This is a preprint of an article published in *European Journal of Personality*, *21*(3), 303-317. The journal can be accessed at http://onlinelibrary.wiley.com/doi/10.1002/per.633/abstract

The gatekeeper:

Individual differences are key in the chain from perception to behavior

Marco Perugini and Andrew Prestwich

Running Head: The gatekeeper model

Address correspondence to:

Marco Perugini

Department of Psychology

University of Essex

Wivenhoe Park

COLCHESTER CO4 3SQ

United Kingdom

email: mperug@essex.ac.uk

WORD COUNT: 6599 words

Abstract

A basic assumption in mainstream social cognition is that the path from perception to behaviour is often automatic and direct, as supported by several experimental studies showing that priming can lead directly to a congruent behavior without any need of conscious awareness of the process. However, we argue that priming of a goal or an object activates individual differences in automatic evaluations at the associative level that in turn are they key predictors of action (gatekeeper model). A study (n=90) on stereotype towards Americans is presented to support the model. The results show that an IAT Americans predicts a relevant action (essay evaluation) but only under condition of priming. Implications for predictive validity of implicit measures are also discussed.

A simple stimulus can have different perceptions, meanings, evaluations, and elicit different behaviors for different people. Imagine that it is 13.00, you did not have lunch yet, are walking down the road and see somebody eating with gusto a hamburger. If you like meat and hamburgers, probably the sight of that person enjoying his hamburger will solicit your taste buds. The likelihood that you will think to buy a hamburger for lunch increases. If you see a McDonald's 50 metres ahead of you, it will become instantly attractive. Chances are that you will enter and buy a hamburger for lunch. Suppose now that you don't like meat and especially hamburgers. The sight of someone visibly enjoying a hamburger when you are hungry does not elicit particularly positive reactions. You will probably feel uneasy. If you see a McDonald's 50 metres ahead of you, it will become instantly repulsive. Instead, you will be more likely to enter in the first vegetarian fast-food that you encounter to buy something to eat. Same stimulus, opposite reactions. If we add that these different reactions can vary systematically among persons in a relatively stable manner, here we enter into the realm of personality psychology. One of the key contributions of personality psychology has been in fact to devise psychometric measures that capture individual differences in basic constructs (traits, dimensions, factors). The underlying assumption is that not everybody thinks, evaluate, and behave in the same way. Rather, there are systematic inter-individual differences that can be helpful to predict what people will do and to understand why they do it.

This paper will focus on the essential role of systematic and reliable individual differences in influencing behavior. More specifically, the focus will be on individual differences in implicit associative evaluations and on under what conditions they can best predict behavior. In so doing, we will propose a theoretical model that specifies the path from perception to behavior, passing through the non-conscious activation of

individual differences in associative evaluations. The main message will be that the influence of non-conscious activation (i.e., priming) on (automatic) behavior is fundamentally moderated by individual differences in the valence of the implicit associative evaluation of the activated concept.

The ubiquitous influence of priming. The last two decades within social cognition have seen a dramatic surge of interest on nonconscious mechanisms leading to instances of so called automatic behavior, namely behavior that it is not performed deliberatively or intentionally (Dijkstheruis & Bargh, 2001; Greenwald & Banaji, 1995). From the original research of Higgins, Rholes, and Jones (1977), dozens of studies have shown that nonconscious activation of psychological concepts (e.g, traits, goals, stereotypes, cultural ideologies) can lead directly to behavior (for reviews, see Bargh, 2006; Higgins, 1996). This is typically achieved through subliminal or supraliminal priming that increases, usually unbeknownst to the person, the accessibility of a psychological construct that in turn influences behavior. The effects of priming have been shown for a variety of psychological constructs such as trait inferences (Winter and Uleman, 1984), stereotypes (Bargh, Chen, & Burrows, 1996) and goals (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Troetschel, 2001). The theoretical mechanism can be described as follows (Bargh & Chartrand, 1999). Behavior results as a consequence of relevant cognitive activities. These cognitive activities can be triggered by either internal (intentional) or external (automatic) sources. Priming is a means through which external sources trigger relevant cognitive activities that in turn activate behavior (cf. Figure 1a). The perception-behavior link is assumed to be highly automatized for most socially relevant situations, implying that the whole process can occur without any need for conscious awareness. From the

perspective of the influence of priming, this model implies that behavior will follow automatically in line with the primed concept. For example, if the prime concerns the concept of being rude, those who are exposed to the prime have an increased likelihood to behave rudely (cf. Bargh, Chen, & Burrows, 1996). This happens because priming makes temporarily more accessible the primed concept and initiates relevant cognitive activities that ultimately lead to action, without need for conscious or deliberative decision making.

--Insert Figure 1 about here--

Individual differences as the gatekeeper to action. The theoretical mechanism depicted in Figure 1a shows how action can be directly elicited from perception. However, it fails to consider the role of individual differences in this process. To illustrate, imagine that Anna is an introverted person whereas John is extraverted. If they are primed with an extraversion-relevant stimulus (e.g., chat) and afterwards they meet a stranger, Anna probably will be even more silent than normally whereas John will be even more outgoing than usual. For the first person, the stimulus will probably activate thoughts relating to the uneasiness of communicating with unknown persons and the unpleasantness of a superficial chat. For the second, however, the same stimulus will probably elicit thoughts concerning the pleasantness of meeting new persons and of chatting with them. These reactions can be systematically different across persons and relatively consistent within the same person. These individual preferences can be conceived in a general sense as a form of valence appraisals that rely upon associative evaluations. They do not require conscious awareness and deliberative decisions, but can directly lead to behavior in line with individual preferences. The theoretical mechanism is depicted in Figure 1b in a general form,

although in this contribution we will focus mostly on the automatic path to action. The key novelty of the proposed mechanism is that external sources (e.g., priming) do not lead directly to action congruent to the prime, but rather they activate fast evaluative associations that in turn lead to behavior. Ample empirical evidence supports the idea that evaluations can be very fast, can occur automatically, and occur very early in the chain towards action (Fazio, 2001), both for known (Bargh, Chaiken, Raymond, & Hymes, 1996) and novel (Duckworth, Bargh, Garcia, & Chaiken, 2002) stimuli. These automatic evaluations therefore critically determine whether behavior will be in line with or in the opposite direction to the primed concept. While part of these automatic evaluations can be a normative property of the activated concept and therefore affecting most people in a similar fashion, a major portion is related to relatively stable individual differences in the underlying associative valence.

Implications. This theoretical model has a number of implications. First, priming can increase or decrease the likelihood of performing a congruent action depending on the associated automatic valence appraisals. The same prime can therefore result in assimilation (i.e., behavior lined up with the primed concept) or contrast (i.e., behavior opposite to the primed concept). The contrast effect can result in two main ways: a) as a decrease of the likelihood of performing the congruent action or b) as an increase of the likelihood of performing an associated but opposite action. To exemplify, priming a hamburger can create a contrast effect that can either lead to a reduced likelihood in buying a hamburger or to an increase in the likelihood of buying a salad. Note that the contrast/assimilation distinction has been already studied in the context of priming but it has been linked to the awareness of the prime, with contrast effects likely to happen if there is clear awareness of the prime manipulation (e.g.,

Strack, Schwarz, Bless, Kubler, & Wanke, 1993). However, the gatekeeper model assumes that even when there is no awareness of the prime, still there can be a contrast effect as a function of individual differences in automatic evaluations. Second, whether there is a direct main effect of priming on behavior will depend also on the average valence of the primed concept/action. In other words, if for most people the associated thoughts initiated by priming are positive, a main assimilation effect will be more likely, whereas a main contrast effect will be more likely if the thoughts are mostly negative. Third, like for many other situational effects, the effects of priming will be moderated by individual differences in the relevant constructs. In this specific case, these will be individual differences in automatic associative evaluations. Therefore, to test the model, it is essential to have a corresponding reliable psychometric measure. As we will argue below, there are a number of such measures, most notably the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998). Fourth, a prime can be successful even without a main effect on behavior. In fact, the model implies that priming mainly activates individual differences in evaluative associations. This may or may not result in a main effect on action but it should be reflected in a significant correlation between such individual differences and behavior. In other words, the key test supporting the gatekeeper model is whether individual differences in associative evaluations moderate the effect of priming on behavior.

Individual differences in automatic evaluations. The last few years have seen the development and widespread adoption of a number of paradigms devoted to measuring individual differences in automatic evaluations, such as affective priming (Fazio, Sanbonmatsu, Powell, & Kardes, 1986), the Go/No Go Task (GNAT, Nosek

& Banaji, 2001), the Extrinsic Affective Simon Task (EAST, De Houwer, 2003), the Masked Affective Priming (MAP, Frings & Wentura, 2003), and the Affect Misattribution Procedure (AMP, Payne, Cheng, Govorun, & Stewart, 2005). However, the most widely used procedure to measure implicit attitudes is the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998). Several studies have shown that the IAT has good internal consistency (usually α=.80) and reasonable test-retest values (usually r=.60). The IAT has also the greatest evidence of construct and predictive validity (for recent reviews, see Nosek, Greenwald, & Banaji, in press; Poehlman, Uhlman, Greenwald, & Banaji, in press). Briefly, the IAT is a computerised method for indirectly measuring the strength of the association between a target concept and a valence attribute via a double-categorization task. It relies on the assumption that, if a target concept and an attribute dimension 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. The IAT needs one target category (e.g., Black), one contrast category (e.g., White), one target attribute (e.g., positive), and one contrast attribute (e.g., negative), each represented by a series of stimuli. In the critical combined task, stimuli from all four classes are presented and participants are asked to assign them correctly to one of the two combined category-attribute pairs (e.g., right key for Black-positive and left key for White-negative). This combined task is successively switched such that the pair category-attribute is the opposite. An IAT score is computed as the difference of the mean response times between the two versions of the combined task. From the perspective of the theoretical model that we propose, the IAT represents a reliable and valid means to tap individual differences in the automatic evaluative associations that are activated by priming.

Aims of the study.

This study aims at testing the model described above. Additional empirical evidence supporting this model is contained in Perugini, Conner, and O'Gorman (2006) and will be briefly reviewed later on in the final discussion paragraph. To recap the main mechanism, priming can be thought of as a means to increase the mental accessibility of the primed concept. One of the very first mental operations performed when a concept is salient is an automatic evaluation. This automatic evaluation is partly determined by relatively stable individual differences in the underlying evaluative associative structures. Behavior will follow either by assimilation or by contrast depending on the valence of this evaluation. Therefore, the crucial hypothesis is that automatic evaluations, as tapped by an implicit measure, will moderate the influence of priming on behavior.

To test this model, we focused on the stereotype towards Americans. The literature on stereotypes mostly focuses on minority or historically disadvantaged groups and therefore the focus has rarely been placed on Americans as a stereotype target. The most conspicuous exception is the classic study by Katz and Braly (1933) on national stereotypes, recently revisited by Madon et al. (2001). Both the original and the revisited study included Americans as a target. The most endorsed adjectives used to characterize Americans were with both positive (e.g., diverse, democratic, pleasure loving) and less clearly positive (e.g., materialistic, individualistic, lazy) valence. However, the studies did not provide overall measures of stereotype content but simply reported lists of adjectives that were endorsed to characterize Americans. Moreover, the samples were composed of American participants. Recently, a model of stereotype content has been proposed by Fiske, Cuddy, Glick, and Xu (2002). The

authors collected empirical evidence that two main dimensions, competence and warmth, can be usefully adopted to describe stereotypes associated with groups. These two dimensions are thought to act as organizers of evaluations – groups are placed on the continuum bi-dimensional space resulting from simultaneously considering warmth and competence. However, the study focused on specific subgroups (e.g., Feminists, Housewives, Elderly) and on racial minority groups (e.g., Jews, Hispanics, Blacks) rather than Americans as an overall group. In this study we will focus on this generic American stereotype and use Australians as a contrast category. There are two main reasons why we chose Australians. First, the IAT needs a contrast category and therefore we had to think about a group that could act as a contrast. Second, the study was performed in UK and, from an English perspective, both Americans and Australians share a common language and ancestry but differ in terms of stereotypes associated to them. Moreover, the subtle differences between American and British (adopted also by Australians) spelling in the written language was a nice feature that helped us to design the experiment, as described below. In the following contribution, we will report the main study, preceded by a short report of a preliminary study aimed at ascertaining the dimensionality of the stereotype content for both Americans and Australians.

Preliminary study

Method

Participants and Procedure

One hundred forty six undergraduate students, prospective students, and accompanying parents of a British university provided data for this preliminary study. Due to a mistake in printing, demographic data were not collected for 66 participants.

For the remaining 80 participants, there were 45 females and the mean age was 32.0 (SD 14.5). These values should be considered as roughly representative of the whole sample. The participants were contacted either individually on campus or as they participated in a visit day in the local Department of Psychology. They were handed the questionnaire and were debriefed afterwards about the purposes of the study.

Materials

The measure derived from Fiske et al. (2002) assessed stereotype content towards Americans and Australians. Specifically, the items: 'As viewed by society, how _____ are American (Australian) people?' was filled in by 12 adjectives, 6 referring to competence (competent, confident, capable, efficient, intelligent, skillful) and 6 to warmth (friendly, well-intentioned, trustworthy, warm, good-natured, sincere). Approximately half of the sample was presented with Australians first and half with Americans first.

Results and discussion

The means of all 12 adjectives were compared with a t-test for Americans and Australians. As depicted in Figure 2, Americans were systematically judged as possessing less of the qualities indicated by the adjectives, with the exception of "confident".

--Insert Figure 2 about here--

All paired t-test comparisons were statistically significant (from $p<.0005$ to $p=.025$) with the exception of "efficient" ($p=.119$). Overall, the American profile was significantly less positive than the Australian one ($t(145)=-5.106$, $p<.0005$, *M=2.89, SD=0.78 vs. M=3.17, SD=0.72*). Next, the factorial structures were ascertained. For both targets, a two-factor solution was not very convincing. For the Americans, the eigenvalues revealed a strong first factor (6.96) explaining 58.0% of the total variance

and a much weaker second factor (1.22; variance explained = 10.2%). An Oblimin rotated solution clarified that whereas one factor mirrored the competence dimension, the second factor had three strong markers of the warmth dimension (friendly, warm, good-natured) and the three remaining items that had sizeable loadings on both factors. The correlation between the factors was quite high (.55). On the contrary, the one-factor solution revealed strong loadings from all 12 items, ranging from .48 (confident) to .88 (skilful). These results were almost identical for the Australian target, with a first factor explaining 61.2% of variance and the second 9.1%, the two factors highly correlated (.66), the same three clear markers for the factor warmth, and strong loadings for all items (from .50, confident, to .85, trustworthy) in a single factor solution.

To sum up, this preliminary study has indicated that the bi-dimensional structure proposed by Fiske et al. (2002) of competence vs. warmth is not well supported by the data. Only one factor is well defined by its markers whereas the second is only partly defined. Moreover, the two factors are highly correlated, in the region of .60. Finally, a single dimension of positivity seems by far the most important factor, explaining around 60% of the total variance. Therefore, also for the sake of simplification, we decided to consider in the main study a single dimension of positivity and selected three markers each from competence and warmth so to have a balanced representation of the positivity dimension.

Main study

Method

Participants

Ninety-two British students (mean age = 25.4, SD=9.4; 53.3% males)[1] at a British university participated in this experiment. Two participants were excluded

from the analysis. One participant was excluded due to a high percentage of quick responses (32.5% of response latencies were under 300ms) and errors (36.6%), indicating a random response pattern. A second participant was excluded because the dependent variable (essay quality, see below) was more than 3 SD below the mean of the group and about 1.4 SD away from the nearest data point[2].

Design and Procedure

All participants were tested individually at a cubicle. They first completed a measure of relative implicit preference for Americans vs. Australians before responding to items assessing their explicit attitudes towards these two groups (first related to Americans and then to Australians). Following this, participants were presented with a scrambled sentence task (Srull & Wyer, 1979). The task presented 20 scrambled sentences comprising five words each from which participants had to construct sentences using four of the words. Participants were randomly assigned to one of two versions of this task. One version used typically Americans words (e.g., subway, elevator, dork, gasoline, cheerleaders, skyscrapers, liqor, thanksgiving) in 15 out of the 20 sentences (priming condition) and a second version contained all neutral sentences (control condition). Immediately after the priming task, participants read and rated the qualities of the essay. This task was designed so that participants would have an opportunity to immediately use the recently activated American stereotype. The essay was one page A4 long and, while discussing pros and cons of the introduction of the Euro in the UK, it finally advocated that the Euro should not be introduced. It was pre-selected in a pilot study with a different sample (n=40) where it was rated on a composite of 2 indicators for importance (importance, relevance) and 3 for qualities (convincing, based on facts, effective). We wanted to select an essay with

moderate importance, so to engage the participants, and good but not excellent quality, so to leave room for individual differences in evaluation. The ratings were on a scale from 1 to 5. Participants judged the essay as moderately important (3.4) and with good qualities (3.65), therefore meeting our requirements. To ensure that the stereotype could be applicable to the essay, two features were incorporated. First, the essay included a paragraph containing a parallel with the North American Free Trade Agreement (NAFTA) agreement among US, Canada, and Mexico. Second, the essay incorporated American spelling. In practice, this meant that words such as analyze, centered, and colors were used in place of their British and Australian equivalents (analyse, centred, colours). There were six such words in the whole essay. Finally participants were probed for suspicion about the nature of the study and debriefed afterwards. No participant expressed suspicion about the nature of the study.

Measures

Implicit attitudes towards Americans relative to Australians were assessed using the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). The IAT followed the standard 5-step format with the first, second and fourth steps being practice and the third and fifth steps including practice and critical test phases. Steps 1, 2 and 4 involved sorting stimuli from two categories to one of two responses (step 1- American targets to the left response button and Australian targets to the right response button; step 2- positive attributes to the left response and negative attributes to the right response; step 4- American stimuli to the right response button and Australian stimuli to left response button). In step 3, participants were required to arrange stimuli from four categories to one of two responses (American and positive stimuli to the left response key and Australian and negative stimuli to the right

response key). In step 5, the pairings from step 3 were reversed, such that the attribute categories remained assigned to the same response key but the target categories were switched. The target categories were American (Buffalo, American, Baseball, Chicago, Yank) and Australian (Kangaroo, Australian, Cricket, Sydney, Aussie) whereas the attribute categories were Positive (Love, Gift, Joy, Rainbow, Pleasure) and Negative (Evil, Cancer, Death, Agony, Vomit). There were 20 practice trials in steps 1, 2 and 4. Steps 3 and 5 contained 20 practice trials followed by 60 test trials (plus two dummy initial trials that were discarded) for the critical steps 3 and 5. The IAT was counterbalanced between subjects for steps 3 and 5.

IAT scores were calculated using the recently proposed D algorithm (Greenwald, Nosek, & Banaji, 2003) and weighting practice and test trials according to their proportion in the overall score. The IAT score was coded such that a positive score reflected implicit preference for Americans over Australians and it was reliable (α =.78).

Explicit attitudes (stereotypes) towards Americans and Australians were measured with a slightly modified and reduced version of the instrument used in the preliminary study. Specifically, the stem was modified in the first person singular and read 'In your opinion, how _____ are American (Australian) people?. The stem was personalized because we wanted to gather their individual opinion. It was followed by six adjectives, three related to competence (competent, capable, efficient) and three to warmth (friendly, well-intentioned, warm) rated on a five-point scale (1=not at all, 5=extremely). The mean average score was calculated by subtracting the evaluation for Australians from the ones of Americans so to generate a relative measure similar to the implicit attitude. Based on the preliminary study, we aimed at an overall

composite measure of stereotype positivity[3]. The measure yielded a satisfactory internal reliability (α=.73).

Five items, along 10-point bipolar scales (1=not at all – 10=extremely), assessed *essay quality*: 'How capable of changing people's opinions is this essay?', 'How trustworthy do you think the writer is?', 'How intelligent do you think the essay is?', 'How well-intentioned do you think the writer is?' and 'How sincere do you believe the essay to be?' A PCA indicated a single factor structure (eigenvalue=2.89 accounting for 57.8% variance) with all items satisfactorily loading onto the factor (all loadings >.63). The measure was internally reliable (α=.81).

Results and Discussion

Participants displayed no implicit preference for Americans or Australians (M=.02, SD=.25; $t(89)=0.92$, $p=.36$. However, participants' explicit attitudes were significantly less positive towards Americans than Australians, $t(89)=-3.72$, $p<.0005$ (M=3.34, SD=0.59 vs. M=3.62, SD=0.54), therefore confirming the results obtained in the preliminary study. The correlations between the variables are displayed in Table 1.

--Insert Table 1 about here--

Priming condition was unrelated to any of the measures (all $p>.31$) suggesting that the participants in the American and neutral prime conditions were equivalent across the predictor variables. Furthermore, priming did not have a direct effect on essay quality, $r=-.11$, $p=.30$. This implies that the essay was rated equally in both conditions (M=6.77, priming, vs. M=6.99, control). Only explicit attitudes, $r=.23$, $p=.03$, was significantly related to essay quality. Implicit attitudes were unrelated to the explicit ones ($r=14$, $p=.19$). To assess how the interrelationships between the

study variables varied across groups, Pearson's correlations were calculated separately for each condition (see Table 2).

--Insert Table 2 about here--

When the statistics were calculated separately for each condition, implicit attitudes were the sole significant predictor of the essay quality but only within the priming condition, r=.31, p=.04. In the neutral condition, implicit attitudes were not related to essay judgment r=-.20, p=.19. To formally test the key hypothesis, a hierarchical regression analysis was conducted in which the standardized implicit attitude measure and priming manipulation were entered on the first step before their interaction on the second step. A significant interaction would imply that the effect of priming on behavior is dependent on individual differences in one's automatic evaluation concerning the target concept (Americans).

On the first step, the model was non-significant, $F(2, 87)=0.73$, p=.48, with neither priming condition, β=-.21, p=33, nor implicit attitudes, β=.07, p=.51, predicting essay quality. However, on the second step the model was almost significant, $F(3, 86)=2.54$, p=.06 and, crucially, there was a significant interaction β=.51, p=.02, indicating that the effect of priming on essay quality was moderated by one's implicit attitudes.

Once the key effect was established, ancillary analyses were conducted to ascertain its uniqueness. First, we checked whether priming had a similar effect on explicit attitudes. They were predictive of essay quality (β=.22, p=.03) but, crucially, this effect was not qualified by a significant interaction with priming (β=-.04, p=.86). Note that this result not only implies that the priming manipulation activated exclusively implicit attitudes, but also that the essay effectively elicited the application of stereotypic opinions, as evidenced by the main effect of explicit

attitudes across conditions. Second, we checked whether the effect of priming on implicit attitudes still remained when including explicit attitudes in the regression equation. This was the case. Explicit attitude predicted significantly essay quality (β=.20, p=.05) and the interaction between IAT and priming remained significant (β=.52, p=.01).

General Discussion

The results confirmed the key hypothesis and supported the proposed theoretical model. The effects of priming on behavior were moderated by individual differences in valence automatically associated to the primed concept. These results mirror those obtained by Perugini et al.(2006) in two studies on spontaneous helping. The first study adopted an IAT attitudinal measure of Altruism and included a subliminal priming manipulation embedded in a lexical decision task whereby masked primes relative to altruism were presented for 42 ms. The dependent variable was the number of video-clips participants volunteered to watch and rate for an unrelated fictitious experiment (spontaneous helping). Similar to the study presented here, the IAT predicted spontaneous helping but only under condition of priming, a germane explicit attitudinal measure was not affected by priming, and there was no main effect of priming on behavior. The second study was identical to the first except that this time the measures were taken one week before the priming manipulation and the dependent variable spontaneous helping. The results were qualitatively identical. The present results add significantly to the empirical evidence in favor of the gatekeeper model by adopting a different priming manipulation (supraliminal), focusing on a different domain (stereotype about Americans) and using a different form of behavior (essay evaluation).

An important result of all these studies is that the effects were exclusive to the IAT and they were not affected when explicit attitudes were also considered. It is useful to embed these results within current theoretical models of decision-making. Strack and Deutsch (2004) recently proposed their reflective-impulsive model that distinguishes between a fast, association-based system (impulsive) and a slow, deliberative-based system (reflective). In terms of measures, an implicit measure like the IAT taps into the associative network whereas an explicit measure impinges upon the reflective system. The results obtained in this study, as well as those concerning spontaneous helping, seem to support the idea that it is possible to selectively activate only one system independently of the other. Specifically, priming had an activating influence only on evaluative individual differences as reflected in implicit but not in the explicit measures that, however, predicted the evaluation of the essay independently of the effects of priming. It seems more reasonable to explain our results by assuming a dual system of attitudes (Wilson, Lindsey, & Schooler, 2000) rather than a single attitude-two measures system (Fazio & Olson, 2003), although probably the two perspectives might be stretched to accommodate most empirical findings.

It should be emphasized that the results can also be read from the perspective of what conditions are needed to increase the predictive validity of implicit measures. This is an essential issue for establishing the theoretical relevance and pragmatic utility of an implicit measure (cf. Perugini, 2005). In fact, the results suggest that the IAT predicts relevant behavior best if the concept associated to the behavior is activated. It is interesting to incorporate this finding with some recent literature. Hofmann, Rauch, and Gawronski (in press) have shown that control resources moderate the predictive validity of the IAT. Specifically, when control resources are low or depleted, an implicit measure predicts relevant behavior more accurately. Therefore, it would seem

that it is possible to increase the predictive validity of an implicit measure both by increasing the availability of the relevant concept in the associative network (impulsive system) and by decreasing the control resources that are needed for deliberative processes (reflective system). In a related vein, Perugini, O'Gorman and Prestwich (in press) have shown that the predictive validity of an implicit measure can be enhanced by activating the self before performing the IAT measure, although it depends on what is exactly activated in the self. In this latter case the emphasis is on conditions that improve the validity of the measure as such. However, in all cases the results start to identify theoretically meaningful conditions that enhance the predictive validity of an implicit measure.

Conclusions

It is worthwhile to highlight that while the study presented here focused on a specific implicit measure of attitudes, the gatekeeper model is more general and does not depend on the application of specific measures. The gist is that perception activates (automatic) cognitive activities that in turn lead to behaviors. However, unlike mainstream social cognitive models, it is maintained that individual differences in these cognitive activities critically influence what kind of behavior will be performed. Among these cognitive activities, automatic evaluations have a prominent role and they occur early in the chain to action. Therefore, individual differences in automatic evaluations have a key role in determining action: they are the gatekeeper. Future research will be needed to replicate the results and to establish precisely the conditions under which these effects are present.

A final thought concerns a certain (admittedly loose) parallel between the gatekeeper model presented herein and the relationships between social psychology and

personality in the 60's and 70's. The critical book of Mischel (1968) paved the way for a ferocious attack on personality research and on its capability to actually explain or even predict behavior and advocated the normative influence of situations. Later research clarified that stable individual differences in relevant personality dimensions, instead, have a central role either directly or in interaction with situational factors for predicting behavior (e.g., Kenrick & Funder, 1988). Today the emphasis of mainstream social cognition is on the automatic (and normative) link between perception and action. The gatekeeper model seeks to establish that this link is fundamentally affected by relatively stable individual differences in automatic evaluations. In its strong version, the gatekeeper model may suggest that the direct perception-action link is an epiphenomenon because the key theoretical mechanism resides in the activation of automatic evaluations that systematically reflect individual differences in associative evaluations. Therefore, from this extreme perspective, there is no such a thing as a direct automatic perception-action path. A weaker version of the model instead would maintain that individual differences fundamentally moderate whatever normative effect is associated with the direct link perception-action. While the first view would suggest that main effects of priming on behavior are more the exception than the rule (the rule being an interaction effect with individual differences in automatic evaluations) and can be found mostly in some specifiable conditions, this second weaker view would suggest that alongside a main effect there will often be a moderation effect of individual differences. Future research will clarify which one of the two versions fits best with available empirical evidence. Either way, the role of individual differences is assumed to be key in the automatic direct link between perception and action. Therefore, the gatekeeper model can be seen as an attempt to establish a foundational role for individual differences, although in the 2000's the

focus is on automatic processes rather than on situations. Blending two famous quotes from ancient philosophers, history repeats itself (Vico) but you can’t step twice in the same river (Heraclitus).

References

Bargh, J.A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology, 36,* 147-168.

Bargh, J. A., Chaiken, S., Raymond, P., & Hymes, C. (1996). The automatic evaluation effect: Unconditional automatic evaluation activation with a pronunciation task. *Journal of Experimental Social Psychology, 32,* 104–128.

Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71,* 230-244.

Bargh, J.A., Gollwitzer, P.M., Lee-Chai, A., Bardollar, K., & Trotschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81,* 1004-1027.

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

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology, 33,* 1–40.

Duckworth, K.L., Bargh, J.A., Garcia, M., & Chaiken, S. (2002). The automatic evaluation of novel stimuli. *Psychological Science, 6,* 515-519.

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

Fazio, R.H., & Olson, M.A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54,* 297-327.

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, 50,* 229-238.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from status and competition. *Journal of Personality and Social Psychology, 82,* 878-902.

Frings, C.G., & Wentura, D. (2003). Who is watching *Big Brother*? TV consumption predicted by masked affective priming. *European Journal of Social Psychology*, *33,* 779-791.

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. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–1480.

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins, & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). New York: Guilford.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141–154.

Hofmann, W., Rauch, W., & Gawronski, B. (in press). And deplete us not into temptation: Automatic attitudes, dietary restraint, and self-regulatory resources as determinants of eating behavior. *Journal of Experimental Social Psychology*.

Katz, D., & Braly, K.W. (1933). Racial stereotypes of one-hundred college students. *Journal of Abnormal and Social Psychology*, *28*, 280-290.

Kenrick, D.T., & Funder, D.C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist*, 43, 23-34.

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., & Jussim, J. (2001). Ethnic and national stereotypes: The Princeton trilogy revisited and revised. *Personality and Social Psychology Bulletin*, *27*, 996-1010.

Mischel, W. (1968). *Personality and assessment*. New York: Wiley.

Nosek B.A., & Banaji, M.R. (2001). The go/no-go association task. *Social Cognition*, *19*, 625-666.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (in press). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Automatic processes in social thinking and behavior*. Psychology Press.

Payne, B.K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology*, *89*, 277-293.

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

Perugini, M., Conner, M., O'Gorman (2006). *The interplay between implicit and explicit attitudes: The case of spontaneous helping*. Unpublished manuscript.

Perugini, M., O'Gorman, R., Prestwich, A. (in press). An ontological test of the IAT: Self-activation can increase predictive validity. *Experimental Psychology*.

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

Strack, F., Schwarz, N., Bless, H., Kuebler, A., & Wanke, M. (1993). Awareness of the influence as a determinant of assimilation versus contrast. *European Journal of Social Psychology, 23,* 53-62

Wilson, T. D., Lindsey, S., & Schooler, T. (2000). A model of dual attitudes. *Psychological Review, 107*, 101-126.

Winter, L. & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47*, 237-252.

FOOTNOTES

[1] Unfortunately a programming mistake meant that gender and age were not recorded for the first 32 participants, after which the mistake was corrected. The descriptive statistics, although relative to this sub-sample of 60 participants, should be considered as approximately reflecting the whole sample.

[2] The results are not affected significantly by including this participant. All the effects that we have discussed in the study remain statistically significant.

[3] We ran the analyses also considering separately a dimension of competence and warmth. The main conclusions are not affected.

Table 1: Correlations between study variables

a) full-sample, n=90

	1	2	3	4
1. Priming Condition (0=Neutral, 1=American)	-	-.06	-.07	-.11
2. IAT score		-	.14	.08
3. Explicit Attitude			-	.23*
4. Essay Quality				-

b) by condition (priming: n=45, neutral: n=45).

	1	2	3
1. IAT score			
Priming	-	.24	.31*
Neutral	-	.05	-.20
2. Explicit Attitude			
Priming		-	.18
Neutral		-	.28
3. Essay Quality			
Priming			-
Neutral			-

Figure 1. Mainstream social cognitive model (a) and the gatekeeper model (b)

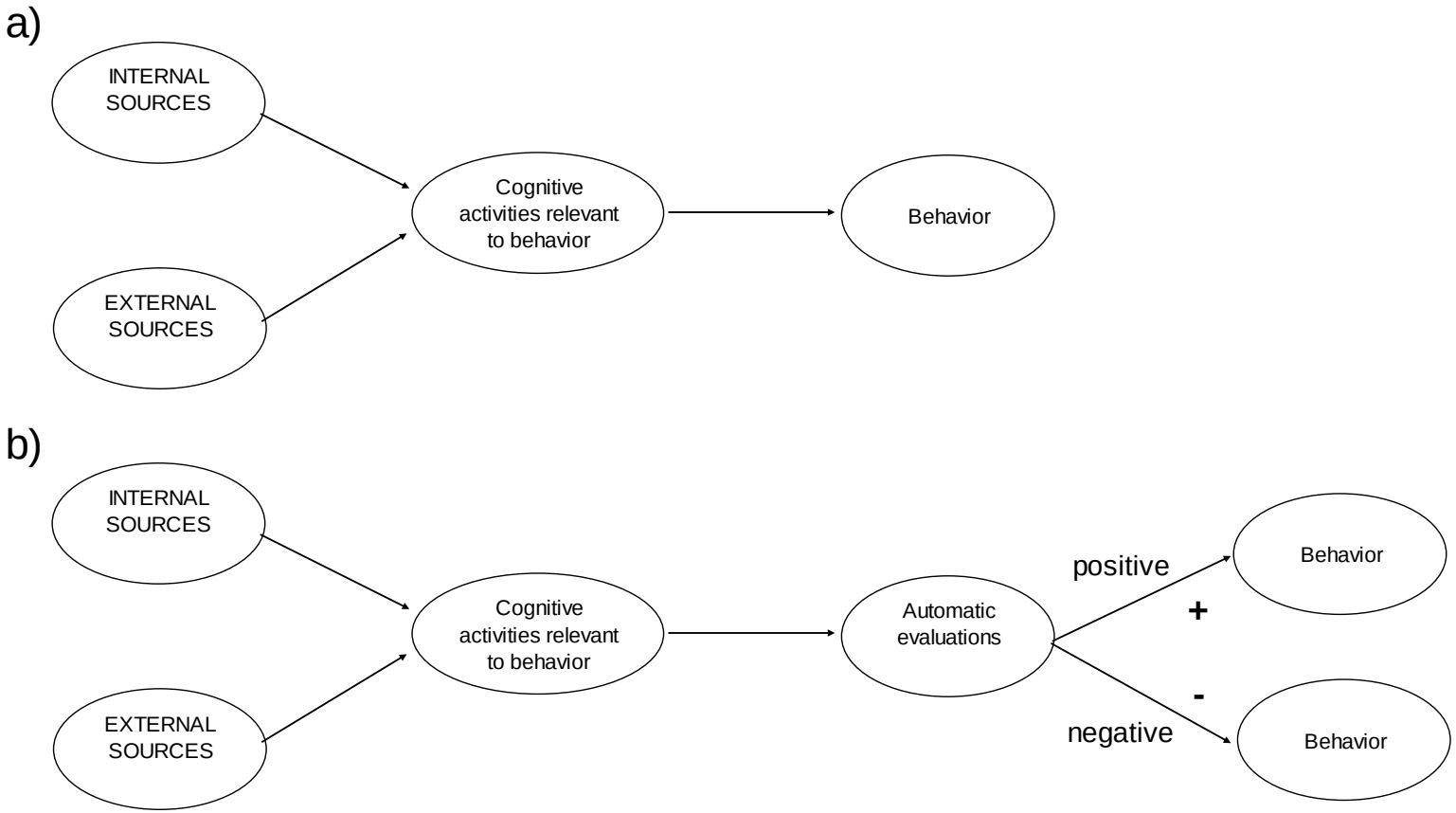

Figure 2. Stereotype content for Americans and Australians (N=146)