addressing this issue has focused on the specifics of the IAT effect (Rothermund & Wentura, 2001), on the influence of confounding effects in the IAT score such as words familiarity (Dasgupta, McGhee, Greenwald, & Banaji, 2000) and extra-personal factors (Olson & Fazio, 2004a), and on the impact of contextual effects on IAT scores (Mitchell, Nosek, & Banaji, 2003). In this contribution we will examine one of the basic assumptions of the IAT, namely that it reflects associative links in the Social Knowledge Structure (SKS; Greenwald, Banaji, Rudman, Farnham, Nosek, & Mellot, 2002).

The SKS Assumption in the IAT

The IAT is a double discrimination task used to measure the relative strength of the associations between pairs of concepts. Even though it is a relatively new paradigm, it has rapidly become a widespread tool in social psychological research. The IAT relies on the assumption that, if a target concept and an attribute concept are highly associated (congruent), the task will be easier, and therefore quicker, when they share the same response key than when they require a different response key (for procedural details, see Greenwald, McGhee, & Schwartz, 1998). The theoretical basis of the IAT, and one of its most important assumptions, relies on its tapping into the SKS. The SKS is a network of variable-strength associations that correspond to social

psychological concepts (self-concept, self-esteem, stereotype, attitude) and attributes (Greenwald et al., 2002, p. 5, Figure 1), presumably stored in long-term memory. The self is a central entity in the SKS. This centrality is represented by "…its being associated with many other concepts that are themselves highly connected in the structure" (ib., p. 5).

The assumption of the SKS is central to the logic of the IAT. It provides both a theoretical foundation and a rationale for its capability to predict behaviors. From a theoretical point of view, the SKS represents the link between measure and concepts. An attitude towards an object is a stored evaluation in memory, relatively stable over time, and can be activated automatically (e.g., Fazio, 1990). The SKS therefore represents a theoretical bed that accommodates the view of attitudes as associations between objects (actions, groups) and valence[1]. From a predictive point of view, an IAT should predict some germane behaviors if it genuinely reflects personal associations between the relevant target and valence. There is at least one sense in which this is an essential requirement. Suppose that the IAT simply reflects specific mechanisms underlying the cognitive operations activated by the task. The ranking of individuals in the resulting IAT score should therefore be affected only by individual differences in the operation of such mechanisms (method variance), for instance, stimulus-response compatibility or task-switching costs (e.g. Mierke & Klauer, 2003). However, it is unclear how individual differences in task switching costs, for example, could predict specific behaviors such as condom use or food choice.

To sum up, a focus on predictive validity appears not only informative about the pragmatic value of the IAT, but also important in terms of its theoretical foundations, namely the SKS assumption.

The Mechanism of Self-activation

Self-activation can be defined as the cognitive activation of any kind of self-related knowledge; it does not necessarily require conscious awareness, and it is characterized by a general heightened state of accessibility of self-related knowledge (Stapel & Tesser, 2001, p. 743). The constructs of self-focus and self-awareness are often used interchangeably with the concept of self-activation. It has been proposed that they should be distinguished from self-activation, mainly because the latter does not require reflective conscious self-attentiveness (Stapel & Tesser, 2001, footnote 1, p. 743). While the argument put forward appears compelling, an inspection of the actual use of the terms in published research produces a far less clear picture. Experimental manipulations that would seem indistinguishable in terms of conscious activation, are labeled as self-focus, self-awareness, or self-activation manipulations in different contributions (e.g., Dijksterhuis & Van Knippenberg, 2000; Macrae, Bodenhausen, & Milne, 1998). Here we will refer to the term self-activation in its generic sense and sidestep these more subtle differences that in practice seem as much real as a question of semantics.

Self-activation typically results from an experimental manipulation that renders self-related cognitions especially salient or accessible. The consequences of this increased accessibility can be diverse. For instance, it has been shown that self-activation increases social comparison (Stapel & Tesser, 2001), the efficiency of self-regulatory processes (Carver & Scheier, 1981), the attitude-behavior consistency (Pryor, Gibbons, Wicklund, Fazio, & Hood, 1977), and decreases stereotyping (Macrae, Bodenhausen, & Milne, 1998). On a different but relevant stream of research, it has been shown that individuals high in self-consciousness, characterized by chronically higher accessibility of self-related knowledge, are particularly sensitive to experimental manipulations such as subliminal priming used to instigate automatic

non-conscious behavior (Hull, Slone, Meteyer and Matthews, 2002). Whilst the specific consequences of self-activation can be diverse and also probably influenced by the demands of the subsequent task, the mechanism is pretty much the same. Self-activation increases the accessibility of self-relevant thoughts and constructs. This affects subsequent activities in the direction of the thoughts and constructs that are momentarily more accessible.

Testing the SKS Assumption via Self-activation

The links between the IAT, the SKS, and self-activation should be apparent at this point. To summarize, the IAT is a task of which the outcome depends on the difference in speed of the motor actions (i.e., movement of the index finger) needed to categorize correctly and which is reflected in different response latencies. This response speed is critically influenced by the relative ease or difficulty in activating the necessary motor command. This relative ease or difficulty, in turn, depends on the strength of the associations between the two pairs of concepts that might interfere or facilitate the use of the same response key. The strength of these associations is reflected in the SKS. Self-activation increases the accessibility of self-related thoughts and concepts. Therefore, it should follow that an IAT completed immediately after a self-activation manipulation should better reflect the SKS and, as a consequence, an IAT score so obtained should be more predictive of actual behavior, as it contains a relatively greater proportion of valid variance (variance that reflects the SKS).

Overview of the Studies

In sum, our key hypothesis is that self-activation should increase the predictive validity of the IAT. This prediction relies on an important assumption underlying the IAT, namely that it reflects concepts and valences as stored in the SKS. We present four studies that test this hypothesis. Study 1 concerns attitudes toward alcohol, Study

2 focuses on attitudes toward academic disciplines and includes students from Arts and from Science departments, Study 3 examines attitudes toward junk food, and Study 4 is about attitudes towards Americans. The key criteria to be predicted are self-reported behaviors (Study 1 and 3), group membership (Study 2) and actual behaviors in the form of judgments (Study 4). The results show that across attitudes, behaviors, and manipulations, self-activation increases the predictive validity of the IAT.

Study 1: Alcohol

Drinking alcohol is a relatively common behavior. The most recent national survey in the United Kingdom estimated that adults aged 14 and over drink on average 11.3 units of alcohol per week (Institute of Alcohol Studies, 2005). Some studies have applied the IAT (or modifications of it) to the issue of drinking alcohol with promising results (for a review, see Wiers, Houben, Smulders, Conrod, & Jones, 2005). For instance, Wiers, Van Worden, Smulders, and De Jong (2002) found a significant relationship (r=.37) between a standard valence IAT and a composite index of alcohol use. Typically, the studies have focused on predicting some kind of composite index of alcohol consumption rather than a consumption index of alcohol relative to soft drinks, even when the IAT measure has been defined using a contrast category of soft-drinks. Therefore, whereas there is some empirical evidence of predictive validity of an IAT for alcohol consumption, little is known to what extent it can predict a relative preference over soft-drinks consumption. The first study explores this issue and tests the key hypothesis that self-activation will increase the predictive validity of the IAT.

Method

Participants. The sample consisted of 60 participants, 27 males and 33 females,

with an average age of 26.2 years (SD=5). Of these, 48 (80%) were successfully re-contacted by e-mail after one week to obtain a second behavioral measure. The participants were predominantly university students and were contacted on campus or through informal networks.

Design and Procedure. The design was a simple 2-condition between-subjects factor. Participants were told that they would be completing two experiments and were randomly allocated to either a self-activation or a neutral condition. Each participant was tested individually with a laptop in different locations. Care was taken that during the experiment no external distractions or noises were present. The first experiment was presented as a pilot study on proofreading and word-search whereas the second was a study about their preferences towards different types of drinks. The first experiment was actually the self-activation manipulation, modeled after Brewer and Gardner (1996) and used in other studies on self activation (Stapel & Tesser, 2001)[2]. Participants were asked to read paragraphs describing a trip to a city and to circle certain words within two minutes. The text was identical, but the words to be encircled were different in the two conditions. In the self-activation condition, the words were "I", "me", "my", and "myself", whereas in the neutral condition the words were "the" and "a". In both cases there were 19 such words. Participants were then asked to perform the next tasks at a laptop with a 14.1-inch display set at a resolution of 1024 x 758, color depth set at 16 bit and refresh rate at 72Hz. The tasks were programmed with Inquisit (version 1.33).

First, participants completed an IAT on alcohol vs. soft drinks. Our implementation of the IAT followed the established format of seven steps (cf. Greenwald, McGhee, & Schwartz, 1998; Greenwald, Nosek, and Banaji, 2003). The target category pairing was Alcoholic drinks (beer, wine, whisky, lager, cider) and

Soft drinks (coke, pepsi, fanta, sprite, juices) whereas the attribute categories were Pleasant (happy, smile, joy, peace, pleasure) and Unpleasant (pain, death, poison, agony, vomit). There were 20 practice trials for the non-critical steps (steps 1, 2, and 5), 20 steps for the training stage of the critical pairs (steps 3 and 6), and 60 trials (plus two dummy initial trials that were discarded) for the critical steps 4 and 7. The stimuli were presented in a random order for all participants. The order of steps 3-4 and 6-7 was fixed for all participants, with Alcoholic drinks paired with Pleasant in step 3-4 and with Unpleasant in step 6-7. Participants were asked to press the left key (letter d) or the right key (letter k) depending on the category of the stimulus. An error message consisting of an acoustic beep was delivered upon incorrect classification. The inter-trial interval was 400 ms.

Participants were then asked their explicit attitude, first, towards drinking alcohol and, then, soft drinks. They responded to the stem "I think that to drink alcohol (soft drinks) is for me:" followed by 7 semantic differential pairs of adjectives (bad-good, foolish-wise, unpleasant-pleasant, negative-positive, unenjoyable-enjoyable, unhealthy-healthy, unattractive-attractive) on a 7-point scale. Next, participants completed a self-reported behavioral grid asking them to report how many units of alcoholic and soft drinks they usually consumed for each day of an average week. The concept of a unit of alcohol is commonly used in the UK and corresponds to specified approximated quantities of different types of alcoholic drinks. For instance, one unit of alcohol corresponds to a small, 125 ml. glass of wine, half a pint (i.e., 284 ml) of beer/cider/lager, and a standard measure (25 ml.) of spirits (e.g., whisky). To further reduce idiosyncratic reporting, a small legend reported the units corresponding to each alcoholic and soft drink. The list of alcoholic drinks included beer, wine, lager, spirits, cider, alcopops, and other alcoholic drinks, whereas

the list of soft drinks included Coke/Pepsi, lemonade, juices, and other soft drinks. Finally, participants were thanked for their participation and were informed that there would be a brief final part of the experiment in one week. They were asked for an e-mail contact address. After one week, participants were sent the previously described self-reported drinking grid and asked for their drinking behavior in the previous, rather than an average, week.

Data analysis strategy. The same data analysis strategy was used in the four studies. We first inspected the psychometric properties of the measures and report relevant descriptive aspects of the data. Next, a regression approach was adopted, centering variables before calculating interaction terms to reduce unessential multicollinearity (Cohen, Cohen, West, & Aiken, 2003). To test the key hypothesis of increased predictive validity of the IAT under self-activation manipulation, we ran regressions in which each dependent variable in the study was predicted by the IAT score, the experimental condition (dummy coded as 0=control and 1=self-activation), and their interaction. The first order effect term for the IAT would reflect the slope of the regression line in the control condition. A positive significant interaction term would signal a successful test of the hypothesis, indicating that the IAT score has higher predictivity in the self-activation condition. The interaction was further probed by reversing the dummy coding to inspect the effect of the IAT in the self-activation group (Aiken & West, 1991).

To establish whether the hypothesized effect was unique to the IAT, three additional sets of regressions were conducted[3]. The first set tested whether the same effect was present for explicit attitudes. The independent variables were therefore the explicit attitude score, the experimental condition, and their interaction. The lack of a significant interaction term in the latter regression set would signal that self-activation

works uniquely with the IAT. The second regression set tested the possibility that the interaction between the IAT and the self-activation condition is due to the shared variance between the implicit and the explicit measures. If that is the case, the effect should vanish if the explicit attitude measure is included in the equation. In other words, we tested whether the moderation effect is mediated or suppressed by the explicit attitudinal measure. Finally, the third set tested the specific issue of whether self-activation increases the correspondence between implicit and explicit measures of attitude. The regression therefore included the explicit attitude score as the dependent variable and the IAT score, the experimental condition, and their interaction as the independent variables. The lack of a significant interaction term would suggest that self-activation does not simultaneously increase the salience of propositional (i.e., explicit) and associative (i.e., implicit) associations. Taken together, once the presence of a significant effect of self-activation on the predictive validity of the IAT is established, these three additional analyses should clarify the extent to which the self-activation manipulation works primarily or uniquely at an implicit level.

Results and Discussion

The IAT score was calculated with the algorithm D (deletion of latencies below 400ms, errors replaced with the mean of the correct responses plus 600ms) developed by Greenwald, Nosek, and Banaji (2003), included all the 80 trials (20 practice and 60 test), and was calculated such that the practice and test stages had a weight proportional to the number of trials included in each (in this case 25% and 75%, respectively). The reliability of the IAT score was good (α=. 80). It was obtained by calculating 80 IAT scores (one for each pair of trials) and using them as items. The IAT score was computed such that higher scores expressed an implicit preference towards alcoholic over soft drinks. The attitude score was calculated as the difference

between the sums of the semantic differentials, with positive scores indicating a preference for alcohol over soft drinks, and showed good reliability (α=. 90). The correlation between implicit and explicit attitudes was not significant (r=.16, p=.234). The means of the measures for the two groups (self-activation vs. control) for all four studies are reported in Table 1.

[Insert Table 1 about here]

The groups did not differ in their explicit (t(58)=0.45, p=.652) and implicit attitudes (t(58)=0.04, p=.970). Mean units of drinks in an average week and in the last week varied between 8.9 (soft drinks, average week) and 11.7 (alcohol, average week). The two groups (self-activation vs. control) did not differ in terms of drinking behavior (all p's >.45). These results suggest that the assignment was effectively random. Two indices of relative preference for drinking alcohol (positive values) or soft drinks (negative values) was calculated by subtracting the total amount of units of soft drinks from those of alcohol, both for an average week and the previous week. The two indices were correlated significantly (r=.52) and aggregated in an overall index of relative alcohol consumption[4].

The multiple regression to test the key hypothesis explained 21.3% of the variance, with a significant effect of the experimental condition (β=.58, p=.039) crucially qualified by the expected significant interaction (β=.57, p=.029). The IAT was not a significant predictor in the control condition (β=-.09, p=.586) whereas it significantly predicted the drinking alcohol index in the self-activation condition (β=.48, p=.017). The two simple slopes are presented in Figure 1.

[Insert Figure 1 about here]

A second set of regressions ascertained that the effect was not present for explicit attitudes. The drinking index was significantly and strongly predicted by

explicit attitudes (β=.54, p<.001), but the interaction term with the experimental condition was not significant (β=.18, p<.459).

The third set of regressions showed that explicit attitudes mediated or suppressed the self-activation effect on the IAT. In fact, the inclusion of the explicit attitudes as a predictor (β=.47, p=.001) rendered the interaction term IAT x Experimental condition no longer significant (β=.30, p=.209).

Finally, we tested whether the correlation between IAT and explicit attitudes changed as a function of the self-activation condition. The results showed that the interaction term involving IAT and experimental condition was significant (β=.58, p=.033). IAT and explicit attitudes were not significantly correlated in the control condition (β=-.09, p=.596) and significantly associated in the self-activation condition (β=.49, p=.020).

The results provide initial support for the idea that self-activation increases the predictive validity of an IAT measure. In fact, under the condition of self-activation, the IAT predicts the relative preference for drinking alcohol over soft-drinks, whereas it does not under the control condition. The additional analyses qualified the effect. Although the self-activation effect was not present for explicit attitudes, they mediated or suppressed the self-activation effect on the IAT. Finally, explicit attitudes and IAT were significantly correlated under the self-activation condition. Taken together these results suggest that the self-activation manipulation simultaneously enhanced both propositional and associative structures concerning preferences for alcohol and soft drinks. Therefore, while self-activation has an effect on the validity of the IAT, this effect seems to be driven by an enhanced salience of germane propositional evaluations.

Study 2: Arts vs. Science

The type of studies students choose to pursue at University level is an important definer of their professional future as well as becoming part of their personal identity. Several reasons underlie what kinds of studies are pursued, including career perspectives and financial success. One of the key reasons is their liking of the type of study that they will pursue. It is reasonable to expect that a student who has chosen to study History, for example, has a stronger preference for arts over science and, conversely, that a student who has chosen Computer Science has a stronger preference for science over arts. This simple argument can be extended to implicit measures like the IAT. Nosek, Banaji, and Greenwald (2002) demonstrated that students, especially women, generally have an implicit preference for arts over science. However, the main focus of the authors was on the stereotypic association between male and science and female and arts and not on the preference of arts students for arts and science students for science. In this study we will investigate this latter issue and test whether self-activation can increase the validity of the corresponding IAT measure.

Method

Participants. The sample consisted of 72 participants, 30 males and 42 females, with an average age of 24.6 years (SD=5.6, two missing values). Participants came from a range of departments classified either as Arts or Science. The most represented departments for arts in the participant pool were Language and Linguistics (10), History (9), and Literature (6) and, for science, Biology (12), Computer Science (10), and Electronics (9).

Design and Procedure. An equal number of participants from Arts and from Science departments was randomly allocated to either self-activation or neutral conditions. Participants were tested in individual cubicles in the laboratory. The instructions

mirrored the ones in the first study. After the self-activation task, they completed an IAT on Arts vs. Science. The IAT had the same format as in Study 1, with the following exceptions.

First, the display was 15-inches and participants responded by use of a Cedrus response box (model RB-730). Second, the IAT was counterbalanced for steps 3-4 and 6-7 (approximately half the participants had Arts paired with positive in step 3-4 and Science paired with positive in step 6-7, and half had the opposite sequence). Third, the error message in the IAT consisted of a red cross displayed below the stimulus and stayed on the screen until participants pressed the correct answer (built-in error penalty). Fourth, the number of trials in the non-critical steps was slightly lower (16 instead of 20). The paired target category of Arts had History, Philosophy, Literature, Language, and Art History as exemplars while the Science exemplars were Biochemistry, Mathematics, Electronics, Computer Science, and Biology. The attribute categories were positive (good, life, pleasure, pretty, friend) and negative (evil, death, pain, ugly, enemy).

Next, participants were asked their explicit attitudes towards, first, science and, then, arts. They were presented with the stem "I think that scientific (artistic) disciplines are:" followed by six semantic differential pairs of adjectives (bad-good, negative-positive, unenjoyable-enjoyable, boring-exciting, unattractive-attractive, worthless-worthwhile) with the same 7-point scale as in the first study. Finally, participants were thanked for their participation, debriefed and paid.

Results and Discussion

The IAT score was calculated with the algorithm D for built-in error penalties and it showed good reliability (α=.93). It was computed such that higher scores expressed an implicit preference towards Arts over Science. The explicit attitude

score was calculated as the difference between the sums of the semantic differentials with the same direction (α=.90). Implicit and explicit attitudes were significantly correlated (r=.50, p<.001). The self-activation group did not differ from the control group in terms of explicit (t(70)=0.57, p=.570) and implicit (t(70)=0.01, p=.989) attitudes.

The main analysis involved a logistic regression with group membership as the dependent variable and IAT, experimental condition, and the interaction term as independent variables. The order of presentation within the IAT (Arts-Positive first vs. last) was included as a covariate in the analysis to partial out its effects (cf. Perugini & Gallucci, 2006). The equation explained 57.7% of variance (Nagelkerke R^2). There was a main effect of order (B=-2.23, p=.008)[5] and no effect for the experimental condition (B=-0.63, p=.445), whereas the IAT was a significant predictor (B=1.48, p=.005). Crucially, this effect was qualified by a borderline significant interaction between the IAT and the self-activation condition (B=2.72, p=.051). The interaction is graphically depicted in Figure 2.

[Insert Figure 2 about here]

While in both conditions the IAT predicts well the probability of being a student of Arts or Science faculties, it does so better in the self-activation condition, as can be evinced by the steeper slope of the curve. Expressing the results differently, the correlation between group membership and IAT scores was r=.36 (p=.030) in the control condition and increased to r=.76 (p<.001) in the self-activation condition.

A second logistic regression ascertained whether the same moderation effect can be found for explicit attitudes. Explicit attitudes had a significant main effect (B=1.63, p=.008), meaning that students of arts had a better explicit evaluation of arts than science students (M=1.44 vs. M=-0.78). However, this effect was not qualified

by a significant interaction (B=2.48, p=.118). A third logistic regression inspected the role of explicit attitudes as a potential mediator of the moderation effect of self-activation on the IAT. Explicit attitudes significantly predicted group membership B=2.15, p=.003) but did not affect the interaction term IAT x experimental condition (B=3.65, p=.030). Finally, the correlation between implicit and explicit attitudes was not greater in the self-activation condition, as evidenced by a non significant interaction term in the appropriate multiple regression (β=-.00, p=.984).

The results therefore confirm the key finding of the first study. Under conditions of self-activation, the IAT was a better predictor of group membership. Moreover, unlike in the first study, the effect here was shown to be unique to the IAT. Specifically, it was not found for explicit attitudes, explicit attitudes did not mediate the effect, and the correlation between implicit and explicit attitudes was not affected by the self-activation manipulation.

Study 3: Junk food

Morgan Spurlock achieved international headlines with his movie "Supersize Me" in 2004. The movie revolves around the adverse health effects of eating junk food by illustrating what happens to the protagonist – Morgan Spurlock – as he goes through a month of eating "super-sized" McDonald's products such as hamburgers and cheeseburgers. The international success of the movie was also due to increasing concerns in Western societies about the negative effects of eating junk food. Obesity, one of the key consequences of an unhealthy diet, now is considered as one of the biggest killers and a public health priority in several countries. Two studies with contrasting results are relevant. Maison, Greenwald and Bruin (2001) investigated in a sample of women whether an IAT with high vs. low calorie food-item categories predicts eating behavior and found a significant positive relation (r=.34). However, it

should be noted that their dependent variable was based on generic self-reported statements (e.g., “I always eat what I want”, “When I buy something, I am always concerned about calories”) rather than on more specific patterns of eating habits. In contrast, Roefs and Jensen (2002) used an IAT with high vs. low fat categories and found that obese people have a significantly more negative implicit attitude towards high fat food than normal weight people, therefore implying a negative relation between IAT and eating behavior. In this study we will investigate a similar issue, focusing, however, on the categories junk vs. healthy food and with the basic hypothesis that self-activation will increase the predictive validity of the IAT.

Method

Participants. The sample consisted of 60 participants, 35 males and 25 females, with an average age of 27.0 years (SD=6.1). One participant failed to answer the questions concerning his/her diet and therefore was not included in the critical analyses.

Design and Procedure. Participants were randomly allocated to either a self-activation or a neutral condition and told that there were two experiments. The first experiment, presented as a pilot study, was a paper and pencil version of Silvia’s self-novelty manipulation (2002), slightly modified for the purposes of this study. Participants in the self-activation condition were asked to answer three questions aimed at explaining what makes them unique as individuals. The questions asked about what makes them different from their family, from their friends, and from their colleagues, respectively. Participants in the control condition were asked to write about one of their university classes and to describe the last time they went out to watch a movie. In both conditions they were provided with an empty box of about 2/5 of an A4 page in which to write their responses. This manipulation has been shown to be a valid manipulation of self-focused attention (Eichstaedt & Silvia, 2003; Silvia &

Eichstaedt, 2004). After the self-novelty manipulation, participants performed an IAT on Junk vs. Healthy food, with the same procedure and number of trials as in Study 2. The exemplars for the target category of Junk food were burger, chips, doughnut, fried breakfast, and chocolate bars while the Healthy food exemplars were salad, vegetables, cereal breakfast, fruits, and yoghurt. The attribute categories were positive (rainbow, happy, smile, joy, peace) and negative (pain, death, poison, agony, sickness). Participants were then asked their explicit attitude towards junk and healthy food. The format for the attitude question was the same as in the two previous studies, followed by seven semantic differential adjective pairs (bad-good, foolish-wise, unpleasant-pleasant, negative-positive, unenjoyable-enjoyable, unhealthy-healthy, unattractive-attractive) on a 7-point scale. Finally, participants were asked to report their eating habits in a usual week. The focus was on foods that are typically included in an Unhealthy vs. Healthy diet. Specifically, they were asked to indicate how many servings a week they had of a series of products. Once completed, they were thanked and debriefed.

For the composite of Unhealthy diet the items were: sausages or beefburgers; beef, pork or lamb; bacon, meat pie, processed meat; any fried food (including cooked breakfast) [all in a 5 step scale: none, < 1, 1 to 2, 3 to 5, 6 or more]. For the composite of Healthy diet they were: Breakfast cereals a) Sugared type: e.g., Frosties, Coco Pops; Rice or Corn type: e.g., Corn Flakes, Special K; b) Porridge or Ready Brek; Wheat type: e.g., Weetabix, Fruit 'n' Fibre; Muesli type: Alpen, Jordan's; c) Bran type: All-Bran, Bran Flakes, Sultana Bran [all in a 5 step scale: none, < 1, 1 to 2, 3 to 5, 6 or more]; Fruit: fresh, frozen or canned [in a 7 step scale: none, <1, 1 to 2, 3 to 5, 6 to 7, 8 to 11, >12].

The items were standardized, averaged within the type of diet, and then a composite index of *Unhealthy Eating* was created by subtracting the Healthy from the Unhealthy diet.

Results and Discussion

The IAT showed good reliability (α=. 85). The IAT score was computed such that higher scores expressed an implicit preference towards junk over healthy food. The explicit attitude score was calculated as the difference between the sums of the semantic differentials in the same direction (α=.80). The two measures were not significantly correlated (r=.21, p=.110). Neither the explicit (t(57)=0.34, p=.735) nor the implicit (t(57)=1.04, p=.305) attitudes of the self-activation group differed from those of the control group. The Unhealthy Eating index also did not differ across conditions (t(57)=0.41, p=.678).

A multiple regression was performed on the Unhealthy Eating index with the order of presentation within the IAT, the IAT score, the experimental condition, and the interaction between IAT and experimental condition as independent variables. The regression explained 10.1% of the variance. The order of presentation (β=-.02, p=.925) and the experimental condition (β=.10, p=.634) were not significant whereas, crucially, the interaction term between IAT and experimental condition was significant (β=.47, p=.048, see Figure 3). The IAT significantly predicted unhealthy eating in the self-activation (β=.30, p=.043) but not in the control condition (β=-.17, p=.412).

[Insert Figure 3 about here]

The influence of the self-activation manipulation did not generalize to explicit attitudes. Explicit attitudes did not significantly predict the unhealthy eating (β=.06, p=.695) and, crucially, the interaction term with the experimental condition was not

significant (β=.19, p=.395). Moreover, the moderation effect of self-activation on the IAT did not vanish when explicit attitudes were included in the regression equation (β=.51, p=.031). Finally, the correlation between implicit and explicit attitudes was not moderated by self-activation, as reflected in a non-significant interaction term between IAT and experimental manipulation in predicting the explicit attitude score (β=-.28, p=.321). The results of this third study fully parallel those of the second study. The analyses showed that self-activation increases the predictive validity of the IAT. A different manipulation of self-activation has proven as effective as the one used in the first two studies. Moreover, the effect has been shown to be exclusive to the associative structures that are reflected in an implicit measure like the IAT.

Study 4: Afro-Caribbean stereotype

Whether the IAT is more predictive when the self-related knowledge structures are activated depends also on exactly what is activated. An important question therefore concerns boundary conditions of the self-activation manipulation. We believe that there are such conditions and one of these concerns stereotype activation. There is evidence that heightened self-focus can lead to spontaneous suppression of stereotypic thoughts through automatic activation of inhibitory thoughts (Macrae, Bodenhausen, & Milne, 1998). An implication of this is that self-activation may not increase the validity of a paradigm like the IAT when the content is stereotype-related. Therefore, if an IAT on stereotype-related content (e.g., a race IAT) were used to predict some prejudiced behaviors or choices, one could expect that under conditions of self-activation its predictive validity may not increase (or may even actually decrease) because of the inhibitory thoughts that can be automatically activated. Inhibitory thoughts can be considered as one of the suppression factors that are involved in the chain from stereotype to action (e.g., Crandall & Eshleman, 2003).

However, not everybody engages in inhibitory thoughts when faced with a situation that activates a stereotype.

One of the most widely validated scales to measure individual differences in the activation of control mechanisms (i.e. inhibitory thoughts) is the Motivation to Control Prejudiced Reactions (MCPR, Dunton & Fazio, 1997). The MCPR is composed of two main dimensions, *concern with acting prejudiced* and *restraint to avoid dispute*. Both dimensions have been shown to moderate the relationship between automatically activated racial attitudes and the expression of prejudice (Dunton & Fazio, 1997; Towles-Schwen & Fazio, 2003). The first dimension seems particularly relevant for our study (see below). The concern with acting prejudiced dimension is strongly related to egalitarianism and implies a particular concern toward negative biases against historically disadvantaged groups such as Blacks (Olson & Fazio, 2004b). Given that the goal of individuals who are high in concern with acting prejudiced is to treat such disadvantaged people more favorably, they may be inclined toward positive judgments and, hence, may over-correct for any negativity that they experience. It follows that for persons who are high in concern, self-activation should simultaneously activate stereotypic and inhibitory/control thoughts, therefore counteracting each other. In contrast, for persons who are low in concern, only stereotypic thoughts will be activated and the IAT should be more predictive of stereotypic-related actions.

The key hypothesis, therefore, is that self-activation will increase the predictive validity of the IAT, but only for people who are low in concern with acting prejudiced. This hypothesis would be confirmed if the corresponding interaction term between IAT, experimental manipulation, and concern is significant.

To test this idea, we focused on the Afro-Caribbean stereotype. In the United Kingdom, the white population has generally a negative stereotype of Afro-Caribbean people. Afro-Caribbeans are usually judged as more dangerous, less friendly, less competent, and more likely to be involved in criminal acts than white people or other ethnic minority groups, such as Chinese. Besides anecdotal evidence and survey studies, the stereotype is reflected in how the English police deal with Afro-Caribbean people. According to official statistics, Afro-Caribbean people are 6 times more likely to be stopped and searched by the police than are white people (Home Office, 2005).

Method

Participants. The sample consisted of 39 White university students. One participant was excluded from the analyses because the pattern of the IAT revealed a high number of rapid responses (27.5% trials <300ms) and a large proportion of errors (29.4%), indicating random responding. The final sample size was thus composed of 38 participants[6].

Design and Procedure. Participants were randomly allocated to self-activation or control conditions. They first completed the paper and pencil version of Silvia's self-novelty manipulation, with the same procedure as in Study 3. Following this manipulation, participants completed an IAT comparing relative preference for Afro-Caribbeans vs. Chinese. The IAT had the same procedure and format of Studies 2 and 3, except that it included 20 trials for the non-critical steps and for the practice stage of the critical steps. The exemplars for the Afro-Caribbean target category were typical names of Afro-Caribbean men (Leroy, Carlton, Winston, Tyrese, Denzil). The exemplars for the Chinese target category were typical names of Chinese men (Chen, Yuan, Ming, Hsin, Chung). The attribute categories were positive (rainbow, love, gift, joy, pleasure) and negative (vomit, death, evil, agony, cancer).

Participants then responded to six items assessing personal attitudes towards first Afro-Caribbean people and then towards Chinese people. Each item took the format, "In your opinion, how _____ are Afro-Caribbean (Chinese) people?" and included in turn the adjectives: competent, capable, efficient, friendly, well-intentioned, warm. These six adjectives were chosen as markers of the two dimensions of competence and warmth (see below).

Next, participants read a hypothetical case study adapted from Bodenhausen (1988): 'On Saturday 20th November a man was physically assaulted in an alleyway in North London. The victim, a young man in his twenties, claimed that he was followed down the alleyway and attacked from behind. After a brief struggle it is claimed that the victim was knocked to the floor and the victim was punched and kicked a number of times before the defendant escaped with his wallet and mobile phone.' Participants then read 12 items of evidence (pre-tested with a pilot study) consisting of both incriminating and acquitting evidence so as to make the case ambiguous overall (e.g., the defendant's ex-girlfriend testified that she had spent time in the bar flirting with the victim and had made plans to meet him later; no eyewitnesses could positively identify the attacker;). Participants were asked to rate how *guilty* they believed the defendant to be along an 11-point scale (definitely not guilty-definitely guilty).

However, before reading the hypothetical case study, all participants completed a masked primed lexical decision task (MPLD) designed to prime participants with the concept Afro-Caribbean. This was necessary in order to activate in a subtle way the Afro-Caribbean stereotype. The MPLD consisted of two blocks. The first block was presented as practice trials and consisted of 10 trials. The second block was presented as test trials and contained 38 trials. Each trial required participants to

decide as quickly as possible whether each letter string represented a word or nonword. There were 8 target words (tree, chair, circle, number, flower, button, square, insect) and 8 pronounceable nonwords (larik, kutred, bengarst, garifu, hustip, finzarit, vugerab, sitab). Within each trial, the target word was preceded by a fixation cross (that appeared for 1 second) that was replaced by a prime word (either an Afro-Caribbean name, Winston or Tyrese; or a neutral word, Neutral or Table) appearing for 42ms then followed by a mask (a row of XXXXXXXX that was longer in length than any primes or targets) that lasted on-screen for 681ms. There was an inter-trial interval of 100ms between the participant's response to the target and the onset of the next trial denoted by the presentation of the fixation cross. Participants were primed with an Afro-Caribbean name on 8 out of 10 practice trials and 24 out of 38 test trials.

After reading the hypothetical scenario and making a judgment of guilt, participants completed a 12-item stereotype scale for both Afro-Caribbean people and Chinese people, adapted from the Stereotype Content Model (SCM; Fiske, Cuddy, Glick, & Xu, 2002) that proposes two main dimensions of competence and warmth. Each item took the form: "As viewed by society, how _____ are Afro-Caribbean (Chinese) people?" along 5-point scales ranging from 'not at all' to 'extremely'. Twelve adjectives were used to assess *competence* (competent, confident, capable, efficient, intelligent, skillful) and *warmth* (friendly, well-intentioned, trustworthy, warm, good-natured, sincere). Then, participants completed the *Concern with acting prejudiced* sub-scale of the Motivation to Control Prejudice Reactions, comprising 9 items (e.g., 'I get angry with myself when I have a thought or feeling that might be considered prejudiced') along 6-point rating scales (strongly disagree-strongly agree). Finally, participants were debriefed and thanked for their time.

Results and Discussion

The IAT score was calculated, as with the previous studies, using a weighted D algorithm and it showed good internal reliability (α=.84). Higher scores reflected positive implicit preference for Afro-Caribbean people. Explicit measures of personal and societal racial views were calculated as difference scores between the sums of the scales in the same direction (Afro-Caribbean minus Chinese: personal competence, PC: α=.69; personal warmth, PW: α=.77; societal competence, SC: α=.69; societal warmth, SW: α=.83).

The IAT was not significantly correlated with any of the explicit measures (PC: r=.11; PW: r=.09; SC: r=.15; SW: r=.22) including Concern (r=-.16). Neither the explicit (PC: t(36)=0.92, p=.364; PW: t(36)=1.21, p=.234; SC: t(36)=0.14, p=.893; SW: t(36)=-0.39, p=.699) nor the implicit (t(36)=-1.03, p=.312) measures of the self-activation group differed from those of the control group. Furthermore, there was no difference in Concern scores across these groups (t(36)=-1.06, p=.297).

A multiple regression was performed on the guilt judgment with the IAT order of presentation, IAT score, experimental condition, concern with acting prejudiced, the two-way interactions between IAT, experimental condition, and Concern, and the relative three-way interaction as independent variables. The regression explained 29.9% of the variance. The order of presentation (β=.03, p=.921), the IAT (β=-.05, p=.780) and the experimental condition (β=.13, p=.474) were not significant. There was a main effect of Concern (β=.40, p=.054). None of the two-way interactions emerged (IAT x self-activation: β=-.12, p=.537; IAT x Concern: β=.07, p=.773; self-activation x Concern: β=-.30, p=.150). Crucially, the three-way interaction term between IAT, experimental condition, and Concern was significant (β=.60, p=.022). An inspection of the three-way interaction revealed that, within the self-activation condition**,** there was a marginally significant interaction between IAT and Concern

(β=.72, p=.062). Simple slopes analyses revealed that when Concern was low, implicit preference for Afro-Caribbeans was significantly negatively correlated with ratings of a guilty judgment (B=-.80,, S.E.=0.32, t=-2.47, p=.026). At intermediate (B=-.08, S.E.=0.24, t=-0.33, p=.749) and high levels of Concern (B=0.64, S.E.=0.52, t=1.23, p=0.236), IAT scores were unrelated to guilty ratings. There was no significant interaction between IAT and Concern in the control condition (β= -.51, p=.147). According to the simple slopes analysis, implicit preference for Afro-Caribbeans was unrelated to ratings of guilt when concern was low, moderate or strong (all p>.215). The simple slopes of the interactions are depicted in Figure 4.

[Insert Figure 4 about here]

To recapitulate, under self-activation, a negative implicit preference for Afro-Caribbean was highly predictive of greater guilt judgment, but only for those individuals who are not concerned with acting in a prejudiced manner--and thus do not engage in inhibitory thoughts--as hypothesized.

The influence of the self-activation manipulation was not present for any of the four explicit indices (PC, PW, SC, SW). Although there was a significant main effect of SW (β=-.40, p=.040) suggesting that those with more negative views of Afro-Caribbeans judged the defendant as more guilty, there were no significant two-way interactions between self-activation and the explicit measures (all p's >.148) or three-way interaction involving Concern (all p's >.319). Moreover, the three-way interaction between self-activation, IAT and Concern not only did not vanish when the explicit indices were included in the regression equation, but actually became marginally stronger (for PC: β=.68, p=.006; for PW: β=.69, p=.012; for SC: β=.66, p=.008; for SW: β=.62, p=.011). Finally, the correlation between implicit and explicit measures was, on the whole, not moderated by self-activation, as reflected in non

significant two-way interactions between self-activation and IAT score (all p's >.526) when predicting the four explicit stereotype measures. Furthermore, there were non-significant three-way interactions involving Concern (p's >.457 for PC, SC and SW), although there was a marginal effect for PW (β=.48, p=.071). A further probing of this effect revealed a nonsignificant tendency for an interaction between IAT and Concern under conditions of self-activation (β=.61, p=.083). Simple slope analyses revealed no significant effects at low, medium, and high values of Concern (all p's >.366).

General Discussion

The results of the four studies taken together provide support for our hypothesis. Using different methods and focusing on different domains, a self-activation manipulation has been shown to increase the predictive validity of the IAT. We will focus the remainder of this paper on some implications of these results.

Theoretical foundations of the IAT

In the last few years empirical evidence concerning the IAT has rapidly accumulated. Fewer studies have investigated the underlying mechanisms, typically focusing on cognitive accounts of the processes involved in the IAT score (e.g., Rothermund & Wentura, 2004) or on contextual effects (e.g., Mitchell, Nosek, & Banaji, 2003). The studies presented herein have tested a key foundational assumption of the IAT, namely that it reflects associations stored in the SKS structure. An implication of this assumption is that the IAT can be more predictive of relevant behaviors when it is gauged in a context with a heightened self. The results have confirmed this hypothesis and therefore support the theoretical rationale underlying the IAT. In other words, the effects of the self-activation manipulation show that the IAT score contains valid variance and not just content-free cognitive processes (that

from this perspective could be classed as method variance, cf. Mierke & Klauer, 2003). These results converge with recent empirical evidence showing that content-free cognitive accounts such as task-switching costs, salience effects, and figure-ground asymmetries cannot fully explain the composition of the IAT scores (e.g., Back, Schmukle, & Egloff, 2005). In fact, the self-activation manipulation is orthogonal to all these accounts of cognitive processes in the IAT.

Mechanisms

The results of the four studies have also provided interesting information concerning potential mechanisms underlying the effects of self-activation on the IAT. Note that the effects of the self-activation manipulations are entirely in terms of variances. It should be highlighted that for all 4 studies the mean IAT scores do not differ significantly due to the self-activation manipulation. While at first this result may appear surprising, we believe that it is fully consistent with the mechanism of self-activation and the SKS assumption. The predictive validity of any measure depends on the portion of valid variance that is shared by the predictor and the criterion. This variance is entirely insensitive to changes in the mean values. The only thing that matters is that the individual ranking in the measure more accurately reflects the differences in the strength of associations as represented in the SKS of different individuals. It follows that the self-activation manipulation should have an effect in terms of covariances (e.g., correlation between IAT and criteria) and not necessarily in terms of means, as we have found in the four studies.

Importantly, in three of the four studies the effects of self-activation have been shown to be unique to the associative structures that underlie the IAT. The effects were neither present for, nor mediated by, explicit attitudes, and there was no evidence of an increased correlation between implicit and explicit measures due to

self-activation. This would suggest that self-activation primarily renders more salient the associative evaluation but does not necessarily activate explicit cognitions and propositional evaluations. There are a number of additional considerations that can be prompted by this finding. First, note that the correlations between IAT and explicit attitudinal measures have ranged from significant to nonsignificant. We believe that this outcome reinforces our idea that using explicit attitudes as criteria to judge the quality of an implicit measure is a red herring. Once again, the focus could be more fruitfully placed on the unique predictive validity of the IAT and not on the convergent validity between measures, which is mostly a descriptive piece of information (cf. Perugini, 2005a,b). If anything, all else being equal, two measures that are more correlated are less likely to contribute unique variance to predict relevant criteria. Second, note that our results imply that self-activation can activate associative evaluations without necessarily also activating the corresponding propositional evaluations. Of course, the results do not rule out the possibility that self-activation can also work at the propositional level. Only further research can increase understanding of this specific aspect. Finally, and consequential to the previous point, our results imply that the self-activation manipulation can be distinguished from the Personalized-IAT (Olson & Fazio, 2004). This is a modified IAT that reduces the relative weight of extra-personal associations, and therefore increases the weight of the personal associations that reflect one's attitude, in the IAT score. Their modification consists of changing the valence dimension and therefore basically adopting a speeded evaluative task ("I like" vs. "I don't like") in place of the original categorization task ("positive" vs. "negative"). However, this procedure relies on systematically producing a stronger association between associative and propositional evaluations. Indeed, the authors use the increased correlation (compared

to the standard IAT) with explicit attitudes as evidence of the superiority of their paradigm.

Optimal testing

The four studies that we have presented suggest some conditions that can increase predictive validity and therefore can represent optimal testing conditions. The motor responses required in the IAT are affected by increased accessibility of self-related knowledge achieved with the self-activation manipulation. The specific temporal procedural sequence should be highlighted. In all studies, the self-activation manipulation was always performed immediately before the IAT measure and not during the whole experimental session (e.g., by using a room with a mirror) or after the measure and before the behavior or dependent variable (e.g., by using a priming manipulation). Therefore, the increase in predictive validity cannot be attributed to the behavior becoming aligned to the pre-existing attitudes as a consequence of increased self-focus (cf. Pryor et al., 1977). The manipulation directly affects the IAT and can be understood in a more general sense as additional evidence of the influence of contextual factors on implicit measures. However, unlike most other studies, what has been shown here is that the effect of self-activation is on the link between predictor (IAT) and criterion. In other words, the effects of self-activation are in terms of increasing the relationship between IAT and the relevant criteria to be predicted by affecting the portion of valid variance contained in its score.

Conclusions

In this contribution we have shown that increasing self-activation can have a significant impact on a subsequent IAT task. Although we have focused on IATs measuring implicit attitudes, it is possible that the findings generalize to other types of mental representations (e.g., self-concepts, non-evaluative stereotypes) as well as to

other types of paradigms (e.g., Extrinsic Affective Simon Task, De Houwer, 2003). Future studies will be needed to establish this generalizability. The impact of self-activation is in terms of changing the portion of valid variance and not in terms of affecting the mean IAT scores per se. As a consequence, the resulting IAT can be more predictive of relevant behaviors and choices. Future studies will be needed to clarify the specific mechanisms involved, to further define the boundary conditions of the effect, and to refine the obtained effects so as to increase understanding of the theoretical foundations of the IAT and of its optimal testing conditions.

References

Back, M.D., Schmukle, S.C., & Egloff, B. (2005). Measuring Task-Switching Ability in the Implicit Association Test. *Experimental Psychology*, *52*, 167-179.

Bodenhausen, G. V. (1988). Stereotypic biases in social decision making and memory: Testing process models of stereotype use. *Journal of Personality and Social Psychology*, *55*, 726-737.

Brewer, M.B., & Gardner, W. (1996). Who is the "we"? Levels of collective identity and self representations. *Journal of Personality and Social Psychology*, *71*, 83-93.

Carver, C.S., & Scheier, M.F. (1978). Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, *36*, 324-332.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed). Hillsdale, USA: Erlbaum.

Crandall, C.S., & Eshleman, A. (2003). A Justification–Suppression Model of the expression and experience of prejudice. *Psychological Bulletin*, *129*, 414-446.

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

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, *37*, 443-451.

De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*, 77-85.

Dijksterhuis, A., & Van Knippenberg, D. (2000). Behavioral indecision: Effects of

self-focus on automatic behavior. *Social Cognition*, *18*, 55-74.

Dunton, B. C., & Fazio, R. H. (1997). An individual difference measure of motivation to control prejudiced reactions. *Personality and Social Psychology Bulletin*, *23*, 316-326.

Eichstaedt, J., & Silvia, P.J. (2003). Noticing the self: Implicit assessment of self-focused attention using word recognition latencies. *Social Cognition, 21,* 349-361.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

Fiske, S.T., Cuddy, A.J., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*, 878-892.

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.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.K.L. (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.

Home Office (2005). *Statistics on Race and the Criminal Justice System 2004: A Home Office publication under Section 95 of the Criminal Justice Act of 1991*. London, UK: Home Office.

Hull, J.G., Slone, L.B., Meteyer, K.B., & Matthews, A.R. (2002). The Nonconsciousness of Self-Consciousness. *Journal of Personality & Social Psychology, 83*, 406-424.

Institute of Alcohol Studies (2005). *Drinking in Great Britain*. St Ives, UK.

Maison, D., Greenwald, A.G., & Bruin, R. (2001). The Implicit Association Test as a measure of implicit consumer attitudes. *Polish Psychological Bulletin*, *2*, 61-79.

Macrae, C.N., Bodenhausen, G.V., & Milne, A.V. (1998). Say no to unwanted thoughts: self-focus and the regulation of mental life. *Journal of Personality and Social Psychology*, *74*, 578-589.

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

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

Nosek, B.A., Banaji, M.R., & Greenwald, A.G. (2002). Math = Male, Me = Female, therefore Math = Me. *Journal of Personality and Social Psychology, 83*, 44-59.

Olson, M.A., & Fazio, R.H. (2004a). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, *86*, 653-667.

Olson, M. A., & Fazio, R. H. (2004b). Trait inferences as a function of automatically-activated racial attitudes and motivation to control prejudiced reactions. *Basic and Applied Social Psychology*, *26*, 1-11.

Perugini, M. (2005a). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology, 44,* 29-45.

Perugini, M. (2005b). Commentary on "Using implicit tasks in attitude research: a review and a guide" by Alexa Spence. *Social Psychological Review, 7, 21-24.*

Perugini, M., & Gallucci, M. (2006). *Order Analysis in counterbalanced measures and experimental designs*. Unpublished manuscript.

Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2005). *Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity*. Unpublished manuscript.

Pryor, J.B., Gibbons, F.X., Wicklund, R.A., Fazio, R.H., & Hood, R. (1977). Self-focused attention and self-report validity. *Journal of Personality*, *45*, 513-527.

Roefs, A., & Jansen, A. (2002). Implicit and explicit attitudes toward highfat foods in obesity. *Journal of Abnormal Psychology*, *111*, 517–521.

Rothermund, K., & Wentura, D. (2001). Figure-ground asymmetries in the Implicit Association Test. *Zeitschrift für Experimentelle Psychologie*, *48*, 94-106.

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

Silvia, P.J. (2002). Self-awareness and emotional intensity. *Cognition and Emotion, 16,* 195-216.

Silvia, P. J., & Eichstaedt, J. (2004). A self-novelty manipulation of self-focused attention for Internet and laboratory experiments. *Behavior Research Methods, Instruments, and Computers, 36,* 325-330.

Stapel, D.E., & Tesser, A. (2001). Self-activation increases social comparison. *Journal of Personality and Social Psychology*, *81*, 742-750.

Towles-Schwen, T., & Fazio, R. H. (2003). Choosing social situations: The relation between automatically- activated racial attitudes and anticipated comfort interacting with African Americans. *Personality and Social Psychology Bulletin*, 29, 170-182.

Wiers, R.W., Woerden, N.V., Smulders F. T. Y., & de Jong, P.T. (2002). Implicit and explicit alcohol-related cognitions in heavy and light drinkers. *Journal of Abnormal Psychology, 111,* 648-658.

Wiers, R. W., Houben, K., Smulders, F. T. Y., Conrod, P. J., & Jones, B. T. (2005). To drink or not to drink: The role of automatic and controlled processes in the etiology of alcohol-related problems. In R. W. Wiers & A. W. Stacy (Eds.), *Handbook of Implicit Cognition and Addiction*. Thousand Oaks, CA: Sage Publishers.

Authors Note

Correspondence concerning this manuscript should be addressed to Marco Perugini, Department of Psychology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom, Email: mperug@essex.ac.uk, Tel. 01206 874330, Fax. 01206 873590. This work was partly supported by a UK Economic and Social Research Council (ESRC) grant R000230104.

Footnotes

[1] Different accounts of the specific mechanisms underlying the IAT have been proposed (e.g., task-switching costs, Mierke & Klauer, 2003; stimulus-response compatibility, De Houwer, 2001; figure-ground asymmetries, Rothermund & Wentura, 2001). A discussion of these accounts is beyond the scope of this contribution, given that they do not challenge the assumption of the SKS but represent different explanations of how it translates into an IAT effect.

[2] We would like to thank Rob Holland for kindly providing us the self-activation manipulation.

[3] We would like to thank anonymous reviewers for suggesting that we investigate this issue.

[4] We would like to thank an anonymous reviewer for suggesting to us this strategy of analysis. The results are qualitatively similar to analyzing the two dependent variables separately.

[5] This effect of order is a spurious result that does not have a substantial interpretation, because it suggests that the order of presentation is not balanced across groups. This is not the case as evidenced by the simple association statistics (B=-.111, p=.814). However, by including the order in the final equation, its effects are partialled out from the relationships between other variables and group membership.

[6] Unfortunately, due to a mistake in the programming software, age and gender were not recorded. The sample was approximately balanced in terms of gender and with an age range between 19 and 25 years old.

Table 1

Descriptive statistics for IAT and explicit attitude scores by experimental conditions in the four studies.

	IAT (D-score)			Explicit measures	
			Study 1		
	M	SD		M	SD
Control	-0.29	0.45		-0.55	1.53
Self-Activation	-0.28	0.39		-0.72	1.40
			Study 2		
Control	0.42	0.66		0.21	1.91
Self-Activation	0.42	0.66		0.45	1.66
			Study 3		
Control	0.98	0.22		-2.78	1.41
Self-Activation	0.91	0.31		-2.75	1.43
			Study 4		
Control	-0.20	0.33	PC	-0.72	1.04
			PW	-0.52	1.10
			SC	-0.90	0.97
			SW	-0.51	1.00
			CON	3.80	1.04
Self-Activation	-0.32	0.38	PC	-0.45	0.78
			PW	-0.10	1.03
			SC	-0.86	0.84
			SW	-0.63	0.96
			CON	3.45	1.01

Note. PC: Personal Competence; PW: Personal Warmth; SC: Society Competence; SW: Society Warmth; CON: Concern with acting prejudiced

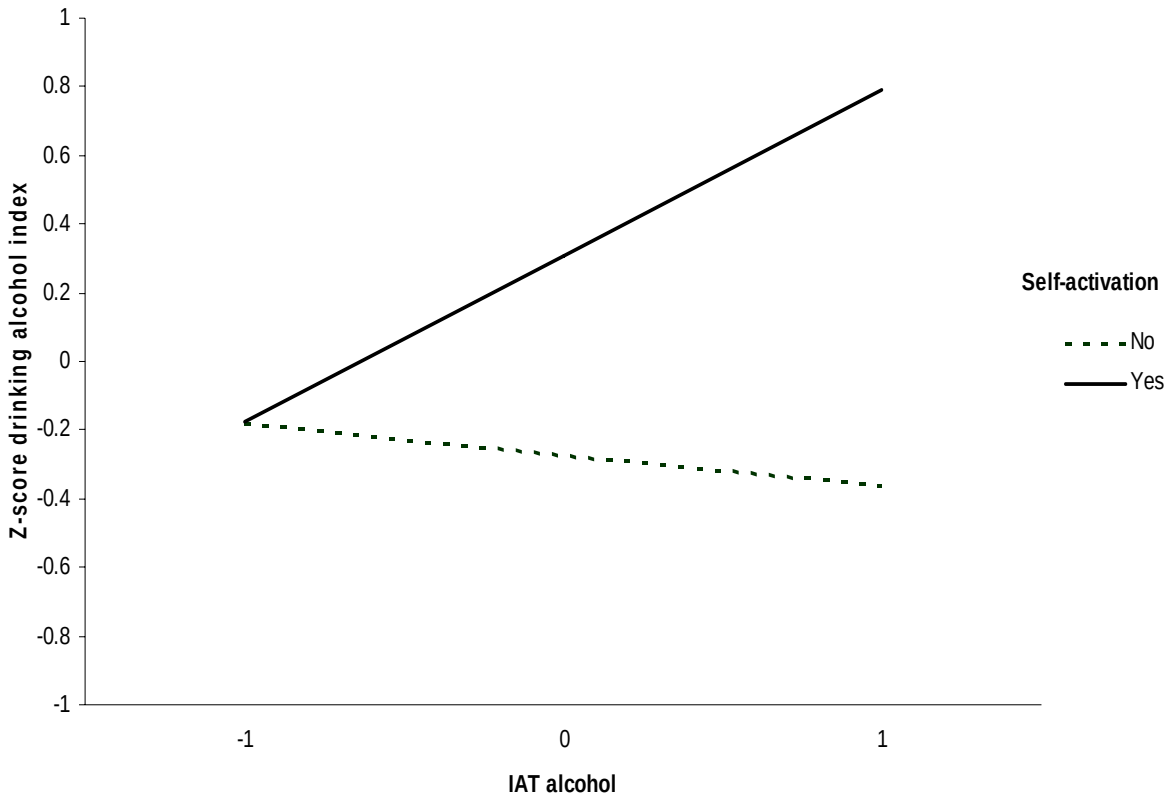
Figure Captions

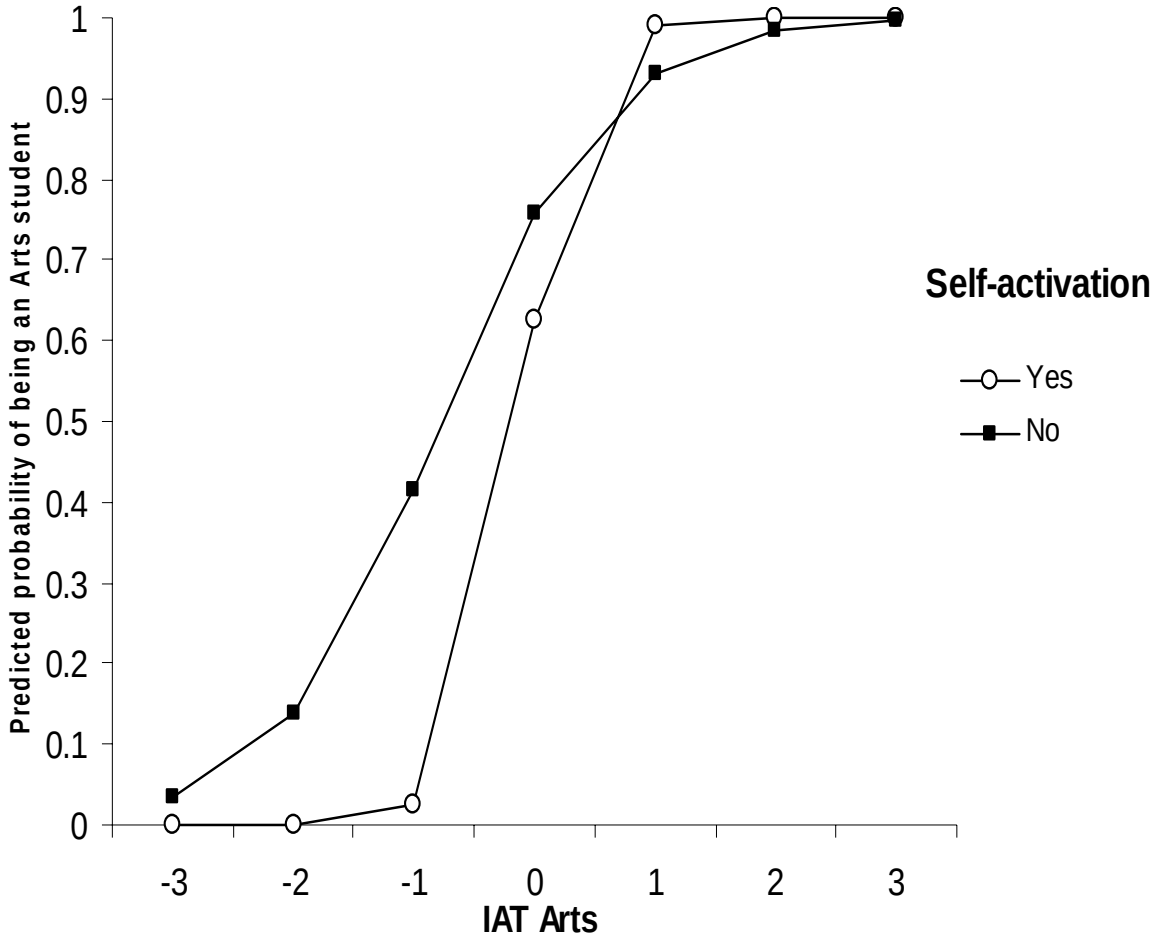
Figure 1. Study 1: Simple slopes for self-activation and control groups for the IAT Alcohol and the drinking alcohol index (standardized scores).

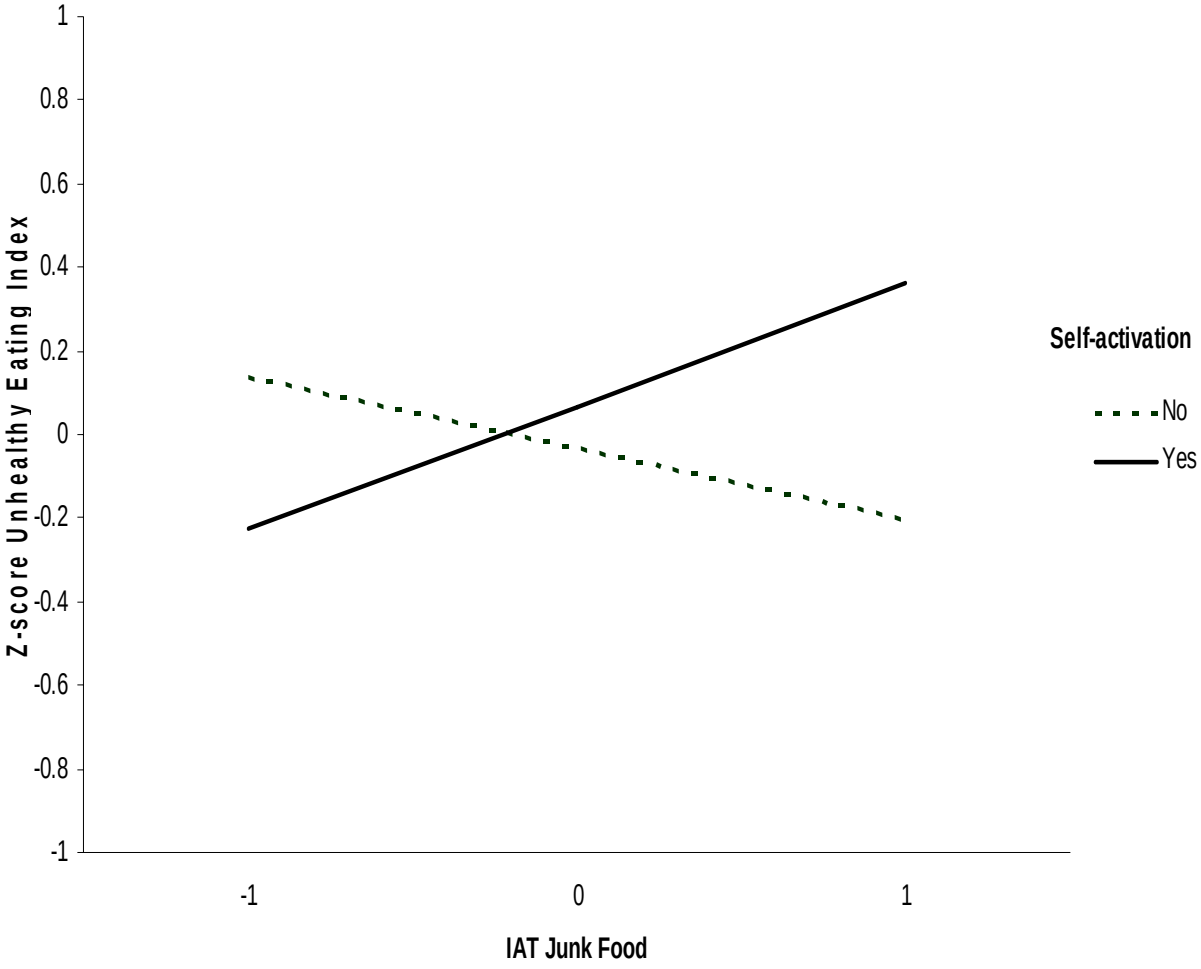
Figure 2. Study 2: Predicted probabilities of being an Arts student for self-activation and control groups as a function of the IAT Arts (standardized scores).

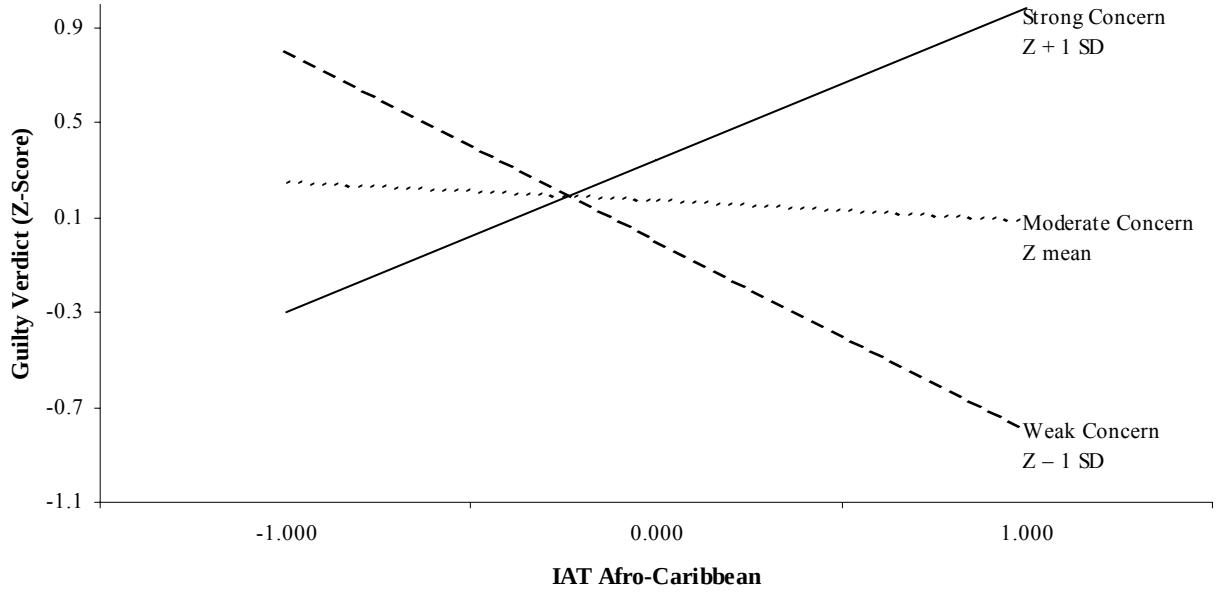
Figure 3. Study 3: Simple slopes for self-activation and control groups for the IAT Junk Food and the Unhealthy Eating index (standardized scores).

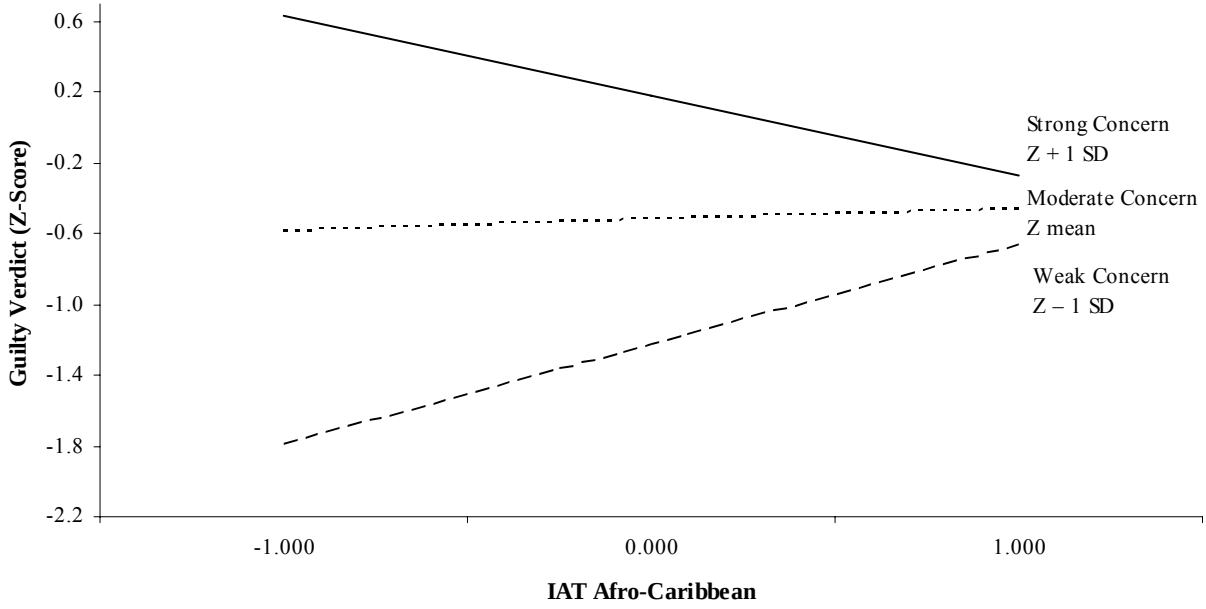
Figure 4. Study 4: Simple slopes for the interaction between IAT Afro-Caribbean and guilt judgment in the self-activation (top) and control (bottom) conditions (standardized scores).


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
Z-score drinking alcohol index
-1
0
1
IAT alcohol
Self-activation
No
Yes


Predicted probability of being an Arts student
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-3
-2
-1
0
1
2
3
IAT Arts
Self-activation
Yes
No


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
Z-score Unhealthy Eating Index
-1
0
1
IAT Junk Food
Self-activation
No
Yes


0.9
0.5
0.1
-0.3
-0.7
-1.1
Guilty Verdict (Z-Score)
Strong Concern
Z + 1 SD
Moderate Concern
Z mean
Weak Concern
Z – 1 SD
-1.000
0.000
1.000
IAT Afro-Caribbean


0.6
0.2
-0.2
-0.6
-1.0
-1.4
-1.8
-2.2
Guilty Verdict (Z-Score)
Strong Concern
Z + 1 SD
Moderate Concern
Z mean
Weak Concern
Z – 1 SD
-1.000
0.000
1.000
IAT Afro-Caribbean