

Moderators of the Relationship between Implicit and Explicit Evaluation

Brian A. Nosek

University of Virginia

Correspondence:

Brian Nosek
Department of Psychology
Box 400400; 102 Gilmer Hall
Charlottesville, VA 22904-4400
Em: nosek@virginia.edu

Fax: 434-982-4766
Phone: 434-924-0666

Abstract

Automatic and controlled modes of evaluation sometimes provide conflicting reports of the quality of social objects. This paper presents evidence for four moderators of the relationship between automatic (implicit) and controlled (explicit) evaluations. Implicit and explicit preferences were measured for a variety of object pairs using a large sample. The average correlation was $r = .36$, and 52 of the 57 object pairs showed a significant positive correlation. Results of multilevel modeling analyses suggested that: (a) implicit and explicit preferences are related, (b) the relationship varies as a function of the objects assessed, and (c) at least four variables moderate the relationship – self-presentation, evaluative strength, dimensionality, and distinctiveness. The variables moderated implicit-explicit correspondence across individuals and accounted for much of the observed variation across content domains. The resulting model of the relationship between automatic and controlled evaluative processes is grounded in personal experience with the targets of evaluation.

Keywords: Implicit Attitudes, Explicit Attitudes, Moderation, Implicit-Explicit Correspondence, Dual-Process Models

A persistent question for psychologists concerns how conscious experience corresponds with the content and processes of the mind. It is clear that conscious experience is not a direct reflection of mental operations and a variety of mental activities are unavailable to introspection (Nisbett & Wilson, 1977). This observation has been extended to attitudes, the association between a concept and an evaluation (Fazio, Chen, McDonel, & Sherman, 1982), by making a theoretical distinction between implicit and explicit evaluations. The former reflects evaluative information activated automatically and perhaps without intention or awareness; and the latter reflects evaluations produced by controlled processes (Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000).

Measurement innovation has paralleled this theoretical innovation with the development of indirect means for assessing concept-evaluation associations. Results from indirect measures have surprised psychologists and respondents by often revealing evaluations that contradict those assessed directly. In comparison to explicit measures, implicit measures assess evaluation in constrained conditions that include one or more hallmarks of automaticity – unawareness, uncontrollability, lack of intention, or efficiency of processing (Bargh, 1994). Comparisons of implicit and explicit measures suggest that they tap related but distinct constructs (Cunningham, Preacher, & Banaji, 2001; Greenwald & Farnham, 2000; Nosek & Smyth, 2005) and that both modes of evaluation may have important implications for social perception, judgment, and action (Banaji, 2001; Chen & Bargh, 1999; Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald & Banaji, 1995; Greenwald & Nosek, 2001; Poehlman, Uhlmann, Greenwald, & Banaji, 2004). The purpose of this paper is to identify moderators of the relationship between implicit and explicit evaluation.

Existing Data and Theory Concerning the Relationship between Implicit and Explicit Evaluations

In 1986, Fazio and his colleagues demonstrated that associations between concepts and evaluations could be measured automatically and without introspection (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). Their evaluative priming procedure assessed the association between a concept and an evaluation indirectly by measuring the time required to evaluate words with good or bad meaning (e.g.,

wonderful, terrible) immediately following a brief presentation of a social object (e.g., a Black face or a White face). Social objects associated with positive evaluation facilitated responses (i.e., faster response times) to words with positive meaning, whereas objects associated with negative evaluation facilitated responses to words with negative meanings. This and related research efforts revealed that evaluative associations can be elicited automatically and are sometimes at odds with respondents' professed attitudes (Dovidio et al., 1997; Fazio et al., 1995; Greenwald, McGhee, & Schwartz, 1998). This suggests that a single individual might employ multiple processes for evaluating information in the environment: some that are conscious, intentional, and controlled and others that are unconscious, unintentional, and automatic (Banaji, 2001).

Theoretical accounts offer at least two interpretations of the general relationship between implicit and explicit measures. One perspective considers them to be assessing distinct constructs (Devine, 1989; Dovidio et al., 1997; Greenwald & Banaji, 1995; Wilson, et al., 2000). These perspectives do not explicitly disallow a relationship to exist between implicit and explicit evaluation, but most have steered clear of providing specific predictions about when they will be related or distinct.

A second perspective hypothesizes a single attitude construct and suggests that the implicit-explicit distinction more meaningfully refers to the method of measurement rather than to dissociable constructs (Fazio et al., 1995; Fazio & Olson, 2003). The activation of an evaluation is presumed to follow a single processing stream and implicit measures tap the evaluation before conscious control processes can be initiated, and explicit measures tap the evaluation after intentional processes have had opportunity to alter the response. From this perspective, implicit and explicit evaluations differ only to the extent that explicit responses are altered through controlled processes (Fazio & Olson, 2003). So, the relationship between implicit and explicit evaluation depends on when people will be motivated to alter their explicit responses and have the opportunity to do so.

Variable relations between implicit and explicit evaluations. Much of the early empirical evidence comparing implicit and explicit measures suggested that they elicit distinct, even conflicting, evaluative responses (e.g., Banaji & Hardin, 1996; Blair & Banaji, 1996; Bosson, Swann, & Pennebaker,

2000; Devine, 1989; Fazio et al., 1995; Greenwald, et al., 1998). More recent evidence reveals that, for some content domains, implicit and explicit evaluations are related (Greenwald & Nosek, 2001; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2004; Nosek, Banaji, & Greenwald, 2002; Nosek & Banaji, 2002; Nosek & Smyth, 2005).

The fact that the strength of the implicit-explicit relationship varies across social objects suggests that one or more factors moderate the relationship. To date, researchers have highlighted two classes of moderators: self-presentation concern, and factors relating to evaluative strength such as attitude importance and elaboration (Fazio, 1995; Karpinski, Steinman, & Hilton, 2005; Nosek & Banaji, 2002; Schuette & Fazio, 1995). The present study examines these factors and introduces two additional moderators – dimensionality (the extent to which target objects conform to a simple, bipolar structure) and distinctiveness (the extent to which one’s evaluation is perceived to differ from cultural norms).

Proposed Moderators of the Relationship between Implicit and Explicit Evaluations

Correspondence between implicit and explicit evaluations may be determined by multiple factors. Selection of candidate moderators for the present study was based on research evidence and an effort to represent both the intrapersonal and interpersonal experience of evaluations. That is, people have an internal experience of liking or disliking social objects and simultaneously experience that evaluation in a social context where it may be compared to the evaluations of others, or potentially adjusted based on concerns about social consequences. The four proposed moderators of implicit-explicit correspondence reflect qualities of motivation and concern about self-reports (self-presentation), the potency and importance of the evaluation (strength), the structural qualities of the evaluation (dimensionality), and the perception of one’s evaluation as unique in comparison to others (distinctiveness). Two of these primarily emphasize intrapersonal qualities of evaluation (strength and dimensionality) and the other two emphasize interpersonal qualities of evaluation (self-presentation and distinctiveness).

While causal relations will not be identifiable in this correlational study, the candidate moderators vary in terms of how they are hypothesized to operate on implicit-explicit relations. That is, controlled processes may influence the development of automatic evaluations, automatic processes may influence

controlled reports, or automatic and controlled evaluations may be similarly influenced by some factors and differentially influenced by others. Next, the rationale for each of the moderators is summarized.

Self-presentation. A commonly hypothesized moderator of implicit-explicit correspondence is self-presentation – the altering of a response for personal or social purposes. Greenwald and Banaji (1995) proposed that implicit preferences are ones that participants are unwilling or unable to report. Self-presentation may capture the unwilling aspect of this definition. People might be unwilling to report an evaluative response that comes to mind because (a) they do not want others to know about it, or (b) the feeling is unwanted in the sense that it is not endorsed or accepted as one's evaluation. The important distinction between (a) and (b) is that self-presentation can be socially deceptive (e.g., a daughter may report liking dad's horrible meatloaf to avoid hurting his feelings), or it can be genuine (e.g., a White store owner may consciously reject her negative reaction to a Black patron as inappropriate and unfair and attempt to replace it with a more positive explicit response).

Self-presentation is often considered as a trait ignoring the particular context, or social objects, of interest (Paulhus, 1984, 1991). Here self-presentation concern is represented in terms of responses to the content domain in particular, rather than characterizing it as a general, content-free personality characteristic. Also, self-presentation has both interpersonal (impression management) and intrapersonal (deception) components, though the former are usually emphasized.

Implicit measures tend to be resistant to deliberate alteration, whereas explicit measures are vulnerable to it (Fornoni & Mayr, 2005; Kim, 2003). When self-presentation concern is high, explicit responses may be altered but implicit responses may not. Therefore, self-presentation should negatively affect the strength of implicit-explicit correspondence. Implicit-explicit correspondence should be higher when self-presentation concerns are weak and lower when self-presentation concerns are strong. This moderator suggests an operative mode in which self-presentation concern causes explicit responses to move away from unadulterated automatic responses (Fazio et al., 1995).

Evaluative strength. Strong evaluations are more stable and consequential than weak ones (Krosnick & Petty, 1995; Petty & Krosnick, 1995). Krosnick and Petty (1995) identified four

distinguishing features of strong evaluations: they are persistent, resistant to change, guide information processing, and strongly predict behavior. Also, there are a variety of features of strength including: importance, elaboration, knowledge or familiarity, frequency of thought, stability, extremity, and ambivalence (Bassili, 1996; Krosnick, Boninger, Chuang, Berent, & Carnot, 1993; Raden, 1985).

A seminal paper by Gordon Logan (1988) developed an instance theory of automatization and suggested that the practicing or repetition of a response increases the likelihood that it will be automatically activated in the future. Practice is more likely to occur for strong evaluations, and, as a consequence, may increase the likelihood that they are automatized. Thus, implicit and explicit evaluations were predicted to be more consistent for strong evaluations than for weak ones. Evaluations that are personally important, highly familiar, frequently thought about, stable, extreme, and unambivalent should elicit stronger implicit-explicit correspondence than those that are unimportant, unfamiliar, infrequently thought about, unstable, insignificant, and ambivalent.

This suggests a mode of operation in which the controlled response shapes the automatic one over time. An alternative causal relation is also conceivable. Fazio, for example, suggested that greater accessibility increases the likelihood that an evaluation will be automatically activated and then influence the explicit response (Fazio, 1995). This perspective emphasizes the relation between automatic and controlled processes in the immediate context. These two perspectives are not mutually exclusive. This study will only identify whether moderation by strength occurs, not the nature of its causal influence.

There is disagreement about whether the various indices of evaluative strength represent one or a few underlying constructs (Abelson, 1988; Verplanken, 1991) or whether they are unique (see Krosnick et al., 1993 for a review). The data suggest that the indices of strength have shared and unique components (Bizer & Krosnick, 2001). At this stage, there is little theory to suggest which of the variety of strength measures will best moderate implicit-explicit correspondence. So, for the purposes of establishing an initial model, a composite measure of evaluative strength measures was examined, followed by an exploratory analysis of the individual factors.

Dimensionality. Judd and Kulik (1980) observed that some evaluations may be mentally conceived as a bipolar continuum defined on one end by information supported and on the other end by information rejected. Pratkanis (1989) extended this idea noting that attitudes can have a bipolar or unipolar structure. Being for gun control usually implies being against gun rights suggesting a negative correlation between the concepts and a bipolar structure. On the other hand, for many people, being pro-men does not imply being anti-women, suggesting that gender attitudes have a more unipolar structure.

Judd and Kulik (1980) hypothesized that attitudes promote selective information processing in a bipolar manner by facilitating encoding, retention and retrieval of both consistent and inconsistent information. They suggested that “judgments of attitude positions should be made more easily, and hence more quickly, if those attitude positions ‘fit’ the bipolar attitude schema” (p. 571). Judd and Kulik’s quote points to the efficiency component of automaticity (Bargh, 1994) suggesting that bipolar evaluations may be automatically activated more readily and consistently than evaluations not conforming to that simple structure. Judd and Kulik supported their hypothesis by showing that bipolar evaluative information was judged more quickly and was much better recalled than information that did not fit the bipolar structure.

This suggests that a bipolar structure may have cognitive benefits that simplify or organize attitudinal information for more efficient and consistent processing. That is, if bipolar evaluations have a simple structure, then evaluations may be activated more consistently across context. Further, a bipolar structure suggests that items assessing the poles are negatively related, so taking the difference of assessments of each pole should increase the reliability of the overall evaluation in the same way that reliability is increased by averaging positively correlated attitude assessments.¹ Conversely, the lack of simple structure for unipolar attitudes may result in a more complex, less stable and more difficult to retrieve mental representation of the evaluation. As such, automatic evaluations for unipolar attitudes may be less reliable or more malleable to situational influence. Hence, evaluations that have more bipolar than unipolar qualities are hypothesized to elicit greater correspondence among automatic and controlled evaluative processes.²

Distinctiveness. In an individual differences approach, high implicit-explicit correspondence indicates that self-reported evaluations are rank-ordered across individuals similarly to automatically-assessed evaluations. That is, high implicit-explicit correspondence suggests that introspection is similar to an indirectly revealed association between the objects and evaluations. What process is used to translate an internally experienced evaluation into a rating on a scale? And, what might facilitate the proper calibration of that introspective act to correspond with the indirectly assessed strength of association?

Two sources of information may contribute to a self-assessment that will influence rank ordering of self-reports. One calibrating factor can occur via an internal comparison with evaluations of other objects. For example, a respondent could respond '5' to ice cream on a liking scale of 1 (hate) – 6 (love) because of a sense that she likes ice cream, but that there are other things that she likes more that would earn a higher rating. A second calibrating factor is to compare the magnitude of one's own evaluation with perceptions of others' evaluations for the same domain. For example, the same respondent might respond '5' to ice cream in part because she recognizes that while she likes ice cream, there are others that like ice cream even more and so the more extreme value would reflect their evaluations.

The first calibrating factor is intrapersonal, the second is interpersonal, and both could moderate implicit-explicit correspondence. Also, the calibrating factors need not be either/or influences on self-reports, both likely operate simultaneously in the translation between an internal experience and an explicit rating. The intrapersonal comparative assessment of evaluations corresponds with the evaluative strength moderator discussed earlier. Part of the interpersonal comparative assessment of evaluations is the focus of this fourth proposed moderator - distinctiveness.

In this paper, distinctiveness is defined as the magnitude of the discrepancy between one's own evaluations and the perceived norm. Higher values indicate greater perceived discrepancy from the norm, suggesting a perception that others' feel differently. A distinctive evaluation should be seen as more unique, like a personal possession (Abelson & Prentice, 1989), whereas non-distinct preferences may

provide little sense that the evaluation is anything more than “what everybody thinks.” Said another way, distinctiveness reflects an interpersonal notion that one’s evaluation is different than the norm.

Unlike the intrapersonally-focused strength variable, evaluations could be distinctive because they are much stronger or much weaker than the perceived cultural norm. Highly distinctive evaluations may be defining features of the self-concept (Blanton & Christie, 2003). A Bostonian may consider his indifference to the Boston Red Sox as self-definitional amid the rabid Sox fans he encounters daily. The distinctly indifferent evaluation in the social context may become self-definitional and an important predictor of the person’s behavior distinguishing him from the nearly universal norms – not going to the bar for the World Series, not talking about it with friends – whereas the casual fan may not have a strong evaluation, but may see herself as normatively indistinct and follow along with the rush of enthusiasm as the norm demands. Interpersonally, the self is defined as much by that which is distinctly cared about as that which distinctly is not cared about.

Distinctiveness examines the notion that personal evaluations can be identified not just in terms of the internal experience, but in relation to a social context. Operatively, this suggests two means by which implicit-explicit correspondence might be enhanced when evaluations are distinct. First, distinct evaluations, even if weakly held, are hypothesized to be more closely linked to the self as a contrast from others, and should allow for more efficient and precise self-observation. This suggests that distinctiveness is an interpersonal complement to the more intrapersonally derived evaluative strength. Second, distinctiveness emphasizes the uniqueness of personal evaluations compared to norms, so accurate self-report of relative standing compared to others may be enhanced when it is clear that one’s own evaluation contrasts from others.

To summarize, accurately reporting an evaluation depends on two things: knowing one’s feelings, and knowing how to accurately characterize one’s feelings on a scale. For correlations comparing across individuals, this latter point depends on an interpersonal factor: people who select more extreme values on a scale should have stronger evaluations than people selecting less extreme values. With implicit measures, there is no decision or inference process – relative standing is a consequence of task

performance. Self-report, however, is critically dependent on judgment processes to translate one's internal feeling into a value on a scale. The more that is known about one's own feelings, and the feelings of others, the more likely that self-report ratings will accurately reflect internal experience rather than response biases or random error. To the extent that those internal experiences correspond to automatic evaluations, the implicit-explicit relationship should then be enhanced.³ Distinct evaluations should be more easily identified in relative standing compared to others than evaluations that are not distinct. Therefore, stronger implicit-explicit correlations between persons should emerge for more distinct evaluations.

The hypothesis that implicit measures are confounded by extra-personal associations, such as cultural knowledge, would lead to exactly the opposite prediction about the relationship between distinctiveness and implicit-explicit correspondence (Arkes & Tetlock, 2004; Karpinski & Hilton, 2001; Olson & Fazio, 2004). From that perspective, implicit measures, such as the Implicit Association Test, are influenced by extra-personal associations, such as cultural norms, that do not contribute to one's evaluation. If implicit associations reflect cultural norms, then the more that one's explicit evaluation is seen as distinct from the norm, the less it should correlate with implicit associations. Likewise, when one's explicit evaluation is the same as the perceived cultural norm, implicit and explicit measures should be strongly correspondent. Nosek and Hansen (2005) raised doubts that implicit measures are influenced by cultural knowledge. The distinctiveness moderator provides another opportunity to test this hypothesis.

Summary of Hypothesized Moderators

This paper reports a study testing four moderators of the relationship between implicit and explicit evaluation. The moderators represent intrapersonal and interpersonal qualities of evaluation including: (a) self-presentation (the degree to which self-reported evaluations are adjusted for personal or social purposes); greater adjustment will be associated with less consistency because implicitly assessed evaluations are more resistant to deliberate alteration, (b) evaluative strength (practice, elaboration, and

related factors); greater strength will be associated with greater convergence of implicit and explicit evaluations because of the automatization of practiced evaluations, (c) dimensionality; the extent to which evaluations are represented with a simple, bipolar structure will be associated with stronger implicit-explicit correspondence because the simple structure of bipolarity will foster efficiency in information processing, and (d) distinctiveness (the extent to which one's evaluations are perceived as distinct from normative responses); greater distinctiveness may facilitate the accuracy of self-assessment through interpersonal comparison. These moderators are proposed to be non-redundant such that, when considered simultaneously, all will make unique contributions to the moderation of implicit-explicit relations.

Using Multilevel Modeling to Identify Moderators of Implicit-Explicit Correspondence

A wide variety of object pairs ($N=57$) and a large sample were used in this study to identify and test features differentiating the target concepts and to demonstrate the generality of effects. The scope of the study and the use of multilevel modeling for analysis enabled (a) simultaneous consideration of multiple moderators to investigate their unique relations to implicit-explicit correspondence, (b) tests of moderation between persons (Do variables that differ among individuals account for variation in implicit-explicit correspondence?), (c) tests of moderation across content domains (Do qualities of the content domains themselves account for variation in implicit-explicit correspondence?), and (d) tests of interactions between domains and person factors (Does variation between persons on a given variable account for implicit-explicit correspondence for some domains but not others?). Addressing all of these questions at once is a strength of multilevel modeling, and provides a methodological basis for testing some of the important conceptual questions about implicit-explicit relations. Further, the broad scope of this approach is comparable to a meta-analysis with the added value of constancy across methodological features and inclusion of all moderators at once for combined analysis that is not possible in meta-analytic methods.

Measuring multiple concepts and moderators at once. Research designs commonly test only one or a small subset of content domains and variables, thus limiting generalizability and sometimes requiring

speculation about the factors that distinguish domains. For example, Karpinski and Hilton (2001) compared implicit and explicit preferences toward flowers relative to insects and asserted that they should show strong correspondence compared to racial preferences because the former was presumed to elicit less self-presentation demand. However, there are many features that could differentiate these object pairs - such as their familiarity, importance, elaboration - making it impossible to pinpoint the role of self-presentation. Also, examining individual moderating variables in isolation prevents testing whether two or more apparently unique moderators are redundant.

The simultaneous examination of multiple moderators across a variety of domains with a large sample should (a) heighten confidence in the generality of the observations, (b) diminish the need for speculation about the operative factors differentiating the domains, (c) facilitate identification of redundancy among moderators, and (d) enable identification of interactions among moderators.

Implicit-explicit correspondence should be moderated by individual features of the preferences, not features inherent to the content domains. An important advantage of multilevel modeling is the capacity to partition variation related to variables nested in multiple levels of analysis. A common example is the influence of student, teacher, and school characteristics on student performance. Students are nested within teachers, which are themselves nested within schools, and features of each may contribute to student performance. With multilevel modeling, predictors of student performance based on student, teacher, and school variables can be examined simultaneously. If the hierarchical structure of the data is ignored, erroneous inferences may be drawn. Following the school example, the hypothetical researchers might, without nesting teachers within schools, observe that teachers with aides are associated with better student performance than teachers without aides. However, with nesting, they might learn that this apparent effect of teacher aides is actually a school effect of financial assets that provide for a variety of resources that together improve student performance. Or, if they considered schools, but used standard regression analysis to represent school-level data at the student-level, then the effects of school variables may be overestimated.

In the current study, individual evaluations are nested within content domains, represented by object pairs (e.g., Black-White, Jazz-Teen Pop, Approaching-Avoiding). Implicit-explicit correlations have been observed to vary from weak to strong depending on the domains examined (Nosek et al., 2002; Nosek & Smyth, 2005). To date, however, it is not clear whether those differences are due to characteristics of the evaluations, the object pairs, or both. With multilevel modeling we can test whether features of the individual evaluations, features of the object pairs, or interactions between the individual evaluation and object pairs are contributing to implicit-explicit correspondence.

There is good reason, both pragmatic and theoretical, to expect that individual characteristics of evaluations should dominate the prediction of implicit-explicit correspondence. Evaluations are personal in that they reflect something about the individual's experience with the social objects. A good theory about the relationship between implicit and explicit evaluations should identify moderators that reflect the individual's experience, not something about the domains being evaluated.

For example, by hypothesizing that racial attitudes will show weaker implicit-explicit correspondence than political attitudes, the hypothesis should not be that there is something inherent to racial and political domains that cause them to differ. Rather, the hypothesis is that the operative influence is an individually experienced variable that tends to differ on average between the domains (self-presentation concern perhaps). But, if the two domains were to elicit opposite amounts of self-presentation concern, then the prediction of which domain would show stronger implicit-explicit correspondence would reverse as well. It is only that the domains tend to differ on something experienced by persons, not a variable inherent to the domain itself that is presumed to cause the differences in implicit-explicit correspondence. A successful theory will identify the person factors that underlie the observed differences in implicit-explicit correlations between content domains and, through multilevel modeling, account for those domain-level differences.

Despite this admonition, it is *possible* that qualities inherent to target domains influence implicit-explicit correspondence. The consequence is that dual-process theories of evaluation would suffer from

lack of parsimony. Different models of implicit-explicit correspondence would be needed for various subsets of evaluative targets.⁴

From the perspective that implicit-explicit correspondence is grounded in individual experience, if evidence for moderation by features of the content domains is observed, it suggests that there are individual difference (person-level) moderators that remain unspecified or are not correctly specified. In this study, four moderators are presented with the expectation that each will contribute uniquely to moderation of the relationship between implicit and explicit evaluations across individuals. Any observation of between-objects moderation suggests that the model is incomplete in its specification of relevant individual experience of the evaluations. The multilevel modeling approach has the advantage of identifying moderating influences while simultaneously specifying the shortcomings of the model and suggesting avenues for improvement. With other approaches such limitations might go unnoticed and foster model misspecification.

Method

Participants

A total of 12,563 tasks were completed by 6,836 volunteer participants. A task is defined as completing implicit preference, explicit preference, and moderator measures for a single object pair. Participants were randomly assigned to one of the 57 object pairs and could complete as many as they wished. Random assignment occurred without replacement so that no participant completed the same object pair more than once. Participants averaged 1.8 tasks completed (Range = 1-56), and 4,439 (65%) of participants completed only one. Substantive results do not differ if analyses are based on the first completed task only.

Of the 6,836 participants who reported demographic information (98% response rate on average) the following was observed: 63% female, 37% male; 1% American Indian, 5% Asian, 5% Black, 5% Hispanic, 73% White, 1% Biracial (Black-White), 4% Multiracial, and 4% Other; 21% Conservative, 32% Neutral or Moderate, and 47% Liberal; 2% under 18, 38% 18-22, 27% 23-29, 17% 30-39, 10% 40-49, and 7% 50 and up; 4% Some high school or less, 6% high school diploma, 50% some college, 23%

bachelor's degree, and 16% advanced degree; 36% not at all religious, 33% slightly religious, 22% moderately religious, and 9% very religious; 4% lower class (self-identified), 16% lower-middle class, 52% middle class, 25% upper-middle class, and 3% upper class; and, 18% family income less than \$25,000, 23% 25K – 49,9K, 16% 50K – 74.9K, 19% 75K – 149.9K, 8% > \$150K, and 16% “don't know”. Following data cleaning (described below), a total of 11,501 completed tasks remained.

Materials and Apparatus

Evaluative target domains. The 57 object pairs were generated by the author on an ad-hoc basis to represent a wide range of topics across domains of social significance such as: social groups (e.g., Jews-Muslims, Tall People-Short People, Whites-Asians); political issues (e.g., Gun Rights-Gun Control, Pro-Choice-Pro-Life); individuals (e.g., Tom Cruise-Denzel Washington); pop culture, products, and things (e.g., Teen Pop-Jazz, Apple-Windows, Tea-Coffee); states and actions (e.g., Abstaining-Drinking, Emotions-Reasons, Hot-Cold); and ideas (e.g., Creationism-Evolution). A complete list of object pairs appears in the Appendix.

Implicit measure: Implicit Association Test (IAT). The IAT (Greenwald, et al., 1998) served as the operationalization of implicit evaluation (see Nosek, Greenwald, & Banaji, in press for a review). The procedure and analysis of the IAT followed the recommendations of Greenwald, Nosek, & Banaji (2003). The IAT requires rapid sorting of exemplars (pictures or words) representing two concept categories (e.g., creationism and evolution) and two attribute categories (good and bad for all IATs in the present study) into their superordinate categories with a standard set of seven response blocks: (B1) 20 trials sorting good and bad words using two response keys - e.g., 'good' words with the 'e' key, 'bad' words with the 'i' key, (B2) 20 trials sorting the two target concepts with the same two keys - e.g., 'creationism' words with the 'e' key, 'evolution' with the 'i' key, (B3) 20 trials sorting items from all four categories with the same two keys alternating by trial between concept and evaluative items - e.g., creationism and good with 'e', evolution and bad with 'i', (B4) 40 trials with same sorting rule as B3, (B5) 30 trials of sorting the concept categories with the reverse key mapping from 2 - i.e., 'creationism' with 'i' key and 'evolution' with 'e' key, (B6) 20 trials sorting items from all four categories with the opposite key pairings from B3

and B4 - i.e., evolution and good with 'e' key, creationism and bad with 'i' key, (B7) 40 trials with same sorting rule as B6. Blocks B3, B4, B6, and B7 comprise the critical data of the task. The concept+evaluation task pairing order was randomized across participants. The difference in average latency between the first sorting condition (B3, B4) and the second (B6, B7) was taken as the relative association strengths between the concepts and evaluations. In other words, participants who find it easier to sort creationism with good (and evolution with bad) compared to sorting creationism with bad (and evolution with good) were said to implicitly prefer creationism to evolution. IAT scores were calculated following the scoring algorithm recommended by Greenwald et al., (2003) with the following features: (1) error trials were removed and replaced with the block mean + 600 milliseconds, (2) response latencies < 400 ms were removed, (3) standard deviations were calculated on all correct response trials before introducing correction for error trials, (4) participants who had > 10% of trials with responses below 300ms were removed, and (5) participants who had an error rate of > 40% in any of the four combined sorting blocks were removed.

The resulting IAT D score is conceptually similar to a Cohen's d effect size measure indicating the direction and strength of associations between the concepts and evaluations. Positive IAT scores indicated a relative preference for the object that was implicitly preferred on average by the sample. The zero-point for the IAT score indicated implicit indifference – no difference in the association strengths between response blocks. The split-half correlation for the IAT was $r = .68$.

The IAT followed the procedure described above and had the following additional features: (a) concept category labels and exemplars were presented in capital letters, and evaluative category labels and exemplars were presented in lowercase, (b) trials in which a participant first made an error were accompanied by a red 'X' immediately below the stimulus item, and the task did not continue until the participant made the correct response, and (c) presentation of stimulus items in combined blocks alternated between concept and evaluative items.

Concept categories were represented by 3 to 8 stimulus images or words for use in the IAT. While IAT effects are primarily driven by the category labels with individual exemplars influencing the

construal of the category (De Houwer, 2001; Mitchell et al., 2003; Nosek, Greenwald, & Banaji, 2005), it is important to avoid items that are confounded with other categories or are difficult to categorize (Steffens & Plewe, 2001). Stimulus items were selected to be readily categorizable as members of the superordinate category and not easily confused as members of other categories in the same task. Seven sets of sixteen attribute items (eight 'good' items: e.g., wonderful, happy, joyful; and eight 'bad' items: e.g., awful, nasty, terrible) were also generated. Each of the 57 IATs was randomly paired with one of the seven evaluative stimulus sets.

IATs were presented via the Internet using a Java applet administered within Project Implicit software (see also Greenwald et al., 2003; Nosek et al., 2002, 2005). The applet used the respondent's computer resources to present stimuli and record response latencies thus avoiding dependence on the stability or speed of the Internet connection. Accuracy of response latencies with browser-administered applets is limited to the operating system's clock rate (e.g., 18.2 Hz for Windows-based machines). This limitation was not an issue for the IAT because of the randomness of the resulting noise, the strong effects elicited by the IAT, and the reduction of unsystematic error by averaging across performance trials.

Explicit measure: Feeling thermometer. Explicit preferences were assessed by calculating the difference between feelings of warmth ratings for the two social objects to conceptually parallel the relative measurement feature inherent in the IAT. Feelings of warmth were assessed on a 9-point scale (1-9) in which 1 indicated very cold feelings and 9 indicated very warm feelings. As such, the difference score between the two ratings had a range of possible values from -8 to +8 with 0 indicating explicit indifference between objects. Positive values indicated an explicit preference for the object that was implicitly preferred on average. Positive implicit-explicit correlations always indicated that stronger implicit preference for one object was associated with stronger explicit preference for the same object.

Candidate moderator: Self-presentation. Self-presentation is a heterogeneous construct represented by factors that may share a core quality of altering a response for personal or social purposes. This investigation deliberately simplified the heterogeneous quality of self-presentation by testing a composite factor of self-presentation with the individual measures examined secondarily on an

exploratory basis. Self-presentation was assessed by calculating the mean of internal motivation to respond without negativity, external motivation to respond without negativity (both adapted from Plant & Devine, 1998), and an estimate of how much the ‘average person’ seeks to avoid appearing biased against the social object ($\alpha = .73$). Internal motivation was assessed as the mean of two items (“Being accepting of <object> is important to my self-concept”, “Because of my personal values, I believe that making negative judgments about <object> is wrong.”), external motivation was assessed as the mean of two items (“I try to hide negative thoughts about <object> to avoid negative reactions from others.”, “I attempt to appear accepting of <object> to avoid disapproval from others.”), and one item assessed concerns of the average person (“How motivated is the average person to conceal negative feelings about <object>?”).⁵ Internal and external items used a 9-point Likert rating scale from 1 = strongly disagree to 9 = strongly agree, ‘average person’ items used a 9-point Likert scale from 1 = not at all motivated to 9 = extremely motivated. Higher values indicated greater self-presentation concern.

Candidate moderator: Evaluative strength. Like self-presentation, evaluative strength is a heterogeneous construct (see Krosnick & Petty, 1995 for a review). Again, because of the lack of theory to predict which particular aspect of strength would moderate implicit-explicit correspondence, this investigation used a mean composite ($\alpha = .72$) with individual measures examined secondarily on an exploratory basis. The items assessing evaluative strength for the present study were meta-attitudinal indices (Bassilli, 1996) and included importance (Cantril, 1944; “How personally important is your attitude toward <object>?”), thought frequency (Bassilli, 1996; “How much do you think about your feelings toward <object>?”), and familiarity (“Compared to other people, how familiar are you with <object>?”). All items used a 9-point Likert rating scale from 1 = not at all (important, often, familiar) to 9 = extremely (important, often, familiar). Higher values indicated stronger evaluations.⁶

Candidate moderator: Dimensionality. Dimensionality was assessed by calculating the mean level of agreement to two items ($r = .75$) for a given task: “Having positive attitudes toward <object X> implies having negative attitudes toward <object Y>” and “Having positive attitudes toward <object Y> implies having negative attitudes toward <object X>”. Both items used a 9-point Likert rating scale from 1 =

strongly disagree to 9 = strongly agree. Higher values indicated that participants perceived the object pair to be more bipolar – evaluations of one of the predicted an opposing evaluation of the other object. The dimensionality items were introduced part-way through data collection ($N = 6,823$).

Candidate moderator: Perceived distinctiveness. Distinctiveness was defined as the extent to which one's own evaluations were thought to differ from the average person's. In addition to rating their own explicit evaluations, participants estimated the average person's evaluations using the same scale. Perceived distinctiveness was calculated as the absolute value of the difference between these ratings. Averaged across the two target objects, possible values ranged from 0 (no perceived discrepancy between self and the average person) to 8 (directly opposing self and perceived other evaluations). The 'average person' items were introduced part way through data collection ($N = 2,957$).

Between-objects moderators. All of the above moderators index an individual's experience of the target objects. Individuals vary in the extent to which they ascribe these qualities to any given object pair, and the mean of responses across individuals reflects the normative response for the sample. Means of each moderator for each object pair index self-presentation, strength, dimensionality, and distinctiveness ratings for potential between-objects moderation of implicit-explicit correspondence. These values are theoretically independent from the individual ratings and reflect qualities of the object pairs.

Procedure

The study was administered via the research website for Project Implicit (version 1.5; <http://implicit.harvard.edu>) between January 10, 2003 and September 6, 2003. It was open to the Internet public and participation was voluntary. Participation in research at the Project Implicit website required identity registration with a demographic questionnaire. After logging in, participants were randomly assigned to a study in the Project Implicit study pool. For the purposes of assignment, each object pair reported in this paper was treated as an individual study. Participants were never assigned to the same study (object pair) more than once. Once randomly assigned, participants completed implicit and explicit measures in a randomized order. Explicit measures including items assessing each moderator were presented on one page. The order of items was randomized for each participant.

Analysis Strategy

A series of two-level multilevel models comprise the primary analyses. Multilevel modeling is a powerful tool for nested data. In the present context, preferences of persons were nested within object pairs. Between-objects and between-persons sources of variability are theoretically independent. The value of multilevel modeling is in the ability to separate these sources of variability and avoid biased inferences in which apparent differences across persons could be due to unidentified causes across objects, or vice versa.

The main multilevel analysis treated explicit preferences as a dependent variable and implicit preferences as an independent variable. Moderators of implicit-explicit correspondence were tested as product terms between the proposed moderator and implicit preferences (Baron & Kenny, 1986).

All object-level moderator variables were centered on the object pair grand mean, and if object-level moderators were significant, person-level moderator variables were centered on their respective object pair means. As such, object-level moderators were entered into models first to test whether variation in implicit-explicit correspondence could be explained by differences across objects (e.g., Do preference for object pairs with greater self-presentation demand on average show weaker implicit-explicit correspondence than object pairs with less self-presentation demand?). As a second step, person-level moderators were entered into the models to test whether variation between persons could explain variation in implicit-explicit correspondence beyond that explained by variation between objects (e.g., Does an individual's concern with self-presentation for a particular object pair predict the strength of his or her implicit-explicit correspondence beyond that explained by differences between object pairs?). As a final step, moderators were tested as random effect predictors of implicit-explicit correspondence. A significant random effect interaction between a moderator and implicit preferences indicates that the extent to which a given variable moderates implicit-explicit correspondence varies as a function of the objects (e.g., Does self-presentation moderate implicit-explicit correspondence between persons for some object pairs but not for others?).

Results

Summary of Implicit and Explicit Preferences Across Content Domains

Implicit and explicit preferences were measured for 57 different object pairs and descriptive statistics for each are summarized in Table 1 sorted by the magnitude of the implicit-explicit correlation. The object of each pair that was implicitly preferred on average appears in the first column. 48 of the 57 tasks revealed a significant implicit preference for one object over the other – i.e., the mean effect significantly differed from zero.

For the explicit preferences, positive mean values indicate that the object implicitly preferred on average was also explicitly preferred. 46 of the 57 tasks revealed a significant explicit preference for one object over the other. Bold typeface indicates object pairs for which the implicit and explicit mean preferences were in opposite directions and both significantly different from zero.

Product-moment correlations between the implicit and explicit measures are listed from highest to lowest. The average correlation was $r = .36$, and 52 of the 57 object pairs showed a significant positive correlation. The correlations range from a low of $r = -.05$ (females-males) to a high of $r = .70$ (Pro Choice-Pro Life). Females-males was the only domain with an implicit-explicit correlation below $r = .10$. Patterns in the females-males data suggested that the correlation was attenuated because participants interpreted the explicit item as either a general measure of liking for the group, or as a measure of sexual attraction to members of the group. Depending on one's gender and sexual orientation, these interpretations would lead to very different evaluations of the target groups.

Stage 1: A General Account of the Relationship between Implicit and Explicit Evaluations

Predictions about the relationship between implicit and explicit evaluations were tested by comparing the fit of successive models. Deviance scores from χ^2 were the basis for comparing models. Higher values indicate greater improvement in fit. Results for Stage 1 are summarized in Table 2. To give a preview of the results in words, the three models of Stage 1 tested and supported three predictions. Establishing a baseline model, Model M_0 showed that some content domains elicited stronger relative preferences on average than other domains. In comparison to the initial model, Model M_1 showed that implicit and explicit evaluations were related. The last model of Stage 1 supported the prediction that the

strength of implicit-explicit correspondence varied as a function of the content domain. These results are readily observable by interpreting data in Table 1, but these models provide the baseline from which to test the individual moderators in Stage 2.

Prediction 0: Explicit preferences vary across content domains. Model M_0 tests a preliminary prediction that some object pairs will, on average, elicit stronger relative preferences for one of the objects (e.g., flowers are likely to be strongly preferred to insects), than will other domains (e.g., pants may only be mildly preferred skirts). Confirmation of this prediction would not be surprising, but it would illustrate the utility of a multilevel approach for these data – i.e., preferences vary as a function of individuals *and* domains. The inferential test of this prediction is indicated by the estimate of the random effects intercept variance in the first column of Table 2. The coefficient was sizable ($B = 1.09$, $SE B = .22$, $z = 5.09$, $p < .0001$) confirming that stronger explicit preferences were evidenced for some object pairs on average than for others.

Prediction 1: Implicit preferences are related to explicit preferences. Compared to Model M_0 , Model M_1 tested whether implicit preferences were related to explicit preferences. This model fit the data much better than Model M_0 , $\Delta\chi^2(1) = 1735$, $p < .0001$. Also, the specific inferential test for this prediction is indicated by the fixed-effects estimate for the IAT in the second column of Table 2. The coefficient was significant ($B = 1.92$, $SE B = .09$, $t = 43.39$, $p < .0001$, $d = .81$) indicating that implicit and explicit preferences were positively related. This demonstration is the conceptual equivalent of a correlation between implicit and explicit preferences across the entire sample ignoring what objects were evaluated.

Prediction 2: The relationship between implicit and explicit evaluations depends, in part, on the content domain. Compared to Model M_1 , Model M_2 tested the prediction that the strength of the relationship between implicit and explicit preferences would vary as a function of the content domains. Previous research suggests that implicit-explicit correspondence varies from near zero to .86 depending on the domain (Greenwald et al., 2003; Nosek & Banaji, 2002; Nosek et al., 2002). This study directly compared the strength of correspondence across a multitude of object pairs. The third column of Table 2

summarizes Model M₂. The addition of the IAT variable as a random effect - that the strength of the implicit-explicit relationship varied as a function of the objects - improved model fit over Model M₁, $\Delta\chi^2(1) = 531, p < .0001$, and that random effect IAT variable was a substantial influence ($B = 1.18, SE B = .24, z = 4.9, p < .0001$). This confirms that the strength of implicit-explicit correspondence varied as a function of the target objects.

Note that if the moderation of implicit-explicit correspondence is person-centered, meaning that variation in implicit-explicit correspondence is caused by individual experiences with the domains not something about the domains themselves, then including the person-level moderators should cause this random effect IAT value to decline in subsequent models. A full account of variation in implicit-explicit correspondence across object pairs would be observed when the random effect IAT variable is reduced to zero.

Summary. Stage 1 resulted in a descriptive model of the relationship between implicit and explicit preferences noting differences between objects and persons (Model M₂). This provides the basis for Stage 2 that introduces the variables hypothesized to moderate implicit-explicit relations.

Stage 2: Testing Moderators of the Relationship between Implicit and Explicit Evaluations

Stage 2 of the multilevel analyses examined each proposed moderator in turn. Model M₂ from Stage 1 served as the base model for comparison. This model was refitted for the sample available for each moderator. Table 3 presents the best model for each proposed moderator (self-presentation, strength, dimensionality, distinctiveness) following analysis of the a priori defined models described in the Method. In all cases, interaction terms between the proposed moderator and IAT effects served as the tests for moderation. Predictors from models were only retained if they were significantly improved overall fit. For simplicity, only the best fitting models are discussed in the text. Parsimony concerns are addressed in Stage 3. Full tables describing all tested models are available at <http://briannosek.com/>.

To summarize the results, all four variables showed some evidence of moderating implicit-explicit correspondence as predicted in the introduction. Higher self-presentation concerns were associated with weaker implicit-explicit correspondence; stronger evaluations were associated with

stronger implicit-explicit correspondence; evaluations conforming to a more bipolar structure showed stronger implicit-explicit correspondence than more unipolar ones; and, evaluations that were more distinct from the norm were associated with stronger implicit-explicit correspondence than were evaluations that were perceived to be normative. Some moderators appeared to operate across object domains and perhaps vary in their moderating influence as a function of domain. These issues are examined more closely in Stage 3 when the moderators are tested simultaneously.

Prediction 3: Variability in implicit-explicit correspondence is predicted by self-presentation concerns. The variation observed in implicit-explicit correspondence is not random. Prediction 3 reflected the expectation that self-presentation concern played a role in accounting for that correspondence. Implicit measures may capture evaluative associations that respondents do not report either because they do not endorse those feelings, or because they are concerned about the social consequences of reporting them.

Of the candidate models, Model SP₂ best reflected the moderating role of self-presentation and showed improved model fit over Model M₂, $\Delta\chi^2(4) = 208$, $p < .0001$. Self-presentation showed significant but weak moderation of implicit-explicit correspondence between-objects ($B = -.39$, $SE B = .19$, $t = -2.12$, $p = .03$, $d = .04$) and between-persons ($B = -.11$, $SE B = .03$, $t = -2.96$, $p = .003$, $d = .06$). The negative estimates for both indicate that, as predicted, higher self-presentation concerns were associated with weaker implicit-explicit correspondence.

Self-presentation was a composite of three measures – external motivation to respond without negativity, internal motivation to respond without negativity, and perceived norms to avoid appearing biased. Follow-up exploratory analyses examined the contributions of each of these factors and are described briefly. Considered separately, two of the three components showed small but significant moderating influence between persons (external motivation $d = -.07$, $p = .0003$; internal motivation $d = -.05$, $p = .02$). The effect of perceived norms was not significant ($p = .31$). When all three components were examined as simultaneous predictors, only external motivation continued to show significant

moderating influence ($p = .005$) suggesting that it was the primary factor illustrating that greater self-presentation concern was associated with weaker implicit-explicit correspondence.

Prediction 4: Variability in implicit-explicit correspondence is predicted by evaluative strength.

Evaluative strength is a general term capturing the conceptual similarities among a variety of features of evaluation (e.g., importance, thought frequency). Stronger evaluations were predicted to elicit greater implicit-explicit correspondence. Of the candidate models, Model ST₃ best reflected the moderating role of strength and improved model fit over Model M₂, $\Delta\chi^2(4) = 424$, $p < .0001$. Strength moderated implicit-explicit correspondence between-persons ($B = .30$, $SE B = .05$, $t = 5.86$, $p < .0001$, $d = .11$) and a significant random effects factor suggested that the moderating role of strength between-persons varied between-objects ($B = .092$, $SE B = .026$, $z = 3.47$, $p = .0003$). In other words, stronger preferences were associated with higher implicit-explicit correspondence for some domains and not for others. The positive estimates for both effects indicated that, as predicted, stronger evaluations were associated with greater implicit-explicit correspondence than were weaker ones.

Evaluative strength was represented as a composite of three measures – importance, familiarity, and thought frequency. Follow-up exploratory analyses examined the contributions of each of these factors. Considered separately, each of the three components supported the general observation that stronger evaluations were associated with greater implicit-explicit correspondence (importance $d = .20$, $p < .0001$; familiarity $d = .11$, $p < .0001$; thought frequency $d = .15$, $p < .0001$). When combined into a single analysis, all three remained significant ($p < .05$) suggesting that each component of the heterogeneous strength measure helped to account for implicit-explicit correspondence.

Prediction 5: Variability in implicit-explicit correspondence is predicted by evaluative dimensionality. Dimensionality reflects the extent to which object pairs were perceived to sit on ends of a bipolar continuum. Evaluations perceived as highly bipolar were predicted to elicit stronger implicit-explicit correspondence compared to evaluations that were more unipolar. Of the candidate models, Model DIM₂ best reflected the moderating role of dimensionality and improved model fit over Model M₂, $\Delta\chi^2(4) = 77$, $p < .0001$. Dimensionality moderated implicit-explicit correspondence between-objects ($B =$

.55, $\underline{SE\ B} = .088$, $\underline{t} = 6.26$, $\underline{p} < .0001$, $\underline{d} = .15$) and between-persons ($\underline{B} = .20$, $\underline{SE\ B} = .026$, $\underline{t} = 7.36$, $\underline{p} < .0001$, $\underline{d} = .18$). The positive estimates for both indicate that, as predicted, evaluations with a more bipolar dimensional structure were associated with greater implicit-explicit correspondence than were more unipolar ones.

Prediction 6: Variability in implicit-explicit correspondence is predicted by evaluative distinctiveness. Distinctiveness reflects the extent to which individuals perceived their evaluations to differ from the average person's. High distinctiveness suggests that the evaluations are likely to be thought of as more uniquely personal than ones perceived to be consistent with group norms. Higher distinctiveness was predicted to elicit stronger implicit-explicit correspondence than lower distinctiveness.

Of the candidate models, Model DIS₃ best reflected the moderating role of distinctiveness and improved model fit over Model M₂, $\Delta\chi^2(6) = 342$, $\underline{p} < .0001$. Distinctiveness moderated implicit-explicit correspondence between-objects ($\underline{B} = 1.10$, $\underline{SE\ B} = .36$, $\underline{t} = 3.07$, $\underline{p} = .002$, $\underline{d} = .12$), between-persons ($\underline{B} = .52$, $\underline{SE\ B} = .08$, $\underline{t} = 6.20$, $\underline{p} < .0001$, $\underline{d} = .24$), and the moderating role of distinctiveness between persons varied as a function of the content domains (i.e., the random effects factor; $\underline{B} = .16$, $\underline{SE\ B} = .07$, $\underline{z} = 2.42$, $\underline{p} = .008$). The positive estimates for all three indicate that, as predicted, higher perceived discrepancies between one's own evaluations and the perceived group norms were associated with stronger implicit-explicit correspondence than were lower discrepancies.

Summary. All four variables showed evidence of moderating implicit-explicit correspondence between persons as predicted. However, in three cases the variables showed additional evidence of moderating implicit-explicit correspondence between objects and, in two cases, the presence of moderation across individuals varied as a function of the content domain. These latter effects suggest limitations for the moderators considered in isolation in accounting for the implicit-explicit relationship. In Stage 3, the moderators are tested simultaneously to test whether the moderators are redundant influences on implicit-explicit relations. This also provided an opportunity to test whether considering multiple moderators at once will diminish the observed differences between content domains. If implicit-

explicit correspondence is rooted in individual experience, as outlined in the introduction, then the presence of variation in correspondence between content domains suggests that single moderator models fail to capture important information that happens to vary across objects, that could be assessed as an individual difference variable if properly identified. So, if including multiple moderators at once eliminates significant differences between content domains, it would show that the individual difference moderators are accounting for differences that had appeared to be due to the object pairs. Before that, the relationships among the moderating variables are presented.

Relationships among Moderators

The product-moment correlations between the moderators are shown in Table 4. Across individuals, there were minimal to moderate positive relations among the moderators. The strongest relationship, $r = .41$, between self-presentation and strength suggests that stronger evaluations are accompanied by greater self-presentation concerns. Interestingly, despite being positively correlated, these two variables have opposite moderating roles for implicit-explicit correspondence. For strength, stronger evaluations were associated with greater implicit-explicit correspondence. But for self-presentation, higher self-presentation concerns were associated with weaker implicit-explicit correspondence. This suggests an interesting interplay between evaluative strength, self-presentation, and implicit-explicit correspondence. Evaluations for which a person experiences self-presentation concerns are also likely to be strong. To the extent that these variables play a role in moderating implicit-explicit correspondence, increasing strength would foster stronger implicit-explicit correspondence while increasing self-presentation concern would simultaneously suppress the magnitude of that relationship. This might explain the rather modest moderating influence of self-presentation observed in Stage 2. When analyzed simultaneously, a suppressive effect may be revealed and ultimately show a more substantial moderating role for self-presentation. In the next section, the moderator variables were investigated together to determine their unique moderating influence on the implicit-explicit relationship.⁷

Stage 3: A Model of the Relationship between Implicit and Explicit Evaluations

The goal of Stage 3 was to establish a model of the relationship between implicit and explicit evaluations. Here, the four moderators examined separately in Stage 2 are combined into a single analysis to test for redundancy among the moderators and examine their combined potency in accounting for implicit-explicit correspondence. As argued earlier, the implicit-explicit relationship should be determined by personal experiences of the concept domain, and not by variables inherent to the object pairs. Support for this perspective would be observed if the moderating influence of the variables persisted between persons, but was diminished or eliminated between objects.

The first model (F_1) included all of the significant moderating factors identified in the four models from Stage 2. The predictors that did not retain significance in the composite model were redundant and removed one by one until only significant predictors remained (Model F_2). Then, a simpler alternative model (Model F_3) was identified as offering a balance of parsimony and explanatory power. As predicted, the final model shows that each moderator is unique in accounting for implicit-explicit relations. Also, inclusion of all four moderators dramatically reduced the variation in implicit-explicit relations between content domains suggesting that a significant portion of the meaningful differences between domains is accounted for by the four moderating variables and that all four are important to reduce model misspecification.

The multilevel modeling results from Stage 3 appear in Table 5. For comparison purposes, the first three columns present Models M_0 , M_1 , and M_2 reanalyzed using only observations for which data for all four moderators was available. All three models remain consistent with the observations from Stage 1. Model F_1 includes all significant moderators from Stage 2 and shows substantially improved model fit over Model (F) M_2 , $\Delta\chi^2(18) = 392$, $p < .0001$. There is also evidence that some of the significant effects from Stage 2 are redundant and should be removed.

Backward stepwise removal of non-significant predictors resulted in the elimination of two between-objects moderators from Stage 2. Model F_2 includes only the moderators that showed a significant interactive effect with implicit evaluations in predicting explicit evaluations. Despite the removal of some factors, estimated model fit was as good as the very inclusive model. Remarkably, even

in this simultaneous analysis, all four variables showed evidence of moderating the relationship between implicit and explicit evaluations between persons. Further, a number of between-object factors were no longer significant predictors suggesting that the inclusion of multiple between-person moderators accounted for variation in implicit-explicit correspondence that might otherwise have been attributed to features of the object pairs.

Additional evidence of the value of these four between-person moderators is seen in the decrease in the IAT random effect factor between Model (F)M₂ ($\underline{B} = 1.16$) and Model F₂ ($\underline{B} = .37$). This random effect factor reflects the variation in implicit-explicit correspondence that was observed between objects, with some objects showing relatively strong correspondence (e.g., Pro Life-Pro Choice) and others showing weak correspondence (e.g., Fat People-Thin People). The fact that the IAT random factor was reduced by 68% suggests that these four moderators were quite successful in identifying meaningful differences between content domains as actually being due to individual differences that happen to differ on average between domains.

In sum, four factors for knowing when implicit and explicit preferences will correspond are the self-presentation, strength, dimensionality, and distinctiveness associated with personal evaluative experiences. Further, differences observed between content domains can be understood as reflecting the fact that domains tend to vary on average in what they elicit from individuals in each of these dimensions rather than there being something inherent to the domain affecting implicit-explicit correspondence.

And yet, Model F₂ still included two small but significant between-objects effects and one person-object interaction (the distinctiveness random effect) suggesting that the model is not a comprehensive account of variation in implicit-explicit correspondence across content domains. As a final step, Model F₂ was compared with the same model after removing the two between-objects effects (Model F₃). Model F₃ is a more parsimonious model than Model F₂ and fit equally well, $\Delta\chi^2(4) = 1$, $p = .91$. This suggests that a more parsimonious model ignoring the weak between-objects effects is as effective as the more theoretically cumbersome model that includes those moderators. The between-object effects suggest that there may be something about individual experience that is not yet captured by

the individual difference assessments and is being mistaken as a difference between content domains.

Their minimal impact on model fit whether included or removed suggests that they are inconsequential.

For this reason, Model F_3 was advanced as the preferred model.

Model F_3 shows that personal assessments of self-presentation ($\underline{B} = -.26$, $\underline{SE\ B} = .06$, $\underline{t} = -4.45$, $\underline{p} < .0001$, $\underline{d} = .17$), strength ($\underline{B} = .22$, $\underline{SE\ B} = .05$, $\underline{t} = 4.04$, $\underline{p} < .0001$, $\underline{d} = .15$), dimensionality ($\underline{B} = .21$, $\underline{SE\ B} = .03$, $\underline{t} = 6.25$, $\underline{p} < .0001$, $\underline{d} = .24$), and distinctiveness ($\underline{B} = .40$, $\underline{SE\ B} = .08$, $\underline{t} = 4.83$, $\underline{p} < .0001$, $\underline{d} = .18$) all moderated the strength of the relationship between implicit and explicit preferences in the same way that they did in Stage 2. This time, they did so in a single model showing that they are non-redundant. Note that the effect size of self-presentation as a moderating factor was larger in the combined model (F_3) than it was in isolation (SP_2) confirming the earlier speculation that the positive relationship between self-presentation and evaluative strength was suppressing its moderating influence.

These results suggest that the relationship between implicit and explicit preferences is determined by multiple individual difference assessments of the characteristics of one's evaluation.⁸

Identifying possible misspecification of distinctiveness factor. The final model (F_3) contained a random effect predictor of the interaction between implicit preference and distinctiveness ($\underline{B} = .15$, $\underline{SE\ B} = .06$, $\underline{z} = 2.52$, $\underline{p} = .006$) showing that distinctiveness varied across content domains in the extent to which it moderated implicit-explicit correspondence. Said another way, distinctiveness was not a consistent moderator; its prediction if implicit-explicit correspondence depended on the assessed domain. Theoretically, this is undesirable from the perspective that the relationship between implicit and explicit evaluation should derive from factors internal to the individual, not inherent in the targets of assessment. If one agrees that a strong theory of implicit-explicit correspondence would make the content domain irrelevant, then this random factor indicates incompleteness or misspecification in the current model. This illustrates an important advantage of multilevel modeling in that it can identify possible sources of misspecification. Approaches that examine a single content domain at a time may promote an 'ignorance is bliss' treatment of these forms of misspecification. If misspecification due to mixing levels of analysis is not measured, then it can not be identified, modeled, or removed.

The significant random factor suggests that distinctiveness, as currently assessed, may conflate two constructs that have different influences on implicit-explicit correspondence. One possibility, implied by the introduction, is that high distinctiveness should foster stronger implicit-explicit correspondence to the extent that norms are perceived accurately on average. If respondents believe their evaluations to be distinct, but are consistently incorrect, then using knowledge of others' evaluations might actually disrupt rather than enhance interpersonal calibration of self-report. Inaccurately perceived distinctiveness may actually predict weaker implicit-explicit correspondence, in opposition to accurately perceived distinctiveness that may be driving observed the overall pattern of predicting strong implicit-explicit correspondence.

It is not obvious if this possibility accounts for the random effect of distinctiveness observed here. A crude follow-up analysis suggests that it may be so. Regression estimates of the IAT by distinctiveness interaction were calculated for each content domain. These estimates reflect the variation in magnitude of distinctiveness moderating implicit-explicit correspondence across content domains. Those estimates were correlated with the absolute difference between the mean explicit preference for each domain and the mean perceived cultural preference. Higher values on this second variable indicate greater systematic misperception of others' evaluations (biased by the selectivity of the sample). The correlation was moderately negative, $r = -.38$, meaning that greater misperception of others' evaluations was associated with a weaker moderating influence of distinctiveness on implicit-explicit correspondence. This suggests that higher perceived distinctiveness will predict greater implicit-explicit correspondence only to the extent that the perceived distinctiveness is not consistently discrepant from reality. When systematically off target, distinctiveness may actually decrease implicit-explicit correspondence by providing an inappropriate point of comparison. This hypothesis requires future systematic scrutiny.

Discussion

Automatic and controlled processes provide distinct assessments of the evaluative quality of social objects. The purpose of the present study was to identify model of multiple moderators to account for when automatic (implicit) and controlled (explicit) evaluative processes will converge or diverge.

Results suggested that implicit and explicit evaluations were related, that the magnitude of correspondence varied between persons and between content domains, and that multiple variables moderated the magnitude of correspondence.

The relationship between implicit and explicit evaluations for a given domain can be understood by considering individuals' experience of (a) self-presentational concern – greater self-presentation concern predicted weaker implicit-explicit relations, (b) evaluative strength – stronger evaluations were associated with greater consistency between implicit and explicit measures than were weaker evaluations; (c) dimensionality – greater implicit-explicit correspondence was observed for social objects perceived to be on either end of a bipolar continuum than those perceived to be more unipolar; and (d) distinctiveness – personal evaluations thought to be distinct showed greater implicit-explicit correspondence than evaluations perceived to be normative. Each variable contributed uniquely to the moderation of implicit-explicit correspondence suggesting that the relationship is multiply-determined, and that the locus of causation for converging or diverging implicit-explicit relations is in individual minds, not in qualities inherent to the evaluated content domains. Further, the variables highlighted the fact that implicit-explicit correspondence is likely determined by both intrapersonal qualities of the evaluation (strength, dimensionality), and interpersonal experiences of one's evaluation in a social context (presentation, distinctiveness).

These results advance the understanding the relationship between automatic and controlled evaluative processes. A comprehensive model can be achieved by broadening the generality of these inferences (e.g., across multiple implicit and explicit measures), increasing the specificity of the operative indices (e.g., what aspects of evaluative strength are involved as moderators?), identifying additional moderators or measurement factors that account for the remaining unexplained variation, and linking these moderators to the mechanisms that give rise to intentional and automatic evaluation.

Implicit and Explicit Evaluations are Related, but Distinct Constructs

Implicit and explicit evaluations appear to be distinct constructs with a relationship that is moderated by intrapersonal and interpersonal evaluative features. This conclusion is bolstered by other

lines of research including: (a) psychometric evidence that implicit and explicit evaluations are related but distinct (Cunningham et al., 2001; Greenwald & Farnham, 2000; Nosek & Smyth, 2005), (b) dual-process theories that suggest distinct evaluative or related processes (Carver, in press; Greenwald & Banaji, 1995; Smith & DeCoster, 2000; Strack & Deutsch, 2004; Wilson et al., 2000), (c) neurological evidence that implicit and explicit measures correspond with distinct cognitive processes (Cunningham, Johnson, Gatenby, Gore & Banaji, 2003; Cunningham, Johnson, Raye, Gatenby, Gore & Banaji, 2004; Phelps et al., 2000), and (d) a recent meta-analysis of predictive validity studies showing that explicit and implicit evaluations both have domains of predictive potency (Poehlman, et al., 2004). The results from the present study bolster this implicit-explicit distinction by identifying factors that provide some account of their differences and similarities.

Implicit Evaluations are Personal

In 1989, Devine proposed that the automatic associations revealed by her evaluative priming measure reflected cultural-shared stereotypes, not individually possessed evaluations. In 1995, Fazio and colleagues showed that there was meaningful variability in automatic associations illustrating that they are personal, individual difference assessments of evaluation (Fazio et al., 1995). Greenwald et al. (1998) interpreted their new measure, the Implicit Association Test, similarly as a measure of individual evaluation. Karpinski and Hilton (2001) disagreed and argued that the IAT measured cultural knowledge, and not evaluations that indicated something about the person.

The present data show that the IAT, this paper's operationalization of implicit evaluations, is an individual difference measure corresponding in varying degrees with self-reported attitudes, and the strength or weakness of that relationship is determined by individually possessed variables such as evaluative strength or perceived distinctiveness. This bolsters other evidence showing that IAT-measured evaluations have predictive validity for individual judgment and action (Poehlman et al., 2004).

With a more nuanced hypothesis, Olson and Fazio (2004) argued that the IAT measures personal evaluations but that it is contaminated by extra-personal associations – such as knowledge of cultural norms – that is irrelevant to the person's evaluation. Nosek and Hansen (2005) measured self-reported

attitudes and extra-personal knowledge across a wide variety of content domains and found substantial evidence for relations between the IAT and self-reported attitudes, but no evidence to support a link between the IAT and extra-personal knowledge. The current data presented an opportunity to replicate this observation with the measure of perceived cultural evaluations used to create the distinctiveness index (see Method). A multilevel analysis regressed IAT scores on self-reported preferences ($\underline{M} = .44$, $\underline{SD} = 3.1$) and perceived cultural preferences ($\underline{M} = .42$, $\underline{SD} = 2.6$) simultaneously. Consistent with the effects reported earlier, self-reported preferences related to implicit preferences across persons ($\underline{d} = .55$, $\underline{p} < .0001$) and the magnitude of that relationship varied across content domains ($\underline{z} = 2.6$, $\underline{p} = .005$). However, cultural knowledge was not related to implicit preferences at all ($\underline{d} = .01$, $\underline{p} = .83$) and that non-relationship did not differ across domains ($\underline{z} = -.74$, $\underline{p} = .23$). The absence of an effect here is remarkable considering the extremely high power with the large sample. This replicates Nosek and Hansen's (2005) evidence that explicit cultural knowledge has no unique relationship to the IAT.

As mentioned in the introduction, the distinctiveness moderator provided an additional opportunity to examine the hypothesis that cultural knowledge is reflected in IAT performance (Arkes & Tetlock, 2004; Karpinski & Hilton, 2001; Olson & Fazio, 2004). High distinctiveness reflected a perceived dissociation between one's own evaluation and the cultural norm, whereas low distinctiveness reflected a perception that one's own evaluation was the same as the norm. If the IAT reflected cultural knowledge, then the distinctiveness moderation of implicit-explicit correspondence should have been negative – explicit evaluations more similar to the cultural norm should have been more strongly related to implicit evaluations. However, the opposite was observed. Highly distinctive personal evaluations were associated with stronger implicit-explicit correspondence, and personal evaluations perceived as relatively indistinct from the norm were associated with weaker implicit-explicit correspondence. This effect is difficult to reconcile with the hypothesis that extra-personal associations, in the form of cultural knowledge, influence IAT performance.

The IAT reflects individual differences in evaluation that are sometimes similar and other times opposing those evaluations that are endorsed and experienced as one's own. The present data suggest that

the difference between explicit and implicit evaluation can not be understood as reflecting the difference between attitudes and knowledge. Rather, understanding the relationship between implicit and explicit evaluation requires consideration of multiple moderating variables – such as self-presentation concern, evaluative strength, dimensionality and distinctiveness – that reflect an individual's intrapersonal and interpersonal experience with the evaluative objects.

Introspective (Un)awareness of Implicitly Assessed Evaluations

Another important issue is whether the associations tapped by implicit measures are available to conscious introspection. These results suggest that implicit measures are positively, though variably, correlated with explicit measures. Kihlstrom (2004) argued that positive implicit-explicit correlations is evidence for awareness of the associations tapped by implicit measures, which would lead one to conclude that the associations measured by the IAT are not unconscious. However, implicit and explicit evaluations can be positively correlated even if implicit measures tap associations that are unavailable to introspection. Consider, for example, making estimates about another person's evaluations. If I was asked to predict the attitudes of my siblings and neighbors, my guesses would positively correlate with their reports, and more so with my siblings than my neighbors. This does not indicate that I have introspective access into the minds of others. A correlation between my estimates and my siblings' reports reflects my observational knowledge of the attitude-relevant actions of my siblings. The fact that the correlations with siblings and neighbors are different reflects the amount and quality of experiences I have had observing my siblings' and neighbors' attitude-relevant actions.

Following self-perception theory (Bem, 1972), the relationship between implicit and explicit evaluations could reflect the quality of self-observation, not introspective access to the associations assessed by implicit measures. Some evaluations are consistent, practiced, clear, and distinct leading, perhaps, to a relatively precise self-observation and strong implicit-explicit correspondence. Other evaluations are inconsistent, novel, ambiguous, and normative, leading, perhaps, to imprecise self-observation and little opportunity for implicit and explicit assessments to converge.

It is also possible that the processes involved in generating an implicit response are influenced by explicit processes and by unconscious sources of information processing simultaneously. As illustration, imagine that you were estimating the proportion of red and green balls in a large black box that was sealed except for two holes for adding balls. You controlled the balls that were put into one hole (explicit), and the other hole accepted balls from an unobservable source (implicit). You would likely estimate the proportion of red to green balls in the box based on what you put into it. While you can't directly observe them, your estimate would be accurate to the extent that the number of balls that you put in was a large proportion of the overall number in the box, or that the proportion that you dropped in happened to match the proportion entered via the unobservable source. For those content domains in which a high proportion of the evaluative information comes from explicit sources, there may be a high implicit-explicit correlation compared to domains in which a high proportion of the evaluative information comes from other sources. In sum, the presence of implicit-explicit correlations does increase the plausibility that implicit measures (the IAT in this case) measure some associative information that is also reflected in subjective experience. However, the present data are non-diagnostic about whether people are aware of implicitly measured associations.

Some evidence suggests that implicit evaluations may be influenced by explicit practice because of chronic goals or motivations (Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Moskowitz, Gollwitzer, Wasel, & Schaal, 1999). This work fits nicely with the evidence for moderation of implicit-explicit correspondence by evaluative strength, and the long-standing evidence for the automatization of explicit processes (Logan, 1988). Much learning includes controlled mental processing at first that, after practice, becomes automatized and is not necessarily available to introspective access, such as learning to ride a bike. In this sense, the implicit associations may be influenced by, but not accessible to, conscious awareness and intention. There can be little certainty about the identity of implicit associations without knowing how much information has been contributed by non-explicit sources.

Limitations and Directions

While these analyses go further than previous investigations, with broad sampling and inferential tests of both individual preferences and content domains, there are additional factors such as personality and situational variables that need to be examined for a complete understanding of the relationship between implicit and explicit evaluation. This study was correlational and used only one implicit (IAT) and one explicit (thermometer ratings) method of assessment. This creates some uncertainty about whether the observed effects are unique to the method of measurement, or are consistent across measures of the construct.

The IAT is, by design, a relative measure in that it assesses evaluations of one object in comparison to another object. This methodological feature may have particular relevance for the dimensionality moderator that showed that implicit-explicit correspondence was enhanced when the object pairs were perceived to provide opposing endpoints of a single continuum. The relevance of dimensionality could be unique to comparative methods rather than a property moderating implicit-explicit correspondence more generally. Future investigations should investigate multiple measurement methodologies to parse those factors that are methodologically peculiar from those that are generally relevant. This study advanced understanding of the relationship between implicit and explicit preferences by investigating characteristics of individual evaluations, characteristics of content domains, and four distinct moderators at once.

Untested moderators. There may be more moderating influences that are not yet identified or were not part of this investigation (see Hofmann, Gschwendner, Nosek, & Schmidt, in press for a review). There are numerous potential avenues for developing a more comprehensive model. One approach is to identify factors that differentially influence implicit and explicit attitude formation (Rudman, 2004). A second is to consider operative processes such as the notion that implicit measures assess 'narrow' associative relations with the target concepts, while explicit measures enable a reflective, broadening of the evaluative space to consider additional beliefs and attitudes that may be conceptually relevant (Strack & Deutsch, 2004). For example, implicit racial attitudes might reflect the direct associations of racial groups and evaluative concepts. Explicitly, however, the respondent might reflect

upon and use more general conceptions of fairness and interest in egalitarianism to generate an evaluation. These and other frameworks may provide additional leverage for clarifying the relationship between automatic and controlled evaluative processes.

Predictive validity as a function of implicit-explicit correspondence. A recent meta-analysis of 86 independent samples found the IAT to have predictive validity across a variety of content domains (Poehlman et al., 2004). An interesting issue to consider for future research is whether the magnitude of implicit-explicit correspondence, or the specific moderators identified here, can clarify when implicit or explicit evaluations will be best predictive of perception, judgment, or action. For example, Poehlman and colleagues observed that implicit evaluations showed better predictive validity than explicit evaluations in investigations of social groups, whereas self-report showed better predictive validity in other domains – such as consumer preferences. They speculated that the tendency for social group attitudes to elicit greater self-presentation concern compared to domains like consumer preferences might account for the differential predictive validity. Using another moderator for illustration, increasing evaluative strength, or reducing attitude ambivalence, is known to enhance the predictive validity of evaluations suggesting that implicit and explicit evaluations will both have stronger predictive validity when they are correspondent compared to when they are discordant. The present data provide a basis for future investigations of the role of implicit-explicit correspondence in predictive validity of implicit and explicit evaluations.

Conclusion

Human behavior is influenced by intentional and unintentional causes. Evaluations derive from routes intended following introspection, and routes unintended following automatic activation. The variable relationship between these distinct paths of evaluation is fascinating because it distinguishes what exists in our minds from what we experience as our minds. The relationship between implicit and explicit evaluations appears to be, in part, a function of the effortful presentation for personal or social purposes (self-presentation), the vigor of the evaluations (strength), the extent to which evaluations are represented with a simple, bipolar structure (dimensionality), and the extent to which one's evaluations

are perceived as distinct from normative responses (distinctiveness). Knowing this provides an opportunity to attempt some control over those parts of the mind that proceed without us.

References

- Abelson, R.P. (1988). Conviction. American Psychologist, *43*, 267-275.
- Abelson, R. P., & Prentice, D. A. (1989). Beliefs as possessions – a functionalist perspective. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude Structure and Function (pp. 361-381). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Arkes, H. R., & Tetlock, P. E. (2004). Attributions of implicit prejudice, or “would Jesse Jackson ‘fail’ the Implicit Association Test?” Psychological Inquiry, *15*(4), 257-278.
- Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), The nature of remembering: Essays in honor of Robert G. Crowder (pp. 117-150). Washington D.C.: American Psychological Association.
- Banaji, M.R., & Hardin, C.D. (1996). Automatic stereotyping. Psychological Science, *7*, 136-141.
- Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In Wyer, R.S. Jr., & Srull, T.K. (Eds) Handbook of social cognition, Vol. 1: Basic processes; Vol. 2: Applications (2nd Ed.) (pp. 1-40). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology, *51*(6), 1173-1182.
- Bassili, J.N. (1996). Meta-judgmental versus operative indexes of psychological attributes: The case of measures of attitude strength. Journal of Personality and Social Psychology, *71*(4), 637-653.
- Bem, D.J. (1972). Self-perception theory. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology Vol. 6, (pp. 1-62). New York: Academic Press.
- Bizer, G.Y., & Krosnick, J.A. (2001). Exploring the structure of strength-related attitude features: The relation between attitude importance and attitude accessibility. Journal of Personality and Social Psychology, *81*(4), 566-586.

- Blair, I.V., & Banaji, M.R. (1996). Automatic and controlled processes in stereotype priming. Journal of Personality and Social Psychology, *70*, 1142-1163.
- Blanton, H., & Christie, C. (2003). Deviance regulation: A theory of action and identity. Review of General Psychology, *7(2)*, 115-149.
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of self-esteem: The blind men and the elephant revisited? Journal of Personality and Social Psychology, *79*, 631-643.
- Cantril, H. (1944). Gauging Public Opinion. Princeton, NJ: Princeton University Press.
- Carver, C. S. (in press). Impulse and Constraint: Perspectives from Personality Psychology, Convergence with Theory in Other Areas, and Potential for Integration. Personality and Social Psychology Review.
- Chen, M., & Bargh, J.A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. Personality and Social Psychology Bulletin, *25(2)*, 215-224.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. Psychological Science, *12*, 163-170.
- Cunningham, W. A., Johnson, M. K., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2003). Neural Components of Social Evaluation. Journal of Personality and Social Psychology, *85*, 639-649.
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. Psychological Science.
- De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. Journal of Experimental Social Psychology, *37*, 443-451.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, *56(1)*, 5-18.

- Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. (2002). The regulation of explicit and implicit racial bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*(5), 835-848.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, *33*(5), 510-540.
- Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In Petty, R. E. & Krosnick, J. A. (Eds) *Attitude strength: Antecedents and consequences*. Ohio State University series on attitudes and persuasion, Vol. 4 (pp. 247-282). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, *18*, 339-357.
- 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.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*, 1013-1027.
- Fazio, R. H. & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, *54*, 297-327.
- Feroni, F., & Mayr, U. (2005). The power of a story: New, automatic associations from a single reading of a short scenario. *Psychonomic Bulletin and Review*, *12*(1), 139-144.
- Gauthier, I., Skudlarski, P., Gore, J.C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191-197.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27.
- Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem

- and self-concept. Journal of Personality and Social Psychology, 79(6), 1022-1038.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. Journal of Personality and Social Psychology, 74(6), 1464-1480.
- Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. Zeitschrift für Experimentelle Psychologie, 48(2), 85-93.
- 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.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2004). A meta-analysis on the correlations between the Implicit Association Test and explicit self-report measures. Unpublished manuscript. University of Trier: Germany.
- Hofmann, W., Gschwendner, T., Nosek, B. A., & Schmitt, M. (in press). What moderates implicit-explicit consistency? European Review of Social Psychology.
- Judd, C. M., & Kulik, J. A. (1980). Schematic effects of social attitudes on information processing and recall. Journal of Personality and Social Psychology, 38, 569-578.
- Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. Journal of Personality and Social Psychology, 81(5), 774-788.
- Karpinski, A., Steinman, R. B., & Hilton, J. L. (2005). Attitude importance as a moderator of the relationship between implicit and explicit attitude measures. Personality and Social Psychology Bulletin, 31(7), 949-962.
- Kihlstrom, J. F. (2004). Implicit methods in social psychology. In C. Sansone, C.C. Morf, & A.T. Panter (Eds.), Handbook of Methods in Social Psychology (pp. 195-212). Thousand Oaks, CA: Sage.
- Kim, D. Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). Social Psychology Quarterly, 66, 83-96.

- Krosnick, J.A., Boninger, D.S., Chuang, Y.C., Berent, M.K., & Carnot, C.G. (1993). Attitude strength: One construct or many related constructs? Journal of Personality and Social Psychology, *65*(6), 1132-1151.
- Krosnick, J.A., & Petty, R.E. (1995). Attitude strength: An overview. In R.E. Petty & J.A. Krosnick (Eds.), Attitude Strength: Antecedents and Consequences (pp. 1-24). Hillsdale, NJ: Erlbaum.
- Logan, G. D. (1988). Toward an instance theory of automatization. Psychological Review, *95*(4), 492-527.
- Moskowitz, G. B., Gollwitzer, P. M., Wasel, W., & Schaal, B. (1999). Preconscious control of stereotype activation through chronic egalitarian goals. Journal of Personality and Social Psychology, *77*(1), 167-184.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. Psychological Review, *84*(3), 231-259.
- Nosek, B. A. & Banaji, M. R. (2002). (At least) two factors moderate the relationship between implicit and explicit attitudes. In R.K. Ohme & M. Jarymowicz (Eds.), Natura Automatyzmow (pp. 49-56), Warszawa: WIP PAN & SWPS.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. Group Dynamics, *6*, 101-115.
- 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.
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Methodological Issues. Personality and Social Psychology Bulletin, *31*(2), 166-180.
- Nosek, B. A., & Hansen, J. (2005). The associations in our heads belong to us: Attitudes and knowledge in implicit cognition. Unpublished manuscript. University of Virginia: Charlottesville, VA.
- Nosek, B. A., & Smyth, F. L. (2005). Implicit and explicit attitudes are related, but distinct constructs. Unpublished manuscript. University of Virginia: Charlottesville, VA.

- Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extra-personal associations on the Implicit Association Test: Personalizing the IAT. Journal of Personality and Social Psychology, 86, 653-667.
- Paulhus, D. L. (1984). Two-component models of social desirable responding. Journal of Personality and Social Psychology, 46, 598-609.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, & L.S. Wrightsman (Eds.), Measures of Personality and Social Psychological Attitudes (pp. 17-59). San Diego, CA: Academic Press.
- Petty, R.E., & Krosnick, J.A. (1995). Attitude Strength: Antecedents and Consequences. Hillsdale, NJ: Erlbaum.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2000). Performance on indirect measures of race evaluation predicts amygdala activation. Journal of Cognitive Neuroscience, 12(5), 729-738.
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. Journal of Personality and Social Psychology, 75(3), 811-832.
- Poehlman, T. A., Uhlmann, E., Greenwald, A. G., & Banaji, M. R. (2004). Understanding and Using the Implicit Association Test: III. Meta-analysis of Predictive Validity. Unpublished Manuscript. Yale University: New Haven, CT.
- Pratkanis, A.R. (1989). The cognitive representation of attitudes. In A. R. Pratkanis, S. J. Breckler, & A. G. Greenwald (Eds.), Attitude structure and function (pp. 71-98). Hillsdale, NJ: Erlbaum.
- Project Implicit. (2002). A virtual laboratory for research in Implicit Social Cognition. Version 1.5. <http://implicit.harvard.edu/>.
- Raden, D. (1985). Strength-related attitude dimensions. Social Psychology Quarterly, 48, 312-330.
- Rudman, L. A. (2004). Sources of implicit attitudes. Current Directions in Psychological Science, 13(2), 79-82.

- Schuette, R. A., & Fazio, R. H. (1995). Attitude accessibility and motivation as determinants of biased processing: A test of the MODE model. Personality & Social Psychology Bulletin, 21(7), 704-710.
- Smith, E.R., & DeCoster, J. (1999). Associative and rule-based processing: A connectionist interpretation of dual-process models. In S. Chaiken & Y. Trope (Eds.), Dual-Process Theories in Social Psychology (pp. 323-338). New York: Guilford.
- Snijders, T.A.B., & Bosker, R.J. (1999). Multilevel Analysis. London, England: Sage.
- Steffens, M. C., & Plewe, I. (2001). Items' cross-category associations as a confounding factor in the Implicit Association Test. Zeitschrift für Experimentelle Psychologie, 48, 123-134.
- Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. Personality and Social Psychology Review, 8(3), 220-247.
- Verplanken, B. (1991). Persuasive communication of risk information: A test of cue versus message processing effects in a field experiment. Personality and Social Psychology Bulletin, 17, 188-193.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. Psychological Review, 107(1), 101-126.
- Yovel, G., & Kanwisher, N. (2004). Face perception: Domain specific, not process specific. Neuron, 44, 889-898.

Footnotes

¹ Thanks to N. Sriram for conducting the simulations that lead to this insight.

² This moderator in particular may be especially suited for understanding comparative evaluations, the focus of this paper, and less so for ‘absolute’ assessments of evaluation. This possibility is covered in the discussion.

³ Note that this is not suggesting that automatic evaluations are therefore more ‘real’ or ‘true’ than explicit reports. It only suggests that explicit evaluations, but not automatic evaluations, are subject to distortion by judgment processes about how to report one’s evaluation. With all response distortions removed, implicit and explicit evaluations can still differ, and both evaluations could still be ‘true.’

⁴ There are examples in the literature proposing that features inherent to content domains are important for understanding cognitive processes. In research on perception, for example, some object classes (e.g., faces) are proposed to engage unique neural processes. Interestingly, there is debate over the appropriate level of analysis – whether the neural area is sensitive to a specific object domain (faces), or to visual expertise, irrespective of domain (Gauthier, Skudlarski, Gore, & Anderson, 2000; Yovel & Kanwisher, 2004).

⁵ Minor edits to phrasing were made as needed to ensure clarity or appropriate grammar depending on the target objects for all explicit measures.

⁶ A fourth measure of evaluative strength, stability (Raden, 1985; “How much do you expect your attitude toward <object> to change over time?”), was included. However, stability was weakly related to the other strength measures (r ’s range = -.08 to .12), and, if anything, showed somewhat stronger correspondence with self-presentation factors (r ’s range = .08 - .23). It was not included in this initial test.

⁷ Of the remaining modest correlations, the most notable were distinctiveness showing positive relations with evaluative strength and distinctiveness. The first reveals the conceptual connection between strength and distinctiveness – evaluations perceived to differ from the norm tend to be somewhat stronger than evaluations perceived to be normative. Also, the positive relation between distinctiveness and dimensionality reveals that bipolar evaluations tend to elicit somewhat greater perceptions of distinctiveness than unipolar evaluations. This is easy to comprehend because strongly bipolar evaluations will elicit polarized preferences that will necessarily result in evaluations that are somewhat more distinct from the norm.

⁸ Though none were hypothesized, it is possible that the moderating variables interact in predicting implicit-explicit correspondence. Analyses of interactions among the four moderators showed that none were qualified by interactions *between persons*. Only one notable effect emerged *between objects* showing that strength and dimensionality combined such that strong bipolar evaluations elicited especially high implicit-explicit correspondence, but its small effect was inconsequential for model fit, $\Delta\chi^2(1) = 0$. The lack of interactive effects between persons suggests that the moderating effects of these variables are relatively independent of one another.

Appendix. Category labels and stimulus items for the 57 object pairs.

Category A	Stimulus Items	Category B	Stimulus Items
Abstaining	Sober, Refrain, Dry Out, Abstain	Drinking	Alcohol, Drunk, Intoxicated, Inebriated
American	3 pictures of American Landmarks, and Detroit, Seattle, Boston	Canadian	3 pictures of Canadian Landmarks, and Montreal, Toronto, Vancouver
American Places	6 pictures of American Landmarks	Foreign Places	6 pictures of Foreign Landmarks
Approaching	Toward, Advance, Forward, Closer, Near	Avoiding	Away, Withdraw, Back, Retreat, Recoil
Books	War and Peace, Iliad, Paradise Lost, Great Expectations, Moby Dick, Ulysses, Great Gatsby, Beloved, Huckleberry Finn	Television	Friends, Survivor, West Wing, Law and Orger, E.R., The Simpsons, Will and Grace, 60 Minutes, Home Improvement
California	San Francisco, Los Angeles, West Coast, Hollywood, Golden Gate	New York	New York City, Big Apple, East Coast, Buffalo, Statue of Liberty
Cats	4 pictures of Cats	Dogs	4 pictures of Dogs
Christian	Church, Minister, Priest, Gospels, Christmas, Christian, Easter, Sunday	Jewish	Synagogue, Rabbi, Torah, Talmud, Hanukkah, Jewish, Yom Kippur, Sabbath
Classical	Beethoven, Mozart, Tchaikovsky, Bach, Vivaldi, Chopin, Stravinsky, Handel	Hip Hop	2Pac, Eminem, DMX, Beastie Boys, Dr. Dre, Snoop Dogg, Jay-z, Wu-Tang Clan
Coke	6 pictures of Coke products/labels	Pepsi	6 pictures of Pepsi products/labels
Cold	Cool, Freeze, Frozen, Ice, Chill	Hot	Warm, Boil, Heat, Steam, Burn
Conforming	Follow, Obey, Yield, Comply, Abide	Rebellious	Question, Challenge, Defy, Disobey, Resist
Creationism	God, Bible, Religion, Creator, Created, Six Days	Evolution	Darwin, Origin of Species, Science, Natural Selection, Eons, Evolved
Democrats	Liberal, Left Wing, Bill Clinton, Donkey, Tom Daschle, Democrat	Republicans	Conservative, Right Wing, George Bush, Elephant, Dick Cheney, Republican
Education	Educate, Schools, Teachers, Books, Teach, Buses, Students	Defense	Military, Soldiers, Combat, Army, Air Force, Navy, Marines
Emotions	Feeling, Impulsive, Spontaneous, Passionate, Volatile, Sensitive, Impassioned	Reason	Sensible, Prudent, Logical, Reasoning, Prudent, Practical, Pragmatic
Euro Americans	3 pictures of Euro American female faces, and 3 pictures of Euro American male faces	African Americans	3 pictures of African American female faces, and 3 pictures of African American male faces
Exercising	Lift Weights, Run, Jump, Aerobics, Crosstraining, Workout	Relaxing	Sit, Rest, Watch TV, Recline, Snooze, Lie Down
Family	Home, Household, Children, Domestic, Kitchen	Career	Work, Business, Job, Profession, Office
Females	Woman, Female, Girl, Aunt, Grandma, Wife, Mother, Daughter	Males	Man, Male, Boy, Grandpa, Uncle, Husband, Father, Son
Feminism	Gloria Steinem, Liberal, Women's Rights, N.O.W.	Traditional Values	Phyllis Schlafly, Conservative, Family Values, Eagle Forum
Flexible	Shifting, New, Different, Variable, Changing, Novelty, Fluctuate	Stable	Same, Familiar, Unchanging, Steady, Fixed, Enduring, Permanent
Freedom	Free, Unregulated, Liberty, Independent	Security	Safe, Secure, Controlled, Protected
Future	Tomorrow, Next Year, After, Ahead, Upcoming, Coming	Past	Behind, Backward, Yesterday, Last Week, Years Behind
Gore	2 pictures of Al Gore, and Al Gore, Gore	Bush	2 pictures of George Bush, and George Bush, Bush
Gun Rights	Gun Rights, NRA, Allow Guns, Bear Arms	Gun Control	Gun Control, Background Checks, Restrict Guns, Waiting Period
Imprisonment	Jail Cell, Locked Away, Rehabilitation, Imprisoned	Capital Punishment	Electric Chair, Lethal Injection, Gas Chamber, Capital Punishment
Jay Leno	Tonight Show, NBC, Leno, Jay Leno	David Letterman	Late Night, CBS, Letterman, David Letterman
Jews	Torah, Moses, Temple, Judaism, Israel	Muslims	Koran, Mohammed, Mosque, Islam, Palestine
Leaders	CEO, President, Boss, Manage, Senator	Helpers	Assistant, Worker, Attendant, Employee, Aide
Letters	ABC, DEF, GHI, BCDE, FGHI, ABCDEF, DEFGHI	Numbers	123, 456, 789, 2345, 6789, 123456, 456789
Liberals	Left Wing, Tom Daschle, Democrats, Al Gore, Bill Clinton	Conservatives	Right Wing, George Bush, Republicans, Dick Cheney, Ronald Reagan
Management	Supervisors, Boss, Administration, Employer, Managers	Labor	Union, Workers, Employees, Staff, Subordinates
Married	Together, Joined, Couple, Wedded, Committed, Hitched	Single	Independent, Autonomous, Unmarried, Unattached, Individual, Solitary
Meg Ryan	6 pictures of Meg Ryan	Julia Roberts	6 pictures of Julia Roberts
Microsoft	6 pictures of Microsoft products/labels	Apple	6 pictures of Apple products/labels
Nerds	Studious, Intelligent, Smart, Study, Calculator, Brains	Jocks	Sports, Strong, Athletic, Exercise, Football, Muscles
Northerners	New York, Maine, Vermont, New Hampshire, Connecticut, Rhode Island, Massachusetts	Southerners	Alabama, Georgia, Mississippi, Louisiana, Florida, Tennessee, South Carolina
Pro-Choice	Pro-Choice, Right to Choose, Planned Parenthood	Pro-Life	Pro-Life, Right to Life, Christian Coalition
Public	Communal, Common, Shared, Exposed, Apparent, Visible, Known	Private	Secluded, Personal, Hidden, Confidential, Covert, Concealed, Secret
Religion	Religious, Theist, Churchgoer	Atheism	Agnostic, Atheist, Skeptic
Rich People	Rich, Wealthy, Affluent, Prosperous, Well Off, Loaded, Fortune, Lucrative	Poor People	Poor, Poverty, Destitute, Needy, Impoverished, Broke, Bankrupt, Penniless
Short People	Small, Tiny, Little, Slight, Petite	Tall People	Lanky, Big, Large, Gigantic, Towering
Simple	Easy, Elementary, Straightforward, Effortless	Difficult	Challenging, Puzzling, Complicated, Baffling
Skirts	5 pictures of Skirts	Pants	5 pictures of Pants
Social Programs	Medicare, Social Security, Build Schools, Welfare	Tax Reductions	Lower Taxes, Tax Credits, Tax Relief, Tax Breaks
Straight People	5 pictures of Heterosexuals and Opposite Sex icon pairs	Gay People	5 pictures of Homosexuals and Same Sex icon pairs
Summer	June, Hot, July, August	Winter	December, Cold, January, February
Tea	Darjeeling, Chamomile, Earl Grey, English Breakfast, Lemon, Sleepytime, Herbal	Coffee	Cappuccino, Espresso, Mocha, Latte, Colombian, Dark-Roasted, French Roast
Teen Pop	N Sync, Mariah Carey, Backstreet Boys, Mandy Moore, Britney Spears, Brandy, Christina Aguilera	Jazz	Ella Fitzgerald, Billie Holliday, Louis Armstrong, Duke Ellington, Miles Davis, John Coltrane, Diana Krall
Thin People	5 pictures of Thin People	Fat People	5 pictures of Fat People
Tom Cruise	6 pictures of Tom Cruise	Denzel Washington	6 pictures of Denzel Washington
USA	Washington D.C., New York, Mt. Rushmore, Chrysler, Ford, America	Japan	Tokyo, Kyoto, Mt. Fuji, Mitsubishi, Honda, Asia
Vegetables	7 pictures of Vegetables	Meat	7 pictures of Meat
Whites	Smith, Jones, Adams, Harris, Miller, Perry, Ward	Asians	Chang, Hwang, Wong, Chen, Nguyen, Fujimoto, Lee
Yankees	7 pictures of Yankees Players/Symbols	Diamondbacks	7 pictures of Diamond Backs Players/Symbols
Young People	6 pictures of Young People	Old People	6 pictures of Old People

Author's note

Brian A. Nosek, University of Virginia.

This research was supported by a grant from the National Institute of Mental Health (MH-R01 MH68447-01). I thank Mary Hunt, Roger Despres, Tim Redding, Patrick Ohiomoba, and Frank Steen for technical support and Mahzarin Banaji, Jeff Hansen, Jim Sherman, Kate Ranganath, Colin Smith, Fred Smyth, N. Sriram, Bethany Teachman and anonymous reviewers for their comments.

Correspondence concerning this article should be addressed to Brian Nosek, Department of Psychology, University of Virginia, 102 Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22911. Electronic mail may be sent to nosek@virginia.edu. Related information can be found at <http://briannosek.com>.

Table 1. Implicit and Explicit Preferences, Implicit-Explicit Correlations, and Means for Four Proposed

Moderators for 57 Object Pairs

Category A	Category B	N	Implicit	Explicit	Implicit-Explicit	Centered Moderator Means			
			Mean (SD)	Mean (SD)	Correlation	SP	ST	DIM	DIS
Pro-Choice	Pro-Life	212	.20 (.59)	1.91 (4.90)	.70*	.12	.88	2.39	2.17
Gore	Bush	188	.13 (.62)	1.45 (3.94)	.66*	-.83	-.74	2.25	.82
Creationism	Evolution	201	.25 (.64)	-1.49 (4.70)	.60*	-.28	.24	2.35	1.85
Feminism	Traditional Values	195	.31 (.61)	.86 (3.24)	.59*	.28	.71	1.37	.79
Democrats	Republicans	164	.23 (.70)	1.62 (3.72)	.59*	-.33	.04	2.33	1.32
Religion	Atheism	180	.58 (.63)	1.25 (4.24)	.58*	.52	.52	2.26	1.88
Gun Rights	Gun Control	185	.08 (.60)	-2.31 (4.42)	.58*	-.59	.10	2.21	1.27
Liberals	Conservatives	186	.34 (.69)	2.02 (3.56)	.56*	.05	.15	1.96	.92
Social Programs	Tax Reductions	168	.47 (.55)	.93 (3.42)	.54*	-.13	.31	.77	-.11
Coke	Pepsi	223	.22 (.59)	.88 (3.30)	.54*	-1.78	-1.35	1.15	-.94
Northerners	Southerners	177	.16 (.65)	.40 (2.40)	.52*	.28	-.47	.09	-1.01
Straight People	Gay People	144	.43 (.54)	1.11 (2.93)	.50*	.62	.46	-1.03	.76
Jews	Muslims	217	.43 (.51)	.91 (1.95)	.46*	.91	-.37	.36	-.92
Vegetables	Meat	199	.41 (.64)	.36 (3.21)	.46*	-.78	-.27	-.70	.25
Teen Pop	Jazz	212	.22 (.58)	-1.33 (3.33)	.44*	-.87	-1.08	.73	-.22
Books	Television	203	.10 (.60)	2.41 (2.67)	.43*	-.46	.39	-.28	2.12
Abstaining	Drinking	211	.32 (.59)	.62 (3.84)	.42*	.31	.14	1.31	.45
Exercising	Relaxing	221	.10 (.75)	-1.28 (2.85)	.42*	.41	.82	-.70	-.15
Nerds	Jocks	203	.36 (.49)	1.38 (2.31)	.41*	.30	-.66	.95	.74
Yankees	Diamondbacks	171	.28 (.54)	.20 (1.94)	.41*	-1.28	-2.75	.53	-1.14
Christian	Jewish	217	.57 (.53)	.04 (2.17)	.40*	.82	.21	-.96	-.31
American	Canadian	263	.35 (.53)	.23 (2.40)	.40*	.25	.34	-1.21	-.57
Classical	Hip Hop	204	.37 (.59)	.82 (3.13)	.39*	-.45	-.79	.47	.32
Cats	Dogs	227	.21 (.56)	-.80 (3.04)	.39*	-.45	.12	-.31	.20
Tea	Coffee	214	.09 (.60)	.19 (3.19)	.38*	-1.41	-1.10	-.88	-.16
Summer	Winter	223	.63 (.60)	1.55 (3.52)	.37*	-1.12	.17	.46	.28
American Places	Foreign Places	178	.54 (.54)	.62 (2.06)	.37*	.11	.25	-1.55	-.53
Emotions	Reason	158	.44 (.52)	-.16 (2.30)	.37*	1.35	1.51	-.09	-.70
Management	Labor	166	.03 (.53)	-.74 (2.58)	.37*	.69	.33	1.00	-.45
USA	Japan	211	.57 (.48)	1.31 (2.46)	.36*	-.08	.08	-1.33	-.73
Leaders	Helpers	236	.03 (.57)	-1.00 (2.15)	.36*	.61	.76	-.33	-.19
Jay Leno	David Letterman	178	.01 (.48)	.00 (2.52)	.36*	-1.44	-2.37	-.23	-1.12
Conforming	Rebellious	194	.64 (.62)	-1.13 (3.09)	.35*	.44	.59	2.10	2.71
Meg Ryan	Julia Roberts	211	.05 (.50)	.02 (1.64)	.35*	-1.30	-2.07	-1.78	-1.76
Education	Defense	190	.79 (.46)	2.27 (2.50)	.33*	.22	1.09	-1.11	-.02
Euro Americans	African Americans	211	.41 (.53)	.41 (1.71)	.33*	1.02	.30	-1.22	-1.04
California	New York	225	.21 (.64)	.01 (2.35)	.33*	-.77	-.95	-.84	-.84
Tom Cruise	Denzel Washington	212	.16 (.51)	-1.04 (2.13)	.33*	-1.10	-2.03	-1.92	-1.11
Microsoft	Apple	177	.04 (.58)	.14 (2.71)	.33*	-1.49	-1.45	.95	-1.08
Flexible	Stable	183	.03 (.64)	.38 (1.92)	.31*	.73	.87	-.46	.42
Imprisonment	Capital Punishment	224	.25 (.48)	1.64 (2.96)	.29*	-.76	.27	-.17	-.08
Young People	Old People	218	.56 (.44)	.14 (2.07)	.26*	.80	.64	-.88	-.49
Simple	Difficult	192	.93 (.47)	.61 (2.94)	.24*	.81	.65	-.04	1.18
Letters	Numbers	205	.41 (.55)	.80 (2.47)	.22*	-.43	-.14	-.73	-.86
Freedom	Security	204	.43 (.47)	1.35 (1.92)	.21*	.59	1.27	-.85	-1.01
Public	Private	175	.33 (.52)	-1.50 (2.05)	.21*	.95	.78	-.91	-.08
Married	Single	217	.28 (.60)	.59 (3.34)	.21*	.63	.53	.44	.89
Skirts	Pants	217	.33 (.52)	-.86 (2.47)	.18*	-.79	-.84	-1.40	-1.01
Family	Career	214	.66 (.49)	1.07 (2.32)	.17*	.97	1.97	-1.99	-.60
Cold	Hot	191	.05 (.85)	-1.26 (3.12)	.17*	-.50	-.04	1.56	-.20
Rich People	Poor People	211	1.00 (.45)	-.44 (2.60)	.16*	.74	.14	1.02	.02
Whites	Asians	239	.36 (.48)	.19 (1.39)	.16*	.60	.12	-1.61	-.88
Short People	Tall People	197	.10 (.70)	-.34 (1.52)	.14	.40	-.87	-1.59	-1.21
Approaching	Avoiding	171	1.11 (.38)	2.18 (3.14)	.13	.83	.17	1.18	1.23
Future	Past	202	.85 (.40)	1.07 (2.16)	.12	.82	1.39	-1.35	-.66
Thin People	Fat People	236	.42 (.50)	.82 (1.93)	.10	.82	-.20	.50	.64
Females	Males	250	.61 (.51)	.40 (2.46)	-.05	.72	1.13	-1.49	-.65

Note. Positive implicit and explicit means indicate a preference for the Category A object (1st column) compared to Category B (2nd column); negative values indicate that the Category B object was preferred. Italicized means for implicit and explicit evaluations indicate that the preference for one object compared to the other was significantly different from zero ($p < .05$). Object pairs in bold reflect those for which one object was significantly preferred implicitly, but the opposing object was significantly preferred explicitly. Asterisks (*) indicate implicit-explicit correlations that were significantly different from zero ($p < .05$). Means for the moderators are centered on grand mean of the moderator: self-presentation (SP: Mean = 3.98, SD = 1.43, Range = 1-9), strength (ST: Mean = 4.89, SD = 1.64, Range = 1-9), dimensionality (DIM: Mean = 4.38, SD = 2.38, Range = 1-9), and distinctiveness (DIS: Mean = 1.82, SD = 1.40, Range = 0-8).

Table 2

Multilevel Models of the Relationship between Implicit and Explicit Evaluations (Stage 1)

Parameter/Statistic	Models			
	M ₀	M ₁	M ₂	
<u>Fixed Effects (Persons)</u>				
Intercept	.38 (3)	-.32 (2.4)	-.25 (2)	
IAT		1.92 (43)	1.87 (12)	
<u>Random Effects</u>				
Intercept Variance	1.09 (5)	.98 (5)	.89 (5)	
IAT			1.18 (5)	
cov(Intercept, IAT)				
<u>Goodness-of-Fit</u>				
-2 Log Likelihood	<i>-2LL</i>	55270	53535	53004
Level 2 Degrees of Freedom	<i>df</i>	56	56	56
change -2LL	<i>LRT</i>		1735	2266
change in df from M ₀			1	2
Intraclass correlation	.12			
Total variance of Y (Explicit)	9.22	7.92	7.92	
Est. var. explained at Level 1		.14	.14	
Est. var. explained at Level 2		.12	.12	

Note. The dependent variable for all models was explicit evaluation. All models were fit with SAS PROC MIXED with $N = 11,166$ and Q (Object Pairs) = 57; Maximum likelihood parameter estimates are listed with approximate t values for the parameters in parentheses. Boldface coefficients are significant at $p < .05$. IAT and Explicit evaluations have rational zero points indicating relative indifference; positive values indicate preference for object implicitly preferred on average. The critical comparative model index is the change in log-likelihood (-2LL; a chi-square value) on the change in degrees of freedom. In the Table, Models 1 and 2 are compared to Model 0. R -sq values were estimated following recommendations of Snijders & Bosker (1999) by first removing random effects and estimating R1 and R2 with only fixed effects in the model. Level 1 refers to between-person effects; Level 2 refers to between-objects effects.

Table 3

Best-Fitting Multilevel Models for Hypothesized Moderators of Implicit-Explicit Correspondence ExaminedIndividually (Stage 2)

Parameter/Statistic	Models				
	SP ₂	ST ₃	DIM ₂	DIS ₃	
<u>Fixed Effects (Persons)</u>					
Intercept	-.25 (2)	-.24 (2)	-.27 (2)	-.16 (1)	
IAT	1.87 (13)	1.84 (13)	1.81 (16)	1.7 (12)	
Moderator (SP, ST, DIM, DIS)	.06 (2.5)	-.02 (.6)	-.03 (2)	-.16 (2)	
IAT * Moderator	-.11 (3)	.30 (6)	.20 (8)	.52 (6)	
<u>Fixed Effects (Objects)</u>					
Mean Moderator	-.02 (0.1)		-.10 (1)	-.03 (.1)	
IAT * Mean Moderator	-.39 (2)		.55 (6)	1.1 (3)	
<u>Random Effects</u>					
Intercept Variance	.91 (5)	.92 (5)	.99 (5)	.90 (4)	
IAT	1.10 (5)	1.01 (5)	.54 (4)	.79 (4)	
Moderator		.05 (4)		.36 (4)	
IAT * Moderator		.09 (3.5)		.16 (2)	
<u>Goodness-of-Fit</u>					
-2 Log Likelihood	<i>-2LL</i>	52796	52580	32361	13990
Level 2 Degrees of Freedom	<i>df</i>	56	56	56	56
change -2LL	<i>LRT</i>	208	424	77	342
change in df from Model M ₂		4	4	4	6
Total variance of Y (Explicit)		7.87	7.87	7.98	7.92
Est. var. explained at Level 1		.15	.15	.15	.17
Est. var. explained at Level 2		.13	.11	.04	.05

Note. The dependent variable for all models was explicit evaluation. All models were fit with SAS PROC MIXED with $N = 11,120$ (SP), 11,131 (ST), 6,781 (DIM), 2,981 (DIS) and Q (Object pairs) = 57; Maximum likelihood parameter estimates are listed with approximate t values for the parameters in parentheses. Boldface coefficients are significant at $p < .05$. IAT and Explicit evaluations have rational zero points indicating relative indifference; positive values indicate preference for object implicitly preferred on average. Self-presentation (SP), strength (ST), dimensionality (DIM), and distinctiveness (DIS) indicators were each centered on its grand mean. Mean self-presentation, strength, dimensionality, and distinctiveness scores were each centered on its group (object pair) grand mean. The critical comparative model index is the change in log-likelihood (-2LL; a chi-square value) on the change in degrees of freedom. For each model presented, the change in log-likelihood is compared with Model 2 from Table 2 refitted for the available data for that moderator. R -sq values were estimated following recommendations of Snijders & Bosker (1999) by first

removing random effects and estimating R1 and R2 with only fixed effects in the model. Level 1 refers to between-person effects; Level 2 refers to between-objects effects. Complete model descriptions for the individual moderators are available at <http://briannosek.com/>.

Table 4

Product-Moment Correlations between Hypothesized Moderators of Implicit-Explicit Correspondence

	Self-Presentation	Evaluative Strength	Dimensionality	Distinctiveness
	N = 11,177	N = 11,187	N = 6,823	N = 2,981
Self-Presentation		.41*	.07*	.02
Evaluative Strength			.06*	.25*
Dimensionality				.19*

Note. * $p < .05$.

Table 5

Multilevel Models Identifying Moderators of the Relationship between Implicit and Explicit Evaluations (Stage 3)

Parameter/Statistic	Models						
	(F) M ₀	(F) M ₁	(F) M ₂	F ₁	F ₂	F ₃	
<u>Fixed Effects (Persons)</u>							
Intercept	.43 (3)	-.30 (2)	-.24 (1.7)	-.13 (.9)	-.11 (.9)	-.13 (1)	
IAT		2.02 (24)	1.97 (12)	1.60 (13)	1.62 (14)	1.62 (13)	
Self-Presentation				.06 (1.4)	.06 (1.4)	.07 (1.7)	
Strength				.02 (.4)	.02 (.4)	.02 (.5)	
Dimensionality				-.02 (-1)	-.03 (1.1)	-.03 (1.3)	
Distinctiveness				-.14 (1.6)	-.15 (1.7)	-.15 (1.6)	
IAT * Self-Presentation				-.21 (3)	-.22 (3)	-.26 (4)	
IAT * Strength				.24 (4)	.24 (4)	.22 (4)	
IAT * Dimensionality				.18 (5)	.40 (5)	.21 (6)	
IAT * Distinctiveness				.38 (5)	.18 (5)	.40 (5)	
<u>Fixed Effects (Objects)</u>							
Mean Self-Presentation				-.08 (.4)	.03 (.3)		
Mean Dimensionality				-.04 (.3)	-.01 (0)		
Mean Distinctiveness				.01 (0)			
IAT * Mean Self-Presentation				-.41 (2)	-.59 (4)		
IAT * Mean Dimensionality				.29 (3)	.35 (4)		
IAT * Mean Distinctiveness				.68 (1.9)			
<u>Random Effects</u>							
Intercept Variance	1.08 (5)	.96 (5)	.85 (4)	.94 (4)	.74 (4)	.73 (4)	
IAT			1.16 (4)	.48 (3)	.37 (3)	.43 (3)	
Strength				.03 (1)			
IAT * Strength				.05 (1)			
Distinctiveness				.35 (4)	.35 (4)	.35 (4)	
IAT * Distinctiveness				.14 (3)	.15 (3)	.15 (3)	
<u>Goodness-of-Fit</u>							
-2 Log Likelihood	<i>-2LL</i>	14922	14414	14307	13915	13907	13908
Level 2 Degrees of Freedom	<i>df</i>	56	56	56	56	56	56
change -2LL	<i>LRT</i>	0	508	615	1007	1015	1014
change in df from Model (F) M ₀	<i>df</i>		1	2	20	16	12
Total variance of Y (Explicit)		9.58	8.13	8.13	7.69	7.61	7.65
Est. var. explained at Level 1			.15	.15	.20	.21	.20
Est. var. explained at Level 2			.11	.11	.07	.14	.13

Note. The dependent variable for all models was explicit evaluation. All models were fit with SAS PROC MIXED with $N = 2,973$ and Q (Objects) = 57; Maximum likelihood parameter estimates are listed with approximate t values for the parameters in parentheses. Boldface coefficients are significant at $p < .05$. IAT and Explicit evaluations have rational zero points indicating relative indifference; positive values indicate preference for object implicitly preferred on average. Self-presentation, strength, dimensionality, and distinctiveness were centered on the grand mean. Mean self-presentation, strength, dimensionality, and distinctiveness scores were centered on the group (object) grand mean. Models (F) M₀, M₁, and M₂ are re-

estimates of the Models presented in Table 2 (for comparison with new models) using the data available for all moderators. The critical comparative model index is the change in log-likelihood (-2LL; a chi-square value) on the change in degrees of freedom. In the Table, both change values are scored compared to the initial model (F) M0. Models are sequential from left to right. Comparison with the adjacent model to the left is reported in the text. R-sq values were estimated following recommendations of Snijders & Bosker (1999) by first removing random effects and estimating R1 and R2 with only fixed effects in the model. Level 1 refers to between-person effects; Level 2 refers to between-objects effects.