Abstract

Moderators of the relationship between implicit and explicit attitudes

Brian Arthur Nosek

2002

Preferences that are products of introspection and endorsed by the respondent (explicit attitudes) can conflict with preferences that are measured indirectly and do not require conscious introspection or endorsement (implicit attitudes). In three studies, two factors are examined that may predict when implicit and explicit attitudes will be associated or dissociated: self-presentation and attitude elaboration. In the first study, evidence that increasing self-presentation demands negatively affected implicit-explicit correspondence was observed through a manipulation of a public and a private context in which the attitude was reported. In the second study, elaborating an attitude for 20 minutes increased implicit-explicit correspondence compared to a control attitude. The third study reports a synthesis of web-based and laboratory tasks that varied in self-presentation and elaboration. Perceived self-presentation and elaboration produced stable differences in implicit-explicit correspondence such that attitude objects with low self-presentation concerns and high elaboration showed the strongest correspondence. These data suggest that existing models cannot sufficiently account for the relationship between implicit and explicit attitudes, and that the relationship between automatic and consciously mediated preferences is both reliable and predictable.

Moderators of the relationship between implicit and explicit attitudes

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

By
Brian Arthur Nosek

Dissertation Director: Mahzarin R. Banaji

September 2002

© 2002 by Brian Arthur Nosek
All rights reserved.

Table of Contents

List of Tables and Figures	Page iv
Acknowledgements	v
Introduction	1
Study 1 – Self-presentation	12
Method	16
Results	23
Discussion	26
Study 2 – Elaboration	28
Method	29
Results	33
Discussion	39
Study 3 – Self-presentation and elaboration	43
Method	45
Results	47
Discussion	49
General Discussion	49
References	61
Attachment A: How can implicit attitudes be malleable and slow to change at the same time?	65
Attachment B: Removing the assumption of homogeneity of variances from the Cohen's d estimate of effect size	67
Appendix A: Labels and stimuli for Implicit Association Tests	70
Appendix B: Elaboration materials for Study 2	71

List of Tables and Figures

Table 1. Two-step hierarchical regressions for four attitude objects: (1) Black/White, (2) Gay/Straight, (3) Science/Humanities, and (4) Republican/Democrat. (* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$) Page 24

Table 2. Internal consistency of elaborated and control, implicit and explicit attitudes Page 37

Table 3. Summary of average implicit-explicit correlations, perceived elaboration, and perceived self-presentation demands for 15 attitude object pairs Page 47

Table 4. A simple taxonomy of attitude measures (rows) and methodological features of measurement (columns). 1 indicates 'Yes' to row question, 0.5 indicates 'Partially', 0 indicates 'No' Page 58

Table 5. Explicit attitudes for five attitude objects (data collected from the Internet) Page 67

Figure 1. A Venn diagram illustrating the hypothesized relationships between implicit attitudes and explicit attitudes assessed either privately or publicly Page 14

Figure 2. Zero-order correlations among implicit and explicit attitudes within and across sessions for elaboration and control conditions. (Study 2) Page 35

Figure 3. Scatter-plot of the relationship between correlations (IE_{corr}) and mean difference relationships (IE_{md}) between implicit and explicit attitudes. Page 54

Acknowledgements

Like any project that bears my name as 'first author' its success is, in large part, due to the extraordinary efforts of research assistants, the generous advice of colleagues, the thoughtful guidance of mentors, and the encouragement of friends and family. I am grateful for the help that I have received in the completion of this project and my graduate education. Several research assistants including Amanda Parsons, Scott Yates, Brian King, Andrew Kamons, Jerry Liang, and Irina Manta recruited subjects, prepared research materials, ran experimental protocols, and brainstormed fixes to errant procedures. The ideas in this dissertation benefitted from many conversations with Bethany Teachman, Kristin Lane, Eric Uhlmann, the rest of the Banaji Lab, and my committee members: David Armor, Peter Salovey, Marcia Johnson, and Geoffrey Cohen. Anything good that comes out of this dissertation, or my career, can be directly attributed to the guidance of my mentor, Mahzarin Banaji. And, I acknowledge the automatic support my family has provided for my pursuit of an academic career. My father did not let me drop Yale from the list of applications, despite the ridiculous $60 application fee. My mother only protested for a moment when I announced in 1994 that I wanted to change my major from computer engineering to psychology (the real protests did not occur until Excite stock hit $90/share). My brother toured institutions with me making the trip East more fun than intimidating. And, by showing genuine interest in it, my sister provided reassurance that the topic of my research was not as geeky as I am. Mostly, I appreciate that whatever challenges I pursue, my family believes in me.

Author's Note

Data for Study 3 are briefly described in a book chapter (Nosek & Banaji, in press). Data from Study 1 were presented at the Society for Personality and Social Psychology Conference in San Antonio, TX on February 2, 2001. This work was supported by grants from the National Institute of Mental Health (MH-57672) and from the National Science Foundation (SBR-9709924) to Mahzarin R. Banaji, and a graduate fellowship from Yale University to Brian Nosek.

Moderators of the relationship between implicit and explicit attitudes

A persistent question in psychology concerns how conscious experience corresponds to the contents and processes of the mind. It is now clear that conscious experience is not a direct reflection of mental operations, and that a variety of mental activities are unavailable to introspection (Nisbett & Wilson, 1977). Recent theorizing extends the apparent dissociation between conscious experience and mental operations to central constructs of social cognition like attitudes resulting in a distinction between evaluations generated intentionally (explicit attitudes), and ones generated without mediation of conscious awareness or control (implicit attitudes). Greenwald and Banaji (1995) identified implicit attitudes as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate favorable or unfavorable feeling, thought, or action toward social objects" (p. 8), distinguishing them from explicit attitudes that are the evaluative products of introspection about an attitude object. A burgeoning literature on implicit measurement of attitudes suggests that they are distinct from explicit attitudes, and that this distinction may have important implications for social perception, judgment, and behavior (Banaji, 2001; Chen & Bargh, 1999; Devine, 1989; Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997; Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald & Banaji, 1995; Greenwald & Nosek, 2001).

Over the past 80 years of attitude research, the overwhelming majority of tools designed to tap attitudes have relied on conscious introspection and have included, perhaps unintentionally, personal endorsement as a functional component of measurement. Interestingly, virtually none of the definitions of the attitude construct

include introspection or endorsement as part of the definition (see Greenwald & Banaji, 1995 for examples). Even so, until the mid-1980s the investigations of attitudes examined evaluative responses that were, by and large, products of introspection and endorsed or acknowledged by responders.

One popular definition of attitude serves the present discussion because of its direct implications for measurement – an attitude is the association between a concept and an evaluation (Fazio et al., 1986). In 1986, Fazio and his colleagues (Fazio et al., 1986) demonstrated that evaluative responses could be measured without introspection, and could reveal preferences that were inconsistent with evaluative responses measured via more traditional means. Fazio's 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 an attitude object (e.g., a Black face or a White face). Attitude objects associated with positive evaluation facilitated responses (i.e., faster response times) to words with positive meaning, whereas attitude objects associated with negative evaluation facilitated responses to words with negative meanings. The results of this and similar research programs revealed that the evaluative associations measured by tasks such as evaluative priming are elicited automatically and are often at odds with respondents' professed attitudes (Dovidio et al., 1997; Fazio et al., 1995; Greenwald, McGhee, & Schwartz, 1998). These demonstrations suggest that a single individual might employ two unique processes for evaluating information in the environment: one that is conscious, intentional, and controlled, and another that is unconscious, unintentional, and automatic.

The purpose of this dissertation is to identify factors that play a role in predicting the relationship between implicit and explicit attitudes. I will argue that implicit and explicit attitudes are neither completely independent nor are they identical assessments of a single construct. The relationship between implicit and explicit attitudes, or (more accurately) the relationship between implicit and explicit measures of attitudes, can be predicted by at least two factors: self-presentation and attitude elaboration.

The recognition that attitudes can exist outside conscious awareness or control has no direct bearing on the validity of explicit measures of attitudes – preferences that are available to conscious awareness. Explicit attitudes are well established as important in guiding person perception, judgment, and action (Eagly & Chaiken, 1998). The existence of implicit preferences suggests that the dependence on tools that require introspection and self-report may unnecessarily limit the attitude construct in ways that constrain its applicability and ignore its complexity. Methods to measure implicit social cognition have broadened the measurement of the attitude construct to better fit its traditional definitions. At the same time, a critical mass of research on implicit and explicit attitudes shows a great deal of confusion about the nature of the relationship between these two modes of evaluation. Are they products of the same mental operations or independent ones? Do they operate in opposition, in tandem, or independently? How are they functionally related? Understanding the nature of the relationship between implicit and explicit attitudes will facilitate the advancement of theory about when and how evaluative associations, of any type, will influence social perception, judgment, and behavior. In this dissertation, I report the beginning of an investigation into the factors that predict when implicit and explicit attitudes will be consistent or divergent, associated or dissociated. At

present, there exist two simple views about the nature of the relationship between implicit and explicit attitudes.

Implicit and explicit attitudes – same or different?

Two views of the relationship between implicit and explicit attitudes could cast them as measures of the same underlying construct, or measures of completely unique constructs. The simple view that implicit and explicit measures tap the same underlying construct presents implicit attitudes as 'true' attitudes, and explicit attitudes as corrupted versions because they can be obscured by self-presentation. That is, because explicit attitudes are mediated by conscious intentions, people can alter their responses in the service of self-serving or interpersonal goals, or in response to social norms. Thus, implicit and explicit attitudes differ only to the extent that explicit attitudes are altered by self-presentation. For example, a person may have a negative attitude toward African Americans but avoid reporting it because such negativity violates social norms of egalitarianism. With explicit measures, a person could alter the attitudinal response so that it appears unbiased. Implicit measures, on the other hand, provide relatively less opportunity for intentional alteration of the response, so biases might still be observed on the implicit measure despite a desire not to reveal them. The assumption, then, is that implicit and explicit attitudes are identical unless conscious intentions alter the explicit response.

A second conceptualization of implicit-explicit (IE) correspondence considers implicit and explicit attitudes to be unique constructs, follows from a body of evidence emphasizing the distinctive contributions of these two modes of evaluation (Devine, 1989; Dovidio et al., 1997; Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler,

2000). In this view, the processes that give rise to implicit and explicit attitudes are thought to involve separate components somewhere in the information-processing stream. It is as if there are two minds in a single brain operating independently. Explicit attitudes are largely a product of our conscious mind and the product of self-reflection, and implicit attitudes are largely a product of the unconscious, which is inaccessible via introspection. As a consequence, implicit and explicit attitudes need not be related at all because they involve different aspects of the learning and production components of the cognitive/affective system.

Both of these views offer relatively simple accounts of the relationship between implicit and explicit attitudes. In the 'same' view, the IE relationship can be completely explained by variation in self-presentation. In the 'separate' view, the relationship, or the lack of one, is 'explained' by postulating separate systems.

Available data regarding the relationship between implicit and explicit attitudes

Much of the early data comparing implicit and explicit attitudes suggested that implicit and explicit measures elicit quite different evaluative responses (e.g., Banaji & Greenwald, 1995; Banaji & Hardin, 1996; Blair & Banaji, 1996; Devine, 1989; Dovidio et al., 1997; Fazio et al., 1995). More recently, a variety of demonstrations indicate that implicit and explicit attitudes are, in some cases, related (Blair, 2001; Cunningham, Preacher, & Banaji, 2001; Greenwald & Farnham, 2000; Lepore & Brown, 1997; Nosek, Banaji, & Greenwald, 2002b; Nosek, Banaji, & Greenwald, 2002c). For example, Nosek et al. (2002b) showed that variation in implicit-explicit correspondence depends, in part, on the attitude object assessed. Across nine measures, implicit-explicit correspondence was observed ranging from a low zero-order correlation of $r = .08$ (attitudes toward

young relative to old) to a high of r = .52 (attitudes toward George Bush relative to Al Gore).[1]

While not completely independent, the available evidence also suggests that implicit and explicit attitudes are not just interchangeable assessments of the same construct. Analyses of the structural properties of implicit and explicit attitudes suggest that they have a common component but each retain unique elements (Cunningham et al., 2001; Greenwald & Farnham, 2000). That is, although a strong correlation between latent constructs measuring implicit and explicit attitudes was found, a single-factor solution did not fit the data. That research indicates that the correlation notwithstanding, the two constructs capture unique aspects of the attitude construct. Further, the unique elements of implicit and explicit attitudes are not simply attributable to differences in their modes of measurement. The unique components of implicit and explicit attitudes can have important predictive qualities. For example, Nosek et al. (2002c) showed that in a simultaneous regression (removing the shared component of implicit and explicit math/arts attitudes) residuals for both attitude types were related to SAT performance.

The variable presence of association and dissociation between implicit and explicit attitudes introduces the possibility that one or more factors could moderate the relationship between these two types of attitudes. Evidence described in this paper suggests that both the 'separate' and 'same' views are insufficient explanations of the relationship between implicit and explicit attitudes. Specifically, I argue that there are at

[1] Note that the .52 correlation for Bush relative to Gore was actually an underestimation of the magnitude of that relationship because of a confounding methodological factor. More recent data suggests that implicit-explicit correspondence for Bush/Gore attitudes (among Internet respondents in 2000 and 2001) actually exceeds .6 (zero-order) and even reaches .8 with an improved scoring algorithm and correction for attenuation (see Study 3 in this dissertation and Greenwald, Nosek, & Banaji, 2002).

least two moderators of implicit-explicit correspondence – (1) self-presentation, or the altering of an evaluative response due to concerns of presentation to oneself or to others[2] and (2) elaboration, or the amount of mental processing devoted to evaluating the attitude object. If true, then a new model can be proposed that will better explain the relationship between implicit and explicit attitudes.

The studies will lead to a proposal in the General Discussion that research on implicit-explicit correspondence should focus on the components of attitude measurement that elicit different aspects of the attitude construct. Future research in this area will manipulate particular aspects of the attitude measurement situation (e.g., awareness of the target of measurement) to (a) observe the effect on the elicited attitude, and (b) test how the conditions of attitude measurement affect the predictive utility of the resulting attitudes (e.g., under what conditions do attitudes measured without awareness of the target have predictive utility?).

Does self-presentation affect implicit-explicit correspondence?

Self-presentation involves the attempt to present oneself as favorably as possible (Paulhus, 1982). In many situations, presenting oneself as favorably as possible can require responses that are in opposition to one's self-concept. For example, a KKK member may avoid expressing hatred of African Americans on a street corner in Harlem, and parents may avoid voicing their dislike of their son's spouse at his wedding.

[2] Some research in self-presentation separates the construct into two components: impression management – altering of responses for concerns relating to social interaction, and self-deception – altering of responses to mislead the self. While the studies in this dissertation emphasize the former, the manipulations and data do not clearly delineate between these alternatives, so the discussion of self-presentation is not specific as to what kind of self-presentation (impression management or self-deception) is occurring.

Vulnerability to self-presentation demands is thought to distinguish implicit and explicit measures of preference. While the expressions of explicit attitudes are vulnerable to self-presentation, implicit attitudes are thought to be impervious to such demands (Greenwald & Banaji, 1995). For example, even though it is often trivial to alter an explicit response, when instructed to alter their implicit responses, respondents are unable to do so (Kim & Greenwald, 1998; but see Lowrey, Hardin, & Sinclair, 2001 for a situation in which such an instruction is effective). The question of whether implicit attitudes are sensitive to other manipulations is an interesting one, but for now we consider the question of whether they are sensitive to change based on a conscious intent to produce a change, and the answer to that question appears to be negative.

Variation in IE correspondence across attitude objects can be explained by self-presentation concerns. When social demands are high (e.g., social reprimand for expressing dislike of African Americans), respondents may alter their response to be more socially desirable in the presence of others, but be unable to mute negative affect when the preference is measured implicitly. On the other hand, when social demands are minimal (e.g., reporting attitudes toward flowers or insects) there is no reason to expect a demand to alter one's response before reporting it.[3] In such circumstances, there would be little reason to expect implicit and explicit attitudes to diverge. If self-presentation were the only factor distinguishing implicit and explicit attitudes, we might consider implicit attitudes the pure form of the attitude, and explicit attitudes just implicit attitudes with adjustments accommodating for self-presentation. This, in essence, is the 'same'

[3] Note that alterations of self-report via self-presentation need not be active and intentional (lying), but could also be produced by more subtle processes (e.g., self-serving bias).

view that implicit and explicit measures assess the same construct. In other words, implicit attitudes are a lie detector for explicit attitudes.[4]

The appealing simplicity of this hypothesis is not sufficient to counter the weight of evidence against it. Implicit and explicit attitudes show low correspondence for a variety of attitude objects in which little to no self-presentation demands are apparent, such as attitudes toward flowers relative to insects or toward fruit and bugs (Greenwald et al., 1998; Nosek & Banaji, 2001). An absence of implicit-explicit correspondence in these conditions suggests that self-presentation is either not involved in moderating implicit-explicit correspondence, or at least is not sufficient to explain the relationship.

It is surprising, given its prevalence, that the assumed role of self-presentation in moderating the relationship between implicit and explicit attitudes has never been subjected to experimental test. The first study adresses this oversight and provides evidence for the role of self-presentation in moderating implicit-explicit correspondence. Specifically, the role of self-presentation is tested with a common, but powerful, procedure for manipulating self-presentational demand – changing the setting of the attitude report from a relatively private report to a more public one. The increased safety of reporting unpopular, undesirable, or unwanted attitudes in private is strained when those reports must be made in a more public setting, directly to another individual, and with possible consequences for maintaining social relationships. This public setting is, in effect, a manipulation of accountability with an audience whose preferences are unknown. If self-presentation plays a role in moderating implicit-explicit correspondence,

[4] Interestingly, though this view is often maligned, it appears to have no prominent supporters. Its origins can perhaps be attributed to the title of the Fazio et al. (1995) paper naming his automatic measure a 'bona-fide pipeline'. However, even in the text of that paper (and more so in subsequent writing), Fazio and colleagues remained uncommitted to whether automatically activated attitudes were merely explicit attitudes with the self-presentation stripped away.

then attitudes reported in a more public setting ought to be more weakly related to implicit attitudes than are attitudes reported in a more private setting.

The purpose of this dissertation is to identify factors that play a role in predicting the relationship between implicit and explicit attitudes. As a consequence, following the first study, I will turn to a second factor, attitude elaboration that might also predict part of the implicit-explicit relationship. The third study will return to self-presentation when the role of self-presentation and elaboration are examined simultaneously.

Does elaboration affect implicit-explicit correspondence?

The lack of relationship between implicit and explicit attitudes toward flowers relative to insects, or fruit and bugs, raises the question of other possibilities, besides self-presentation, that may underlie such lack of correspondence. As a second factor of interest, I propose to examine a construct that captures a lack of cognitive elaboration that usually accompanies deliberative thought. We have a general sense that flowers are good and insects are bad, but the lack of effortful consideration of those evaluations may prevent implicit and explicit modes of evaluation from converging.

Attitude elaboration, the amount of evaluative processing about an attitude object, is known to play an important role in determining the state of an explicit attitude. Petty and his colleagues have advanced a program of research that identifies the role of elaboration in attitude change. A central element of Petty's program of research is that extensive and effortful information processing actually engages cognitive restructuring through mechanisms such as thought rehearsal and reflection time. This cognitive restructuring then leads to attitude changes that are more temporally stable, resistant to persuasion, and predictive of behavior (Petty & Cacioppo, 1986; Petty & Wegener, 1999). The increased

stability and predictive capacity of attitudes after elaboration may translate to increased consistency across different forms (explicit and implicit) of attitude assessment. Using the language of rule-based (explicit) versus associative (implicit) processing, Smith and DeCoster (1999) suggest that the repeated use of a rule could allow a person, over time, to learn to use the rule associatively (Logan, 1988). In essence, the repeated expression of an attitude may lead to stronger encoding of that attitude such that, when automatically assessed, the same response is observed. In the second study of this paper, we consider whether the act of elaboration can provide the opportunity for implicit and explicit attitudes to achieve greater correspondence either by the automatization of explicit responses (Logan, 1988), the greater accessibility of automatic evaluations, or simply that the two usually independent modes of evaluation are similarly and simultaneously shaped by the exercise of elaboration.

Although no studies testing self-presentation as moderators of the IE relationship exist, it is frequently suggested as a possible moderator. With elaboration, the idea that such a moderator may exist is originally presented here. Hence, this research does not directly investigate the mechanism by which elaboration might influence IE correspondence, just whether or not elaboration can function as a moderator of that correspondence. The second study examines the possible influence of elaboration on IE correspondence with an experimental test in which correspondence is compared between attitude objects that have been elaborated with ones that have not.

These studies present a predictive relationship of self-presentation and elaboration for IE correspondence in order to provide an initial demonstration that a relationship does, in fact, exist. In the psychological literature, self-presentation and elaboration are

manipulated in a variety of ways, and there are some conditions in which we might predict the opposite of the proposed relationships (e.g., increasing self-presentation could increase IE correspondence in situations where accuracy accountability is high). The purpose of this dissertation is not to pursue the predictive subtleties of each factor. Rather, given the state of the literature, the focus is on whether IE correspondence exists, and if so, that there are factors that can predict (and perhaps explain) the relationship between implicit and explicit attitudes.

Study 1: Self Presentation

Self-presentation influences public reports more than private reports

People are more likely to express unpopular attitudes and admit to undesirable behaviors in private versus public contexts (for a review, see Schwarz, Groves, & Schuman, 1998). Whether the audience's own attitudes or beliefs are known or not, the mere presence of others is presumed to increase the accountability of responding and consequently influence self-report (Guerin, 1993; Lerner & Tetlock, 1999). Public respondents typically try to seek approval of audience members (Baumeister & Leary, 1995), such that if audience members' attitudes are known, responses shift to more closely approximate these preferences. When audience members' attitudes are not known, respondents engage in what Tetlock and his colleagues call 'preemptive self-criticism' (Tetlock, 1983; Tetlock, Skitka, & Boettger, 1989). This involves an internal consideration of multiple perspectives and anticipation of objections to one's response (see Lerner & Tetlock, 1999, for a review). Public expression of attitudes is thus affected by numerous presentational demands that do not appear to be influential factors for private expression of attitudes. The public or private reporting of attitudes should be

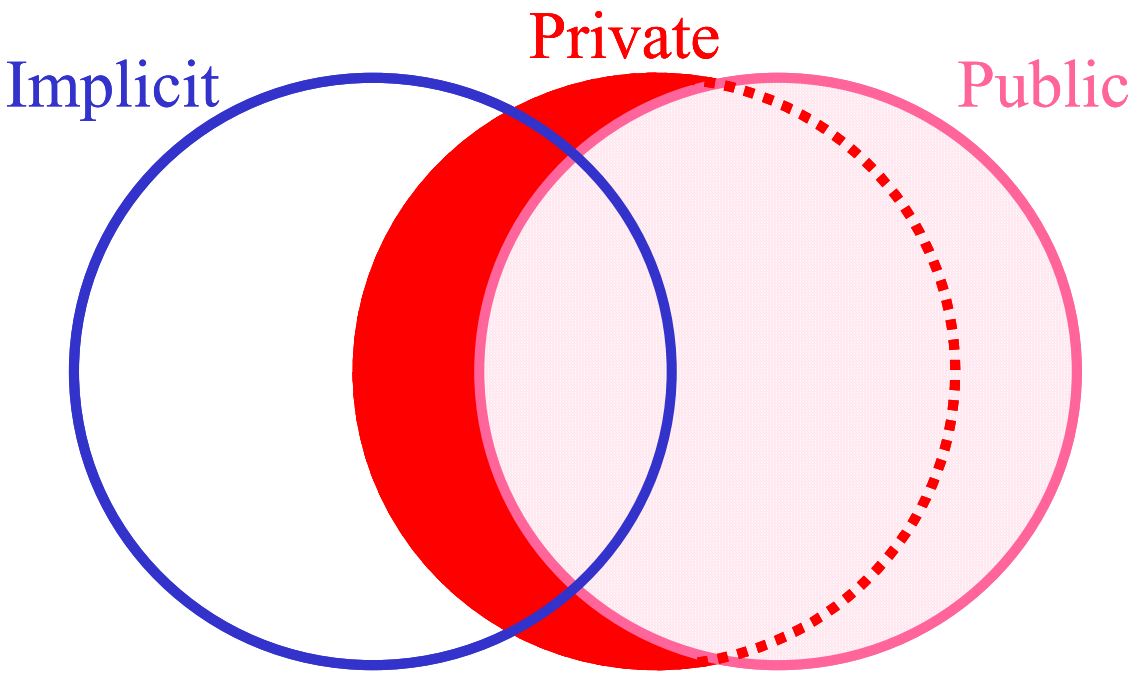
considered a manipulation of the relative magnitude of self-presentational concerns, not the presence or absence of such concerns.

Overview

In Study 1, shifts in implicit-explicit (IE) correspondence were compared under two conditions of explicit attitude assessment – an interview (public condition), and a computer report (private condition). If increasing self-presentational demands attenuates IE correspondence, then private attitudes should completely account for the relationship between implicit and public attitudes. That is, public attitudes should relate to implicit attitudes only to the extent that public and private attitudes are related. This prediction is presented graphically, as circles of overlapping variance, in Figure 1.

This hypothesis was tested with a hierarchical regression analysis. In the first step, public attitudes were submitted as a predictor of implicit attitudes. In the second step, private and public attitudes were simultaneously submitted as predictors of implicit attitudes. If the increased self-presentational demands of the public context reduce IE correspondence, then private attitudes should continue to predict implicit attitudes in the second step of the regression. But, public attitudes should not relate to implicit attitudes when private attitudes are included in the model. This finding would suggest that increasing self-presentational concerns (i.e., making a reporting situation more public) attenuates correspondence between implicit and explicit attitudes.

Figure 1. A Venn diagram illustrating the hypothesized relationships between implicit attitudes and explicit-private attitudes and explicit-public attitudes



A precondition for testing this hypothesis is that implicit and explicit attitudes (public or private) must be significantly related. If no IE correspondence were observed for a given variable, then the test would be non-diagnostic about whether self-presentational demands can influence IE correspondence. Also, in this study, public and private attitudes were assessed in the same session making it likely that they would be highly related. This requires negotiating a delicate balance between selecting attitude objects that will show some IE correspondence while simultaneously eliciting sufficient presentational concern to shift attitude reports between the public and private reporting sessions (separated by only five minutes). High multicollinearity between predictors in a regression decreases the power to detect relationships between the predictors (public and private attitudes) and the dependent variable (implicit attitudes). The five minutes between public and private reports does increase the likelihood that participants will be

motivated to be consistent and lower the overall power of the test. Despite this, I predicted that, to the extent that attitude reports change between public and private sessions, the private reports will relate to implicit attitudes more than public reports. That is, the increase of self-presentational demands in public compared to private reports will lead to decreased IE correspondence when explicit attitudes are measured publicly compared to privately.

Four target pairs (Black/White, gay/straight, science/humanities, Republican/Democrat) were selected to represent a range of attitude objects that could both show IE correspondence and change enough between public and private reports to produce a relatively sensitive test of the hypotheses. Two (Black/White and gay/straight) are thought to involve a good deal of self-presentation demand, but tend to show fairly low IE correspondence. The other two (science/humanities, Republican/Democrat) are typically assumed to invoke less self-presentational demand, but tend to show moderate to strong IE correspondence.[5] Attitudes toward the four target pairs were assessed under three conditions (public self-report, private self-report, implicit). Each target pair provides an independent test of whether making the reporting situation more private increases IE correspondence.

Method

Participants

Eighty-one undergraduates from Yale University participated for partial course credit or $8; one participant was removed for not following instructions.

[5] This tradeoff between the strength of self-presentation demands and the magnitude of IE correspondence is, itself, indirect evidence for the moderating role of self-presentational demands on IE correspondence. This variation across attitude objects does lower the power of the test for this study, and will be examined more directly in Study 3.

Explicit attitude measures

Attitudes toward four-target concept pairs (Black/White, gay/straight, science/humanities, and Republican/Democrat) were assessed with ratings on four nine-point semantic differential scales assessing evaluation (warm/cold, good/bad, positive/negative, and favorable/unfavorable). The difference between the ratings of target concept pairs on each scale was taken as a measure of explicit attitude (e.g., Black/White attitude = Black rating on warm/cold –White rating on warm/cold). Two semantic differentials were administered privately on a computer (private condition); the other two were administered in an interview with the experimenter (public condition). Selection of scales administered publicly versus privately was counterbalanced across participants. Following standard interview practice, experimenters were trained to exhibit neutral verbal and non-verbal reactions to all responses.

Ratings on the four semantic differential scales (2 administered publicly, 2 privately) were standardized and averaged within each condition resulting in a public and a private attitude score for each target concept.[6]

Implicit attitude measures

Implicit attitudes toward the target concept pairs were assessed with the Implicit Association Test (IAT; Greenwald et al., 1998; Greenwald & Nosek, 2001; for reviews of the measurement and assessment of implicit, or automatic, evaluation see Banaji, 2001).

---

[6] In addition to semantic differential ratings, participants publicly and privately reported agreement and disagreement to a variety of statements about each target concept (e.g., Science is the only way to learn what is real and what is not). Also, items adapted from Plant and Devine's (1998) motivation to respond without prejudice scale were administered at either the beginning or end of the study. These data are not discussed in this report, but the order of tasks did not qualify any of the results reported here.

Four IATs were presented to participants to assess Black/White, gay/straight, science/humanities, and Republican/Democrat attitudes.

The task operates on the basis of a principle that it ought to be easier to pair concepts (e.g., any attitude object such as *flowers* or *insects*) with attributes that have come to be associated through experience (e.g. the qualities of *good* or *bad*, *strong* or *weak*, *dangerous* or *safe*), than it is to pair concepts with attributes that are less or not at all associated. For instance, the concept flower and the attribute pleasant (e.g., represented by items such as *wonderful*, *happy*, *rainbow*) ought to be easier to pair mentally than the concept flower and the attribute unpleasant (e.g., *disgust*, *hatred*, *gun*). The extent to which it is easier to pair flower+pleasant (in the presence of a contrasting pair such as insect+unpleasant) compared to the opposite pairings (e.g., flower+unpleasant/insect+pleasant), the stronger is the assumed positive implicit evaluation of flowers relative to insects. In this task, ease or strength of association is measured by the speed to respond to a type of pairing (e.g., flower+pleasant) compared to another type of pairing (e.g., flower+unpleasant).

To perform the IAT, participants placed one finger on the 'a' key of the keyboard (the "left" key), and another finger on the '5' key (the "right" key) of the keypad. For half of the task, they were instructed to respond by pressing a key (say their left key) each time an item that represented the category flowers (words such as *rose, tulip*) and the category pleasant (words such as *peace, love*) appeared in the center of the screen. At the same time, they were asked to press a second key (say the right key) for all items that represented the category insects (e.g., *fly, mosquito*) or words that were unpleasant in meaning (e.g., *hatred, bomb*). If participants made an error, an 'X' appeared below the

item and they had to correct the error before moving on. For the other half of the task, one of the categories was switched such that flowers and unpleasant category words were classified on the same key, while insects and pleasant category words were classified on the other key.

The critical dependent variable is a difference in response latency computed by measuring the average speed to respond to flowers and pleasant items when they were paired together (while also paring unpleasant items with insects) and the average speed to respond to flowers and unpleasant items when they were paired together (while also pairing pleasant items with insects). The difference in average response latency was taken as an assessment of implicit preference (e.g., liking for flowers relative to insects). The IAT procedure followed the standard format outlined by Greenwald and colleagues (Greenwald et al., 1998) with the procedural variations described below.[7]

Stimuli representing the categories good and bad were taken from a list compiled by Bellezza, Greenwald, and Banaji (1986) and from Nosek and Banaji (2001) in which good and bad items were primarily positive and negative adjectives (e.g., wonderful, beautiful, terrible, awful). To assess automatic Black-White preferences, the categories African Americans and White Americans were represented by Black and White faces (Nosek et al., 2002b). To assess gay-straight preferences, the items 'gay' and 'homosexual' represented the category gay; the items 'straight' and 'heterosexual' represented the category straight (Lemm & Banaji, 2001). To measure science-humanities preferences, the categories science and humanities were represented by exemplars from those academic domains (e.g., physics, history; Nosek et al., 2002b).

[7] Demonstration IATs can be performed at http://www.yale.edu/implicit/.

Finally, to assess Republican-Democrat preferences, the categories Republicans and Democrats were represented by exemplars and descriptive items (e.g., George Bush, conservative, Al Gore, liberal). For a complete list of stimuli see Appendix A.

A warm-up IAT (flowers/insects evaluation) was administered before the critical blocks to familiarize participants with the task following the format described by Greenwald et al. (1998). The four implicit measures immediately followed the warm-up task. Unlike the standard Greenwald et al. (1998) procedure, only two blocks (e.g., one sorting Gay with bad [and Straight with good] and a second sorting Gay with good [and Straight with bad]) were administered for each target pair. For all blocks, 20 practice trials preceded the 36 critical trials to allow participants to become acquainted with the response configuration (a trial involved the classification of a single item into its appropriate category). Also, the eight blocks were presented in a randomized order to minimize task order effects.

Distracter task

A cognitively demanding distracter task was introduced between public, private, and implicit tasks. The distracter task was based on the popular game 'Boggle' (Parker Brothers, 1977). Each distracter session consisted of two Boggle tasks of three minutes each.

Procedure

The study was administered by one of two professionally dressed White male experimenters. The experiment consisted of three segments separated by two six-minute distracter tasks (Boggle game). Each segment consisted of one of three tasks – private attitude reporting (private), public interview attitude reporting (public), and implicit

attitude assessment (implicit). Implicit attitude measures were counterbalanced with explicit (public and private) attitude measures; and public attitude assessment was counterbalanced with private attitude assessment.[8] Except for the interview, all tasks were administered on a PC using Inquisit software (Draine, 1998). Before debriefing, participants reported general demographic information (e.g., age, major, ethnicity).

Analysis Strategy

One approach to analyzing these data would be to test for a difference in the magnitude of the IE correlations between the public and private conditions using a z comparison of correlations. There are two problems with the z-test approach for the current purposes. First, a z-test comparison of correlations does not test the hypothesis as described in the introduction of this study. The z-test only compares the magnitude of correlations, and not whether the variance of one relationship (implicit-private) is sufficient to explain the variance of a second relationship (implicit-public), but not vice-versa. The hypothesis outlined in Figure 1 is ideally suited for hierarchical regression. Any relationship between implicit and public attitudes should disappear when private attitudes are included in the model, although private attitudes should continue to be a predictor of implicit attitudes.

Second, even if a z-test were conceptually the appropriate test, it would be a severely underpowered test in the current experimental design. The implicit measure is known to have only moderate internal consistency making it difficult to obtain large zero-order correlations with other measures. Previous research suggests that the zero-order correlations between implicit and explicit attitudes will range from .1 to .5. Also,

[8] Across attitude objects, no systematic effects of order were observed.

participants report attitudes publicly and privately within a five-minute span. Any change in attitude reporting will have to be sufficiently strong to overcome ordinary consistency motivations. Any changes in the magnitude of the IE correlation between public and private conditions is not likely to be high. People who privately report strong liking for science over humanities are not likely to suddenly profess undying love for humanities and dislike of science in public. Rather the changes in attitude reports are likely to be more subtle; participants may simply temper their responses, or avoid expressing dislike of any discipline.

These factors suggest that shifts in correlations between public and private settings will be rather weak. For example, an IE correlation of 0.4 in the private condition might decline to 0.3 in the public condition. Using Cohen's (1988) terminology, this difference should be considered a small effect ($q = .114$).[9] A power test (assuming $p < .05$, and $n = 80$) reveals that there is only a 15% chance of detecting an effect of this magnitude. To achieve acceptable power (.80), we would need more than 1000 participants ($n = 1000$, power = .72).

A more generous estimate of the magnitude change in correlations is not much more encouraging. For example, shifts in IE correspondence between private and public conditions might be $r = 0.5$ to $r = 0.3$, or $r = 0.4$ to $r = 0.2$. These correlation differences would translate into effect sizes between small and medium in magnitude (q's = .239 and .221 respectively). In this case, the power to detect a significant difference would be approximately 36%. A sample size of 300 would be required to achieve power of .80.

[9] Cohen (1988) suggests that $q = .1$ is a small effect, $q = .3$ is a medium effect, and $q = .5$ is a large effect.

Clearly, the z-test approach is not the appropriate test for the current hypothesis and experimental design.

Hierarchical regression using implicit attitudes as the dependent variable and public and private attitudes as predictors is a direct test of the relationship between implicit, public and private attitudes as described in Figure 1. This test will demonstrate whether private attitudes can completely explain the implicit-public relationship, and simultaneously predict implicit attitudes beyond what can be explained by public attitudes. The first step of the hierarchical regression will account for the relationship between implicit and public attitudes, and private attitudes will be entered as a second step. At this second step, all of the variance shared between the two predictors (public and private attitudes) will be removed. The remaining portion of variance will represent only what changes as a function of reporting the same attitude in a public versus a private condition. According to the hypothesis, the unique portion of public attitudes will not be related to implicit attitudes, but the unique portion of private attitudes will still show a significant relationship. If true, this will support the hypothesis that private attitudes can completely explain the relationship between implicit and public attitudes. That is, increasing self-presentation demand weakens the relationship between implicit and explicit attitudes.[10]

---

[10] A typical concern in multiple regression is multicollinearity between the independent variables. Specifically, multicollinearity can affect sampling stability and make the prediction less stable and inflate the standard errors. These factors reduce the possibility of rejecting the null hypothesis (Cohen & Cohen, 1983). In this experiment, we expect that collinearity between the two predictors, public and private attitudes, will be quite high (indeed it was over .8 for each attitude object). The concerns about multicollinearity in regression applications are not an issue for the present research. In fact, they work against the current hypothesis, which depends on rejecting the null hypothesis that there is no relationship between private and implicit attitudes after controlling for public attitudes. In other words, it will not be surprising to see that public attitudes do not continue to relate to implicit attitudes in the second step of the hierarchical regression, but it

Results

Data preparation

Data were prepared following the conventions established by Greenwald et al. (1998). Errors were coded but retained in the analysis (5.2%; 1302/24948). Trial latencies shorter than 300 ms and longer than 3000 ms were recoded as 300 ms and 3000 ms respectively (1.6%; 388/24948). Latencies were not transformed.

Private attitudes account for the relationship between implicit and public attitudes

I hypothesized that privately assessed attitudes would fully explain the relationship between public and implicit attitudes. That is, public attitudes would only relate to implicit attitudes to the extent that public attitudes are similar to private attitudes. Private attitudes, on the other hand, would continue to associate with implicit attitudes even after removing variance shared with public attitudes. If neither public nor private attitudes continued to relate to implicit attitudes when included as simultaneous predictors, then they could be considered redundant measures (at least as far as their relationship with implicit attitudes is concerned). If both public and private attitudes continued to relate to implicit attitudes when included as simultaneous predictors, it would suggest that public and private responding represent unique elements involved in implicit-explicit correspondence. Results for the four hierarchical regressions are reported in Table 1.

Table 1. Two-step hierarchical regressions for four attitude objects: (1) Black/White, (2) Gay/Straight, (3) Science/Humanities, and (4) Republican/Democrat. (* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$)

Black/White	B	SE B	Beta	R^2
Step 1				.001

would, given all the issues of power stacked against it, be quite surprising to see that private attitudes continue to significantly relate to implicit attitudes.

Public Attitudes	7.57	23.8	.036	
Step 2				.002
Public Attitudes	4.56	28.5	.022	
Private Attitudes	3.44	17.7	.026	

Gay/Straight	B	SE B	Beta	R^2
Step 1				.111
Public Attitudes	45.16	14.5	.333**	
Step 2				.162
Public Attitudes	1.22	24.8	.009	
Private Attitudes	35.74	16.6	.395*	

Science/Humanities	B	SE B	Beta	R^2
Step 1				.152
Public Attitudes	26.91	7.2	.390***	
Step 2				.198
Public Attitudes	-0.32	14.7	-.005	
Private Attitudes	31.2	14.8	.449*	

Republican/Democrat	B	SE B	Beta	R^2
Step 1				.173
Public Attitudes	33.45	8.28	.416****	
Step 2				.185
Public Attitudes	17.26	17.15	.215	
Private Attitudes	15.89	14.73	.230	

No relationship was observed between implicit and explicit (public or private) Black-White attitudes, which is consistent with numerous previous demonstrations (Fazio et al., 1995; Greenwald et al., 1998). This lack of relationship prevents this attitude object from providing a test of the primary hypothesis. In addition, it points out that manipulating self-presentation demand with a public-private manipulation is insufficient to explain all of the variability in IE correspondence. It is possible that, with additional assurance of privacy, self-reported racial attitudes would correspond with implicit racial attitudes. It is

also possible that implicit and explicit racial attitudes will never be strongly related, no matter how little self-presentation demand is present.

For gay/straight and science/humanities attitudes, a significant relationship between implicit and public attitudes in step 1 of the hierarchical regression (gay/straight ß = .33, p = .003; science/humanities ß = .39, p = .0004) was reduced to near zero in the second step of the regression (gay/straight ß = .01, p = .96; science/humanities ß = -.005, p = .98). At the same time, in the second step a significant implicit-private relationship remained after controlling for public attitudes for both gay/straight (ß = .39, p = .03) and science/humanities (ß = .45, p = .04) attitudes. This effect is particularly impressive considering the strong relationship between public and private attitudes for both tasks (r's = .82 and .88), which left very little variance for an independent relationship between implicit and private attitudes to persist. In other words, despite the fact that less than 35% of the variance in private attitudes was still available after controlling for public attitudes, private attitudes were still significantly related to implicit attitudes.

The regressions for Republican/Democrat attitudes did not support the hypothesis that implicit attitudes would be more robustly related to private than public attitudes. Public and private Republican/Democrat attitudes were similarly related with implicit Republican/Democrat attitudes.

Discussion

Study 1 provided evidence that implicit-explicit correspondence is moderated by self-presentation. For two of the three attitude objects (in which a significant relationship existed between implicit and explicit attitudes), private attitudes predicted implicit

attitudes even after controlling for public attitudes. Further, in those cases, the private attitudes completely accounted for public attitudes' relationship with implicit attitudes.

It is notable that these effects were observed despite the fact that public and private attitudes were highly correlated (r's > .80). In most circumstances, correspondence of this magnitude would be taken as evidence that the same construct is being measured. Even so, the small amount of variance in private attitudes unaccounted for by public attitudes was significantly related to implicit attitudes.

A variety of possibilities could explain the failure of political (Republican/Democrat) attitudes to replicate the effects observed with gay/straight and science/humanities attitudes. First, the test had low power because of high collinearity between public and private attitudes (r = .88), which may have produced Type II error. Second, self-presentational concerns for reporting political attitudes may be weak in both public and private reporting conditions leading to little difference across conditions. It is possible that a stronger manipulation of self-presentational demands (i.e., requiring attitude report in front of a large group of strangers) would demonstrate attenuation similar to the other attitude objects. In fact, all attitude objects should be vulnerable to self-presentation given the right conditions. Our 'soft' manipulation involved a single interviewer whose own attitudes were unknown. If the experimenter was known to be an enthusiastic Republican, participants may have been less willing to report dislike of Republicans, and Republican-Democrat attitudes ought to have shown the same pattern as the other attitude objects.

Limits on the explanatory power of self-presentation

Despite evidence that self-presentation moderates implicit-explicit correspondence, these data do not suggest that self-presentation is sufficient to explain everything about

the relationship between implicit and explicit attitudes. For one, racial (Black/White) attitudes did not show any IE correspondence whether attitudes were assessed publicly or privately. Also, when implicit-explicit relationships were observed, the relationships were still only moderate in size (even for implicit-private relations). These relationships are attenuated to some extent by measurement error (Cunningham et al., 2001). Even so, these findings are consistent with other research suggesting that implicit and explicit attitudes have both shared and unique components (Cunningham et al., 2001; Greenwald & Farnham, 2000). The lack of a complete account for IE correspondence by self-presentation allows the possibility that multiple factors moderate IE correspondence.

Another issue for the role of self-presentation in moderating IE correspondence concerns the shift in reliability of attitude measures in public versus private settings. In these data, public reports (mean split-half $\underline{r}$ = .80) were slightly less reliable than private reports (mean split-half $\underline{r}$ = .84), which may be psychologically meaningful (e.g., public situations cause greater anxiety, which may cause greater inconsistency in attitudinal reports) or may be statistical artifact. In any case, differences in reliability may affect the strength of the correlations with implicit attitudes, which could play a role in producing the observed effects. Future investigations should directly examine the affects of public-private manipulations on the reliability of self-reports, and then determine whether those shifts in reliability are responsible for the effects reported here.

This first experiment suggests that a situational manipulation of self-presentation concerns moderates implicit-explicit correspondence. In the next study, we examine whether a second factor, elaboration, might also contribute to the relationship between implicit and explicit attitudes.

Study 2: Elaboration

Extensive and effortful information processing produces cognitive restructuring through mechanisms such as thought rehearsal and reflection time (Petty & Cacioppo, 1986; Petty & Wegener, 2001) leading to increased attitude stability and predictive utility. The increased stability and predictive capacity of attitudes after elaboration may translate to increased consistency across different forms (explicit and implicit) of attitude assessment either by the automatization of explicit responses (Logan, 1988), the greater accessibility of automatic evaluations, or simply that the two usually independent modes of evaluation are similarly and simultaneously shaped by the exercise of elaboration.

In Study 2, we tested whether merely thinking about an attitude object was sufficient to increase correspondence between implicit and explicit attitudes. Participants elaborated on one of two issues: (1) should children with disabilities be integrated into the standard classroom environment (mainstreaming or inclusion) or be in a school environment that is separate and specially designed for their needs (special education), or (2) what is more important, conserving our natural resources or developing lands to support our energy needs (specifically with regard to the Artic National Wildlife Refuge)?[11] The elaboration exercise involved reading factual materials and writing arguments for and against the two positions presented in the materials. Immediately following elaboration, participants' attitudes were assessed for both topics with the unelaborated topic serving as a control attitude. We hypothesized that implicit and explicit attitudes would show stronger correspondence for the elaborated attitude than for the control attitude. To test whether

[11] This study was conducted in the Fall of 2000, about the time that the political debate over the use of ANWR was covered in the news media. As a consequence, most participants had heard of the topic, but did not yet know a great deal about it.

the elaboration exercise had a lasting effect on implicit-explicit correspondence, participants also completed a follow-up assessment three to four days later via the Internet. Being exploratory, no specific hypothesis was generated for implicit-explicit correspondence at the follow-up session.

Method

Participants

Seventy members of the Yale community were recruited to participate for $10. Of the 70 participants, 61 logged in to the website for Session Two. Of the 55 participants who completed the second session, 1 returned two days after Session One, 24 returned three days later, 22 returned four days later, 7 returned five days later, and 1 returned six days later.[12] Data are reported for the 55 participants who completed both sessions. Results for Session One are not qualified when data from all 70 participants were included.

Materials

Elaboration materials and measures. The experimenters prepared two 2-page single-spaced readings each describing one issue: either educating children with disabilities, or the relative importance of conservation and environment in the recent debate about oil drilling in the Artic National Wildlife Refuge (ANWR). Information for these readings were culled from published news articles available on the Internet and from public documents distributed by organizations promoting one of the two solutions for each topic. The experimenters edited and informally pretested the readings to create a balanced

[12] Six participants logged into the website but did not complete the second session. Specific reasons for premature termination of participation are unknown but could include: equipment malfunction, boredom, or distraction (see Nosek, Banaji, & Greenwald, 2002a for a discussion of early termination in web-based research).

set of arguments that did not noticeably favor one position over the other. The materials avoided connecting the debate to other well-elaborated concepts (e.g., Republican/Democrat attitudes) by providing no information about the identities of the protagonists for each position. In addition, four factual questions were presented immediately after the reading to insure that participants were attending to the reading materials.[13] The readings and fact questions are available in Appendix B.

Implicit attitude measures. Implicit attitudes for mainstreaming versus special education and conservation versus development were assessed with the Implicit Association Test (Greenwald et al., 1998). To assess implicit attitudes toward mainstreaming (inclusion) versus special education, an IAT was created with the categories 'mainstreaming' and 'special education' and evaluative categories 'good' and 'bad'. To assess implicit attitudes toward conservation versus development, an IAT was created with the categories 'conservation' and 'development' and evaluative categories 'good' and 'bad'. Stimuli representing the attitude objects were culled from the reading materials to be clearly representative of their categories. Labels and stimuli for all categories are listed in Appendix A.

Explicit attitude measures. In evaluating whether mainstreaming (inclusion) or special education was preferable for children with disabilities, participants made ratings on three scales anchored by mainstreaming/special education, all together/separate tracks, and same schools/unique schools. In evaluating whether conservation or development was preferable in the context of the ANWR debate, participants made preference ratings on

[13] All participants answered at least 3 of the 4 questions correctly indicating that they did attend to the experimental materials.

three scales anchored by conservation/development, environment/industry, and maintain habitat/promote growth.[14]

Procedure

Session One. Before agreeing to participate in the experiment, participants were informed that the study involved two sessions: one in the laboratory, and a follow-up session between three and four days later administered via a website. After obtaining informed consent, participants were seated in a private room with the experimental materials and a Windows based PC running Inquisit software (Draine, 1998). First, participants rated their familiarity with both elaboration topics (educating children with disabilities and relative priorities of development and conservation) on a 1 to 5 scale in which 5 indicated strong familiarity and 1 indicated little to no familiarity.[15]

Second, participants were randomly presented with one of the two sets of elaboration materials. Participants were instructed to read the materials carefully and then answer the factual questions about the reading. Participants were told that they could review the materials to answer the factual questions. After reading the passage and responding to the fact questions, participants were instructed to spend 20 minutes generating arguments for and against the two positions presented in the reading materials. The experimenter impressed upon the participants that it was important that they think critically about the topic and the arguments presented for and against each position saying, “Please take the

[14] An unnoticed difference in the first and second session materials resulted in the presentation of 7-point semantic differential scales in the first session, and 9-point semantic differential scales in the second session. While this error does not threaten any comparisons between elaboration and control conditions, it could result in slight underestimations of the reliability of the explicit measures between sessions.

[15] Initial perceived familiarity did not moderate observed IE correspondence in either the elaboration or control condition, or for either attitude object.

critical thinking part of this study seriously, it is very important for our study that you get engaged with the issues and consider as many of the possible consequences of the various courses of action." In addition, the experimenter made it very clear that participants were to generate arguments for and against both positions, considering the positives and negatives of each course of action. Participants were encouraged to think beyond the materials presented to them and generate arguments of their own. Participants were left alone to complete the materials and were informed when they had five minutes left.

Third, participants were seated in front of the computer and informed that they would complete two types of computer tasks: one involving judgments on rating scales and another involving classification of words into groups (IATs). The procedures were explained and a list of the category labels and stimulus items were presented for both the studied topic (elaborated) and non-studied topic (control) to ensure that participants knew how to perform the task. Presentation of the elaborated and control attitude measures was counterbalanced, as was the order of implicit and explicit measures within topic.

Finally, participants completed a demographics form, were paid, and were reminded that they were to return to the Internet for a 15 minute follow-up session in three or four days. Participants left with information for connecting to the website for the second session, and each participant was emailed on the third day to remind them to complete the second session.

Session Two. Participants connected to a private Internet address at the research branch of the experimenter's research website (see Nosek et al., 2002b). Before completing the implicit and explicit attitude measures, participants entered a code to link the data between Sessions One and Two. Following that, the labels and stimuli for both

topics were presented as well as the instructions for completing the tasks. No mention was made of the elaboration exercise from Session One. Participants simply completed the same implicit and explicit attitude measures as before. The presentation order of topics was randomized, as was order of implicit and explicit measures within topic.

Results

Data preparation

Data for the implicit measures were prepared using conventions established by Greenwald et al. (1998). Error trials were coded but retained (Session One error rate: 6.8%, 599/8800 trials; Session Two error rate: 9.3% error rate, 767/8284 trials). Response latencies below 300 milliseconds and above 3000 milliseconds were recoded as 300 and 3000 milliseconds respectively (Session One: 1.2%, 107/8800 trials; Session Two: 2.4%, 200/8284 trials). Data for one implicit measure for four participants were discarded because of a high proportion of error rates and fast/slow response latencies (> 25%), which make the measurement unstable.

IAT effects were calculated by taking a difference score between the average latency for one pairing (e.g., conservation+good/development+bad) and the average latency for the second pairing (e.g., conservation+bad/development+good). After conducting tests for the simple effects, the difference scores were standardized (Mean = 0; SD = 1) separately for each attitude object. Then, standard scores were resorted into elaborated and control attitude categories. This way, IE correspondence could be assessed for elaborated and control attitudes without reference to the specific attitude objects themselves.

Simple effects

No differences in main effects of implicit and explicit attitudes were observed between elaboration and control conditions, so simple effects are reported for the whole sample. Participants preferred conservation to development both implicitly (M = 108ms, SD = 183; $t(54) = 4.37$, $p < .0001$) and explicitly (M = 1.02, SD = 1.42; $t(54) = 5.30$, $p < .0001$). Also, while participants implicitly preferred mainstreaming to special education (M = 74, SD = 212; $t(54) = 2.59$, $p = .01$), they showed no explicit preference (M = .07, SD = 1.46; $t(54) = 0.37$, $p = .71$). No differences were observed in comparisons of main effects for attitudes assessed in Session Two compared to those assessed in Session One.

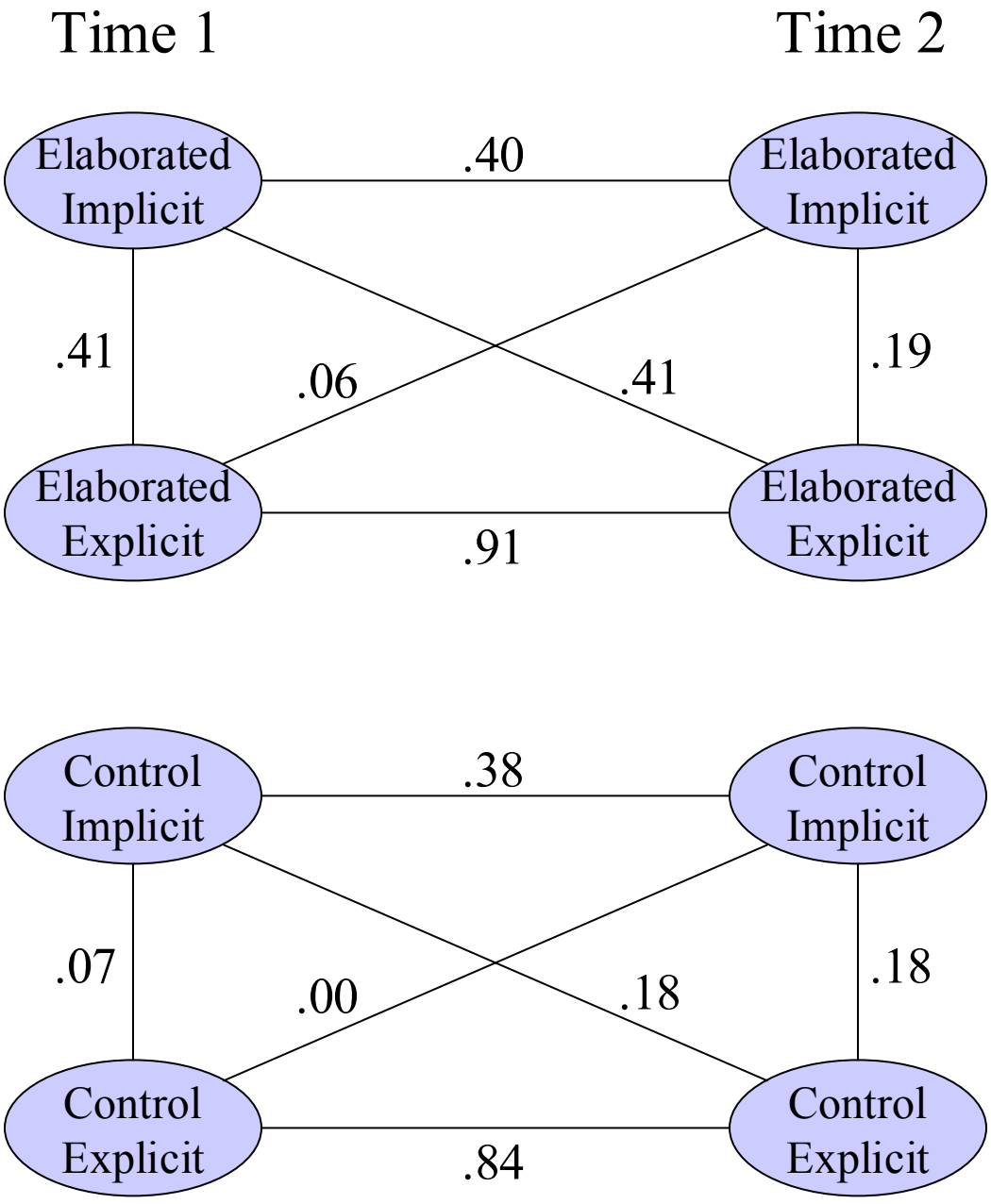
Implicit-Explicit correspondence immediately following elaboration

We predicted that implicit and explicit attitudes would show greater correspondence following an elaboration exercise compared to no elaboration. In this sample, a significant zero-order correlation between implicit and explicit elaborated attitudes was observed ($r = .41$, $p = .002$), but no relationship was observed between implicit and explicit control attitudes ($r = .07$, $p = .59$; see Figure 2). A z-test comparing these correlations confirmed that IE correspondence was stronger for elaborated than control attitudes ($z = 2.57$, $p = .01$) immediately following the elaboration exercise.

Implicit-Explicit correspondence over time

Participants completed implicit and explicit attitude measures in a second session a few days after the elaboration exercise. Because of the exploratory nature of the time course for IE correspondence, no specific predictions were made for the presence or absence of correspondence at Session Two.

Figure 2. Zero-order correlations among implicit and explicit attitudes within and across sessions for elaboration and control conditions. (Study 2)



Implicit and explicit elaborated attitudes were not significantly related at Session Two (r = .19, p = .15), nor were implicit and explicit control attitudes (r = .18, p = .21; see Figure 1). Obviously, these correlations were not significantly different from one another (z =

.15, p = .88). IE correspondence that had been produced by elaboration in Session One was not apparent a few days later in Session Two.

Figure 2 also presents zero-order correlations that index the stability of implicit and explicit measures from Session One to Session Two. Elaborated explicit attitudes (r = .91) were more stable over time than control explicit attitudes (r = .84; z = 2.17, p = .03)[16] consistent with previous research demonstrating a link between elaboration and attitude stability (Petty & Wegener, 1999). However, the benefits of elaboration for attitude stability did not translate to implicit attitudes. Elaborated implicit attitudes (r = .40, p = .003) were not more stable over time than the control implicit attitudes (r = .38, p = .005; z = .17, p = .87).[17]

Internal consistency of elaborated and control implicit and explicit attitudes

Though no hypotheses were generated about the effects of elaboration on the internal consistency of implicit and explicit attitudes, observation of changes in internal consistency could shed some light on the mechanisms by which elaboration is affecting IE correspondence.[18] For example, the elaboration exercise could momentarily crystallize implicit (and explicit) attitudes making them more internally reliable, which

[16] Readers may be surprised that the z-test is significant given the r's of .91 and .84. Remember that r's are not distributed linearly. The difference in variance explained between correlations of .91 and .84 is 12.25%, which is the same as the difference between r's of .00 and .35.

[17] Readers may note that explicit attitudes show much greater correspondence across time than implicit attitudes. This is probably due, in part, to the greater reliability of explicit measurement compared to implicit measurement (Cunningham et al., 2001; Bosson, Swann, & Pennebaker, 2000), but could also indicate more interesting psychological mechanisms at work, such as greater instability of implicit attitudes compared to explicit attitudes over time, or greater vulnerability of implicit attitudes to situational variations. This, itself, is another form of evidence against the 'same' attitude perspective of implicit and explicit attitudes – they differ in the extent to which particular situational constraints influence implicit or explicit measures.

[18] Thanks to Geoff Cohen for making this suggestion.

could improve IE correspondence. This would be a relatively clear and direct mechanism by which elaboration increases IE correspondence.

Table 2. Internal consistency correlations for elaborated and control implicit and explicit attitudes. Implicit attitude correlations are a split-half correlation (zero-order), and explicit attitudes correlations are the average zero-order correlation among the three items assessing preferences for the elaborated attitude or the control attitude.

		Session	
		1	2
Implicit	Elaborated	.56	.46
	Control	.76	.80
Explicit	Elaborated	.83	.90
	Control	.54	.73

The data in Table 2 indicate that the elaboration exercise did increase internal consistency, but only for explicit attitudes. In the first and second session, elaborated explicit attitudes were more internally consistent than their control counterparts. Unexpectedly, the exact opposite pattern emerged for implicit attitudes. In both sessions, attitudes that were elaborated (Session 1 = .56; Session 2 = .46) were less internally consistent than control attitudes in the same session (Session 1 = .76; Session 2 = .80). It appears that the elaboration exercise actually increased the instability of the implicit attitudes. This raises an interesting question about the mechanism by which elaboration affects IE correspondence. Despite the fact that IE correspondence increased as a function of elaboration, the internal consistency of implicit attitudes actually declined. Consider an analogy likening implicit attitudes to the global economy, and explicit attitudes to the U.S. economy (analogy for relationship, not for relative size). When many countries are doing well, and the global economy is stable, the link between the U.S.

economy and the global economy is moderately strong. However, when the global economy gets more unstable, its dips and rises are more of a function of the U.S. economy because of the significant size of the U.S. economy. In a similar vein, the current data suggest that the increasing instability of implicit attitudes because of elaboration was accompanied by a stronger link between implicit and explicit attitudes. Understanding the process that led to this shift is left for future research.

The decrease in internal consistency of implicit attitudes may be, in part, a consequence of the elaboration exercise selected for this study. In this elaboration exercise, participants were instructed to consider both sides of the debate and generate pros and cons for each position. The increase in internal consistency of elaborated explicit attitudes suggests that this exercise helped participants to generate a specific and consistent position on the issue. This is consistent with our understanding of explicit attitudes, which ought to be the products of our reasoned consideration of multiple viewpoints and perspectives. Having considered multiple viewpoints and generated a position, participants could be more confident and consistent in their explicit attitudinal responses. Implicit attitudes, on the other hand, showed much less stability after this elaboration session. Perhaps the exercise of generating pros and cons for both positions of an issue made all of those elements (even the one's that the person explicitly disagrees with) more automatically accessible and, as a consequence, led to greater instability in the automatically activated evaluations. If this hypothesis is true, the internal consistency of implicit attitudes might increase following an elaboration exercise in which a single position is elaborated because the accessible cognitions would be evaluatively consistent.

The intriguing effects of elaboration on internal consistency will need to be pursued in future research.

Discussion

The results of Study 2 demonstrate that an elaboration exercise increases the correspondence of implicit and explicit attitudes measured immediately after the elaboration exercise. Spending 20 minutes thinking about the pros and cons of two competing positions on political issues (conservation versus development of natural resources and mainstreaming versus separate schooling for children with disabilities) was sufficient to increase IE correspondence from near zero to moderate levels. This is the first known evidence that elaboration may be involved in the production of correspondence between implicit and explicit attitudes. Though the mechanism is far from clear, this study supports further study of elaboration as a predictor of IE correspondence. In sum, self-presentation is not the only factor involved in predicting IE correspondence; attitude elaboration may also be involved.

Typical conceptions of automatic processes hypothesize that they are slow to develop and resistant to rapid change (Logan, 1988; Wilson et al., 2000). This experiment, however, showed that elaboration between implicit and explicit attitudes could be increased with a short (20 minute) elaboration exercise and that shift was relatively short-lived – certainly less than three days and possibly only a few minutes. So, if implicit attitudes were altered as a consequence of the elaboration exercise, how did it happen?

One possibility is that implicit attitudes didn't shift at all – the increased correspondence could be exclusively due to alterations of explicit responses to be more consistent with implicit ones. This possibility would fit neatly into current conceptions of

explicit attitudes as easily changed relative to implicit ones that are slow to form and slow to change (Wilson et al., 2000). The data, however, suggest that this did not occur. Elaborated explicit attitudes were very consistent between sessions ($\underline{r}$ = .91) suggesting that, if they had shifted to be more in line with implicit attitudes, then they should continue to relate to explicit attitudes during the second session. The data pattern in Figure 2 clearly shows that explicit attitudes assessed in Session Two did not continue to correspond to implicit attitudes in Session Two. Rather, they related to the implicit attitudes assessed in Session One. This, coupled with the modest reliability of the implicit attitudes between sessions, suggests that changes in the strength of IE correspondence between Sessions One and Two was due to temporary shifts in implicit, not explicit, attitudes. This does not suggest that explicit attitudes were not affected by the elaboration exercise, which occurred before attitude report in Session One. But, it does suggest that that the small shifts in explicit attitudes between sessions were not responsible for the changes in IE correspondence.

The short-term influence of elaboration observed in Study 2 may be less indicative of actual change to representations about the target concepts, and more indicative of the malleability of implicit social cognition (see Banaji, in press). A number of recent papers suggest that implicit attitudes are more malleable than previously thought (Blair, Ma, & Lenton, 2001; Dasgupta & Greenwald, 2001; Lowrey, Hardin, & Sinclair, 2001; Mitchell, Nosek, & Banaji, 2002; Nosek & Banaji, 2002b). Banaji (in press) argued that it might not be accurate to think about attitude change for implicit social cognition in the same way that we think about change for explicit attitudes. Implicit attitudes are evoked from a network of associations that exist as patterns of activations, not actual

representations of concepts (Smith, 1996). Simple changes to the context of evaluation may alter the pattern of activation for the particular attitude object, such that it is less appropriate to think about shifts in implicit evaluations as changes to the specific evaluation of the target, and more appropriate to consider implicit evaluations as reflections of the target in its context of evaluation.

Minimal changes to the experimental context are sufficient to alter the expression of implicit attitudes in dramatic ways. For example, Nosek and Banaji (2002b) observed that just changing the gender of the experimenter could eliminate gender differences in automatic math attitudes. When the experimenter was male, women showed significantly greater negativity toward math than men did. When the experimenter was female, women and men were not different in their implicit evaluations of mathematics. Yet, when no experimenter is present (as in Internet-based research) stable gender differences are observed in implicit math attitudes (Nosek et al., 2002b).

In the current study, the elaboration exercise may have made a particular set of cognitions about the target attitude object very accessible. Those accessible cognitions may have similarly influenced both the explicit and implicit responses. Interestingly, the time course data would then suggest that the elaboration exercise had impact in terms of long-term stability of explicit responses, but much less so for implicit responses. This is consistent with current theory about the slow responsiveness of implicit associations to lasting change (Logan, 1988; Wilson et al., 2000).[19] Additional investigations are necessary to determine what is actually changing as a consequence of the elaboration exercise and how that change is affecting IE correspondence.

[19] Is the evidence for greater malleability of implicit than explicit attitudes in conflict with evidence that implicit attitudes are slower to change than explicit attitudes? Though this topic is beyond the purposes of this dissertation, I provide a brief discussion of it in Attachment A.

The current study does not provide direct evidence for causal priority between the generation of implicit and explicit attitudes, but it does offer the opportunity for some speculation. The high degree of consistency between sessions for elaborated explicit attitudes (r = .91), the fact that implicit attitudes from Session One related to explicit attitudes from both sessions (r's = .41), and the fact that implicit attitudes from Session Two were not related to explicit attitudes from either session, together implies that implicit attitudes were temporarily shaped to be more consistent with explicit attitudes in Session One. These suggest that this elaboration exercise may be effective in bringing implicit attitudes into line with one's explicit attitudes. In the present case, the impact of the elaboration exercise was short-lived, but repeated elaboration sessions may show greater effectiveness at instilling long-term change. In other words, elaboration may engage cognitive resources that can cause implicit attitudes to become more similar to explicit ones. It is not clear from these data whether the opposite may be true – can explicit attitudes be shaped to be more consistent with implicit ones?

That the observed effects may be indicative of the malleability of implicit cognition raises questions about whether the results of Study 2 will generalize to extended elaboration exercises and predictions of stable differences in IE correspondence. Obviously, effects of situational malleability on implicit attitudes could not easily explain why stable differences in IE correspondence is observed between math/arts and Republican/Democrat attitudes on the higher end, and flowers/insects and young/old attitudes on the lower end. In the next study, we examine whether elaboration can help to account for the stable differences observed across attitude objects. Also, in Study 3, we

examine whether self-presentation and elaboration are unique predictors of IE correspondence by investigating them simultaneously.

Study 3: Self-presentation and Elaboration[20]

In the previous two studies, I experimentally manipulated self-presentational demand (Study 1) and elaboration (Study 2) for a specific attitude object and observed that both manipulations moderated IE correspondence. Much of the interest in IE correspondence, however, comes from the observation that some attitude objects (e.g., math/arts attitudes [r = .48]) have stronger IE correspondence than other attitude objects (e.g., Black/White attitudes [r = .18], old/young attitudes [r = .01]; Nosek et al., 2002b and this study). Informally, the attitude objects that have been subjected to experimental test appear to vary in the amount of self-presentation demand inherent in their expression (e.g., Black/White attitudes engenders strong presentation demands, while math/arts attitudes does not), and in the amount of elaboration that is likely to have been invested (e.g., flowers/insects attitudes are not likely to be highly elaborated, but Republican/Democrat attitudes are probably more elaborated). In Study 3, we test whether self-presentation and elaboration can explain variation in IE correspondence across a variety of attitude objects.

Study 3 also offers an opportunity to examine self-presentation and elaboration at the same time. Do self-presentation and elaboration simultaneously explain variation in IE correspondence, or are they redundant moderators? Also, do self-presentation and elaboration interact to predict IE correspondence? When self-presentation demand is high, attitude elaboration may be irrelevant for IE correspondence because there are still many reasons to alter explicit responses (without the comparable opportunity to alter

[20] A simplified report of this study appears in Nosek and Banaji (in press)

implicit responses). However, when presentation demand is low, there is little motivation to alter one's explicit responses, so variation in IE correspondence may be more heavily influenced by other factors, like attitude elaboration.

In this study, we examine the role of self-presentation and elaboration in predicting variation in IE correspondence across attitude objects. This was accomplished by synthesizing data collected in the traditional laboratory and from two demonstration websites. Because the analysis uses IE correlations for attitude objects as the data points, the power of the test depends on the number of unique attitude objects that have been systematically investigated with similar implicit and explicit measurement tools. With these conservative inclusion requirements, the sample of IE correlations is relatively small ($\underline{N} = 15$) resulting in a low power test, but (hopefully) compensated for by the consistency of experimental designs and measurement tools, which should decrease random error.

The studies included in the synthesis were not directly designed to test the research questions as stated. No objective measures of elaboration of self-presentation demand were collected with the implicit and explicit attitude data. For this study, raters were recruited to estimate the amount of self-presentation demand and elaboration inherent in each attitude object "for the average person." As is evident from high inter-rater reliability (>.9), there was strong agreement about the degree to which varying attitude objects invoked high and low self-presentation demands and involved greater or lesser elaboration. Even so, the estimates of elaboration and self-presentation are subjective rather than objective assessments.

Because of the small sample and the subjective assessments of elaboration and self-presentation, this study only provides preliminary evidence for the viability of self-presentation and elaboration as explanatory factors in determining the relationship between implicit and explicit attitudes across attitude objects.

Method

Inclusion criteria

Over the past six years, we have collected measures of implicit attitudes via two means: traditional laboratory and web-based data collections. The web-based data described in this paper were collected at two demonstration sites (http://www.yale.edu/implicit/; http://www.tolerance.org/) in which visitors have an opportunity to examine their own implicit biases for a variety of topics. We summarized subsets of these data, the procedures used to collect and analyze them, and the methodological issues to consider in their interpretation, in Nosek et al. (2002b). Other portions of the data appear in two additional papers (Greenwald, Nosek, & Banaji, 2002; Nosek, Greenwald, & Banaji, 2002).

Studies were included in this synthesis if they met the following conditions: (1) the Implicit Association Test (IAT; Greenwald et al., 1998; Greenwald & Nosek, 2001) served as the measure of implicit evaluation; (2) the explicit measure either mapped directly on to the relative assessment of the IAT (e.g., I strongly prefer White people relative to Black people), or could easily be calculated by taking the difference between two single-item rating scales (e.g., warm-cold ratings for Black and White people); (3) the tasks assessed evaluative (good/bad) associations of the target concepts; and, (4) the

target concepts were topics that would be familiar to the average American (to be sure that the raters could make informed presentation and elaboration ratings).

With these criteria, this review included eight web-based tasks and nine tasks collected in the traditional laboratory. For tasks appearing in multiple experiments, implicit-explicit correlations were calculated independently for each experiment and then averaged.[21] This conservative analytic approach combines multiple tests of the same attitude objects into a single index to avoid artificial inflation of the sample size. Combining data resulted in IE correlations for 15 unique attitude objects: Arab-Muslims/Other people (web), creationism/evolution (2 studies), Republican/Democrat (4 studies), fruit/bugs (1 study), science/humanities (3 studies), thin people/fat people (web; 2 studies), Black children/White children (web), dark-skin/light-skin (web), math/arts (web; 3 studies), Black people/White people (web; 3 studies), Gay people/Straight people (3 studies), flowers/insects (1 study), old people/young people (web), George Bush/Al Gore (web), and self/other (web).

Rating perceived social desirability and elaboration

Ten American raters were recruited to provide estimates of (a) the extent to which college-aged people in general would be likely to conceal preferences for one concept over the other for each of the pairs, and separately (b) the extent to which college-aged people in general would have spent time thinking about each of the object pairs.

[21] Sample sizes for the individual traditional laboratory experiments ranged from 49 to 83. Web-based samples were first constrained to consist exclusively of college-aged participants (18-21 year olds) to maximize the comparability to the laboratory samples. This resulted in sample sizes ranging from 678 to 16797. If a task appeared on the Internet and in multiple experiments in the traditional laboratory, the overall implicit-explicit correlation reflected equal weighting of laboratory and web-based data.

Reliability among the judges was in excess of 0.9 for both self-presentation and elaboration estimates.

Results

Table 3 summarizes the simple zero-order correlations between implicit and explicit attitudes toward 15 attitude object pairs. Across attitude objects, the relationship between implicit and explicit attitudes varies from a paltry r of .01 (age attitudes) to a strong r of .62 (Bush-Gore attitudes). The relationship between implicit and explicit attitudes is reliable, but variable, across attitude objects.

Table 3. Summary of average implicit-explicit correlations, perceived elaboration, and perceived self-presentation demands for 15 attitude object pairs

	Implicit-Explicit correlation	Perceived elaboration	Perceived presentation demand
George Bush / Al Gore	.62	4.5	2.1
Science / Humanities	.51	4.5	2
Math / Arts	.48	4.5	2.2
Creationism / Evolution	.47	3.2	2.4
Democrats / Republicans	.45	5.3	2.6
Gay people / Straight people	.32	5.3	5.4
Flowers / Insects	.26	1.4	1.2
Thin people / Fat people	.22	4.9	5.3
Black children / White children	.21	3.8	6.1
Black people / White people	.18	4.9	6.1
Self / Others	.18	6.7	4.8
Fruit / Bugs	.18	1.3	1.3
Dark-skin / Light-skin	.14	3.8	5.4
Arab-Muslims / Other people	.09	4.3	5.5
Old people / Young people	.01	3.2	3.3
Average	.29	4.1	3.7

Implicit-explicit correspondence is a function of self-presentation and elaboration

To test whether perceived presentation demand and elaboration can predict variation in IE correspondence, we conducted a simultaneous regression on IE correspondence with self-presentation and elaboration ratings as predictors. These two factors explained 55% of the variance in IE correspondence ($F(2,12) = 7.22$, $p = .009$, adj $R^2 = .47$). Both perceived self-presentation demand ($B = .08$, SE $B = .022$, $p = .003$, $ß = .84$) and elaboration ($B = .08$, SE $B = .028$, $p = .02$, $ß = .61$) explained significant variation in IE correspondence, despite the small sample.

Given the low power of this analysis, adding the interaction term of perceived self-presentation demand and elaboration to the above model, is unlikely to produce a reliable effect. With this recognition, we conducted a second regression analysis including both the main effects of presentation and elaboration along with the interaction of those two factors. This model explained 62% of the variance in IE correspondence ($F(3,11) = 5.87$, $p = .01$, adj $R^2 = .51$). The interaction term ($B = .02$, SE $B = .016$, $p = .19$, $ß = .32$) was of moderate size but, as expected, did not reach statistical significance. Shown another way, for the seven attitude objects lowest on self-presentation demand, variation in elaboration was strongly predictive of IE correspondence ($r = .85$, $p = .01$), but for the seven attitude objects highest on self-presentation demand, the predictive power of elaboration was much weaker ($r = .29$, $p = .52$). Though underpowered, the increase in overall R^2, the moderately sized ß, and the pattern of correlations suggest that the weaker the self-presentation demands, the greater is the role elaboration might play in predicting IE correspondence. This finding should be viewed as no more than a strong speculation regarding the interplay of two moderators of the implicit-explicit relationship.

Discussion

This initial research synthesis suggests that the relationship between implicit and explicit attitudes is determined by (at least) two factors – self-presentation and elaboration. A new formulation of the nature of the relationship between implicit and explicit attitudes must be devised to handle these observations. The 'separate' attitudes view does not explain how reliable and predictable relationships can exist between implicit and explicit attitudes. And, the 'same' attitude view does not anticipate that factors other than self-presentation will moderate IE correspondence. Developing strong theory about the functional relationship between implicit and explicit attitudes requires examination of other factors that could moderate the relationship (e.g., attitude accessibility, attitude importance), and attention to the cognitive processes that give rise to these two modes of evaluation.

General Discussion

In three studies, I examined two moderators of the relationship between implicit and explicit attitudes – self-presentation and elaboration. The data support the view that multiple factors are involved in the prediction of implicit-explicit correspondence. These results challenge existing views of the relationship between implicit and explicit attitudes and have important limitations that obscure a complete understanding of the relationship between these two modes of evaluation.

Implications for views of implicit-explicit correspondence

That IE correspondence can be predicted by at least two factors – self-presentation and elaboration – is incompatible with the simple views of the relationship between implicit and explicit attitudes. The 'separate' attitudes view does not predict any

systematic relationships between implicit and explicit attitudes. This is not to say that these studies indicate that implicit and explicit attitudes are products of the same system or processes. Indeed, these two forms of evaluation could involve very different components at encoding, maintenance, or retrieval and yet still come to be related through factors external to the evaluative systems. The data reported here are agnostic to the functional processes involved in generating implicit and explicit attitudes, but do point out that a systematic relationship does exist between the outputs (evaluative responses) of implicit and explicit processes.

The 'same' view suggests that self-presentation is the only factor that will distinguish implicit and explicit attitudes. Data from Studies 2 and 3 suggest that, in addition to self-presentation, elaboration may be involved in moderating IE correspondence. As such, self-presentation explains implicit-explicit correspondence similar to the way height explains being in the NBA. Height is only one of multiple factors involved in predicting whether someone will play in the NBA. Being tall is neither necessary, nor sufficient, to guarantee that a person will be in the NBA. Lack of self-presentation demand may be necessary (these data do not clarify this proposition), but it is certainly not sufficient to explain all of the difference between implicit and explicit measurement of attitudes.

The lack of support for the two simple views of IE correspondence suggests that a new model is needed to explain this relationship. A new model will need to be sensitive to the variation in IE correspondence both across attitude objects and across situations. The data from Study 3 suggests that two factors (self-presentation and elaboration), and their interaction, may go a long way to explaining this relationship. In fact, those two

factors and the interaction predicted 62% of the variance (multiple R = .79) in IE correspondence, an impressive amount for virtually any psychological model. There are numerous other candidate factors that may also play a role in moderating IE correspondence including: attitude strength, conviction, accessibility, and knowledge of other's attitudes.[22] Below, two issues are raised that limit the conclusiveness and generality of the current studies, and outline research directions to better understand the relationship between implicit and explicit social cognition.

Multiple forms of association and dissociation

Two attitude measures are dissociated when they reveal divergent, or unrelated, evaluations of the same target. For the studies in this paper, comparing individual differences (correlations) served as the assessment of dissociation. Low IE correlations indicated dissociation, high IE correlations indicated association. An alternative, and popular, approach to examine dissociation is to compare the average magnitude of attitudes between conditions (see Greenwald & Banaji, 1995 for a review). For example, explicit attitudes toward African Americans tend to be positive, but implicit measures reveal negativity toward African Americans, showing evidence of dissociation between self-reported and automatically activated preferences. In other words, mean dissociations reveal that implicit attitudes are different than explicit attitudes on average; individual difference dissociations reveal that, relative to other people's attitudes, an individual's explicit attitudes are unrelated to his or her implicit attitudes.

[22] As you might expect, this list does not come out of thin air. In fact, there are two notable features of the items in this list. One, many have support as factors that moderate attitude-behavior relationships. And, two, we have preliminary evidence that some of these are predictors of IE correspondence from other research that is not reported in this dissertation.

The individual difference and mean difference approaches for examining association and dissociation have been used interchangeably with an implicit assumption that they are equivalent. There is, however, some evidence to suggest that observing implicit-explicit dissociation in individual differences does not necessarily translate to dissociation in mean preferences. For example, the correlation between implicit and explicit attitudes toward flowers relative to insects is modest (~ .20 in general). Yet, attitudes toward flowers relative to insects are strongly positive whether measured implicitly or explicitly (Greenwald et al., 1998). That is, implicit and explicit flower/insect attitudes are associated in the average magnitude of preference (mean comparisons), but dissociated in individual differences (correlations).

The data from Study 3 also reveal that implicit-explicit correlations can be quite strong even when mean preferences are largely dissociated. Stronger self-reported (explicit) positivity toward creationism corresponds with stronger implicit positivity toward creationism ($\underline{r}$ = .47; Study 3), suggesting that implicit and explicit creationism/evolution attitudes are associated. However, a comparison of implicit and explicit mean preferences suggests dissociation. Yale undergraduates report liking evolution much more than creation (M = 2.39, SD = 3.38; 0 = equal liking), but implicitly show a slight preference for creation over evolution (M = -33.5ms, SD = 217; 0 = equal liking). In other words, implicit and explicit creationism/evolution attitudes are associated in individual differences (correlations), but dissociated in average magnitude of preference (mean comparisons).

These two examples suggest that associations assessed in individual differences need not be identical to associations assessed as comparisons of mean preferences. But, is

there any relationship between these two methods of assessing dissociation? As correlations between implicit and explicit attitudes decline, will a larger dissociation in mean preferences between implicit and explicit attitudes emerge? Likewise, for attitude objects that elicit very similar liking whether implicit or explicit measures are used, will the correlations between implicit and explicit attitudes also tend to be high? If they are not related, then two separate models may be needed to understand each type of dissociation.

I conducted a preliminary comparison of mean and individual difference dissociations with some of the data described in Study 3. For individual differences, the correlations continued to serve as the measure of association (IE_{corr}). For mean comparisons, the index of association/dissociation was the absolute value of the difference between mean implicit attitude toward a target concept and mean explicit attitude toward the same concept.[23] Because implicit and explicit attitudes are assessed on different scales (the former in milliseconds, the latter on ratings scales), each was converted to an effect size before calculating the difference.[24]

The index of dissociation for mean differences is reflected by the following formula:

$$IE_{md} = abs(\underline{\rho}_{implicit} - \underline{\rho}_{explicit})$$

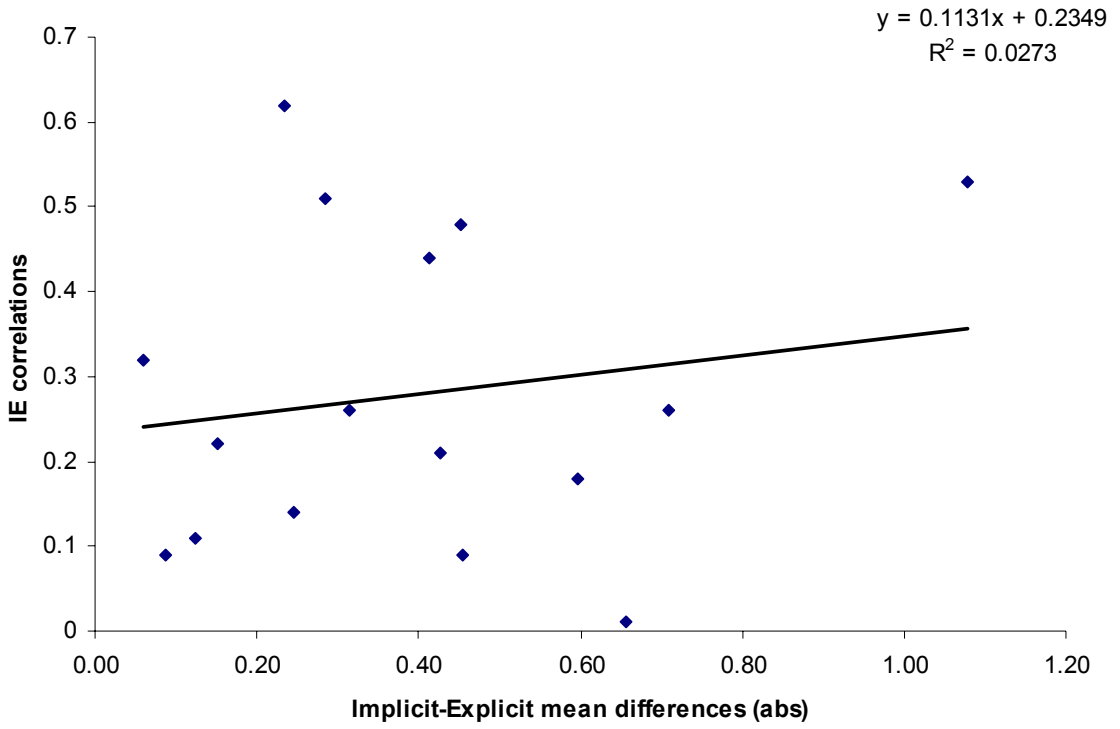
[23] The absolute value is important because dissociation is observed when implicit and explicit attitudes are different, but it is irrelevant which one is more positive or negative than the other.

[24] In conducting this analysis, I realized that the traditional approach for calculating effect size (i.e., Cohen's d) carried an assumption about homogeneity of variances that was not appropriate for this analysis (and probably many others). I generated a variation of Cohen's d that does not depend on the assumption of homogeneity of variances that I refer to as $\underline{\rho}$ for lack of a better symbol. A description of the logic and calculation of $\underline{\rho}$ is in Attachment B. Essentially, $\underline{\rho}$ is a standardized measure of the magnitude of an effect that uses a single index of variability for all measures that use the same scale. For example, all explicit attitude ratings that were made with a single 9-point Likert scale were combined to create a single index of variability of responding. This approach removes the assumption of homogeneity of variances across the attitude objects because the same variance is used across measures.

When IE_{md} is 0, then there is a perfect association between the magnitude of the implicit attitude and the explicit attitude. For values > 0, the higher the value, the greater is the dissociation between mean implicit and explicit attitudes.

If dissociations in mean differences are the same as dissociations in individual differences, then we should expect a strong negative correlation between IE_{md} and IE_{corr} because greater dissociation is reflected by larger values in the former and smaller values in the latter. Figure 3 presents a scatterplot of the relationship between IE_{md} and IE_{corr} using data collected from the laboratory and the Internet (as described in Study 3).

Figure 3. Scatter-plot of the relationship between correlations (IE_{corr}) and mean difference relationships (IE_{md}) between implicit and explicit attitudes.



Contrary to what would be expected if these two modes of association/dissociation were comparable and consistent, the relationship was not strongly negative. In fact, it was slightly positive ($\underline{r}$ = .16).[25]

The two dominant approaches to examining association and dissociation are not related, at least in these data. This suggests that the production of differences in implicit and explicit attitudes at the group level might involve very different mechanisms than production of differences across individuals. In this dissertation, we have identified two factors, self-presentation and elaboration, that predict association and dissociation in individual differences. Further investigation is needed to see whether those factors are also predictive of variation in mean level dissociation.

Multiple measures of implicit and explicit attitudes

Of the many ways to measure the association between a concept and an evaluation, this dissertation involved only two, the IAT and self-reported affective judgments. Each of these measures has numerous features that might affect the process and result of an evaluation. Evaluations tapped by the IAT, for example, are uncontrollable in the sense that the respondent cannot opt to express positivity or negativity. Also, respondents do not seem to have conscious access to the evaluations except as performing the task itself reveals them to him or her. Even so, respondents are (usually) aware that evaluations are being measured as well as the target of those evaluations because the task itself is quite transparent. This is not true for another implicit measure, subliminal priming. In that task, respondents are not only unaware of the content of their evaluations but they are also unaware of the target of evaluation or even that evaluations are being assessed. Clearly,

[25] When the outlier (creation-evolution attitudes) is removed, the relationship is slightly negative ($\underline{r}$ = -.11) but still close to 0.

the two measures used in this dissertation do not represent the breadth of methods available to capture the association between a concept and an evaluation. As a consequence, the generality of these findings may be limited to the relationship between these two measures and not a more general explanation for the relationship between evaluations that occur within and without conscious awareness and conscious control.

A more complete analysis of this relationship will require less of an emphasis on the two constructs – implicit and explicit attitudes – and more of an emphasis on the processes involved in generating evaluations consciously and unconsciously. That is, the study of the relationship between implicit and explicit attitudes could be reframed as the study of how implicit and explicit processes alter the expression of attitudes. This shift in emphasis will help solve some of the perplexing issues regarding the relationship between various measures of evaluation. Bosson et al. (2000), for example, observed that implicit measures of self-esteem show little to no relationship with one another. While some of the observed lack of relationship between implicit measures is surely due to measurement error (Lane, Brescoll, & Bosson, 2001; Cunningham et al., 2001), it is quite likely that variations in the measurement procedures affect the way that implicit evaluations are elicited from the mind, and thus assess different aspects of the network of cognitive associations. Each of the implicit measures used by Bosson involved unique procedures to tap implicit self-esteem. For example, the name-letter effect demonstrates liking for self based on rating liking of the letters of one's own name versus other letters, while the IAT taps associations between self and good by requiring rapid sorting of words into groups. The fact that both are implicit measures, in the sense that they tap self-associations without awareness or control, should not lead to an automatic assumption

that these disparate methods will reveal identical results. Rather, there are many other features of measurement that might dramatically alter one's evaluation of oneself.

A successful analysis will clarify how various mental processes (e.g., awareness of the target of evaluation, control over the expression of evaluation) affect the generation and presentation of attitudes. The variety of measurement tools that exist emphasize the use of different mental processes, and provide an opportunity to observe how and when each feature of the evaluation process plays a role in altering the ultimate response.

A first step in this investigation will be to identify the features of measurement that are meaningfully related to the process of generating an evaluative response. Table 4 presents a simplistic taxonomy of a variety of methods for measuring evaluation (in rows) by a selection of methodological features that might be involved in generating an evaluation. The methods are sorted from the most direct measure at the top (respondent is aware of all aspects of the evaluation situation and can control his/her response), to the most indirect measure at the bottom (respondent is unaware that evaluations are being tapped [or even of the target of evaluation] and cannot control his/her response). The values of 1, .5, and 0 indicate whether the given measure has the feature (1), seems to partially have the feature (.5), or does not have the feature (0).[26]

For example, the first column asks whether the evaluative response is under the intentional control of the respondent. This is at least partially true for the first four measures that all involve self-report (typically considered 'explicit' though they vary in how direct they are), but the last five measures do not provide respondents the opportunity to control their evaluative responses. If this factor is involved in

[26] The 0, .5, 1 ratings are not established in the psychological literature; they reflect my personal judgment of each measurement type.

distinguishing evaluative responses across measures, then we might expect the measures to be more strongly related to others that share that quality than ones that do not.

Table 4. A simple taxonomy of attitude measures (rows) and methodological features of measurement (columns). 1 indicates 'Yes' to row question, 0.5 indicates 'Partially', 0 indicates 'No'

	Controllable?	Aware that evaluations are being assessed?	Aware of the target (source) of evaluation?	Reported (verbal or written)?	Is response free of distraction?	No other cog demands present during eval.?	Total
Reported affect toward a specific group	1	1	1	1	1	1	6
Reported affect toward a specific group (speeded)	0.5	1	1	1	1	0	4.5
Reported affect toward a target (critical masked)	1	1	0	1	0	1	4
Reported beliefs that are apparently not affective	1	0	1	1	0	1	4
IAT / GNAT	0	1	1	0	1	0.5	3.5
Supraliminal priming (evaluative judgments)	0	1	1	0	1	0	3
Supraliminal priming (lexical decision)	0	0	1	0	1	0	2
Subliminal priming (evaluative judgments)	0	1	0	0	0	0	1
Subliminal priming (lexical decision)	0	0	0	0	0	0	0

This is not meant to be a comprehensive list of measures or methodological features, it is merely intended to illustrate that there are a variety of features of attitude measurement that may alter responses, and as a consequence, that predict how and when the elicited evaluations will be associated or dissociated across measures.

Understanding the processes that conspire to generate or alter an evaluative response may then help to identify which measures will be most useful for predicting perceptions, judgments, or behaviors in particular situational conditions. For example, an employer's explicit racial attitudes may predict a hiring decision when the employer attends to the applicant's ethnicity, unless the employer is unaware that the applicant's ethnicity may be playing a role in shaping the overall evaluation of the application. In that case, a more

indirect measure of racial attitudes (one that does not reveal the target of evaluation to the respondent) may be a better predictor of the ultimate judgment. Systematic investigations of the processes engaged by attitude measurement will clarify how each component of the evaluative system participates in the generation of an evaluation, and how and when factors such as elaboration or self-presentation will result in divergent responses between more implicit and more explicit measures.

Conclusion

The relationship between attitudes that lie largely outside conscious awareness and those that reside within the conscious mind is fascinating because it distinguishes what exists in our minds from what we experience as our minds. That there is not a perfect relationship between what one intends, endorses, and reports as his or her attitude with what emerges from the mind without intention, endorsement, or self-report is as intriguing as it is surprising. Introspection cannot resolve this mysterious relationship because many components of the evaluative process are simply outside of the mind's eye. We can see as plain as data that the relationship between implicit and explicit attitudes is, in part, a function of the effort to present oneself in a positive light to others, and of how much effort has been expended thinking about the attitudes themselves. Knowing this provides an opportunity to have some control over those parts of the mind that proceed without us.

References

Banaji, M. R. (in press). The opposite of a great truth is also true: Homage to koan #7. In J. Jost, D. Prentice, and M. Banaji (Eds.) The Yin and Yang of Social Cognition. American Psychological Association.

Banaji, M. R. (2001). Implicit attitudes can be measured. In H. L. Roediger, III, J. S. Nairne, I. Neath, & A. Surprenant (Eds.), The nature of remembering: Essays in honor of Robert G. Crowder (pp. 117-150). Washington, DC: American Psychological Association.

Banaji, M. R., & Hardin, C. D. (1996). Automatic stereotyping. Psychological Science, 7, 136-141.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motive. Psychological Bulletin, 117, 497-529.

Bellezza, F. S., Greenwald, A. G., & Banaji, M. R. (1986). Words high and low in pleasantness as rated by male and female college students. Behavior Research Methods, Instruments, & Computers, 18(3), 299-303.

Blair, I. V. (2001). Implicit stereotypes and prejudice. In G. B. Moskowitz (Ed). Cognitive Social Psychology: The Princeton Symposium on the Legacy and Future of Social Cognition (pp. 359-374). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Blair, I. V., & Banaji, M. R. (1996). Automatic and controlled processes in stereotype priming. Journal of Personality and Social Psychology, 70, 1142-1163.

Blair, I. V., Ma, J., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of automatic stereotypes through mental imagery. Journal of Personality and Social Psychology, 81, 828-841.

Bosson, J.K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? Journal of Personality and Social Psychology, 79(4), 631-643.

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.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. Journal of Personality and Social Psychology, 81, 800-814.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56(1), 5-18.

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.

Draine, S. (1998). Inquisit software. Millisecond: www.millisecond.com.

Eagly, A. H., & Chaiken, S. (1998). Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds) The Handbook of Social Psychology, Vol. 1 4th ed. (pp. 269-322). New York, NY: McGraw-Hill.

Fazio, R. H. (1986). How do attitudes guide behavior? In R. M. Sorrentino and E. T. Higgins (Eds.) Handbook of motivation and cognition: Foundations of social behavior. (pp. 204-243). New York: The Guilford Press.

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(6), 1013-1027.

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(2), 229-238.

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, 85-93.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2002). Scoring Procedures to Improve Implicit Association Test Measures. Unpublished manuscript.

Guerin, B. (1993). Social Facilitation. New York: Cambridge University Press.

Kim, D-Y., & Greenwald, A. G. (1998, May). Voluntary controllability of implicit cognition: can implicit attitudes be faked? Paper presented at meetings of the Midwestern Psychological Association, Chicago.

Lane, K. A., Brescoll, V.B., and Bosson, J.E. (2001). The relationship between implicit and explicit identity. Poster presented at the 2001 conference for the American Psychological Society. Toronto, Canada. June, 2001.

Lemm, K. M., & Banaji, M. R. (2001). Personal and social motivation to respond without prejudice: Implications for implicit and explicit attitude and behavior. Unpublished manuscript.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? Journal of Personality and Social Psychology, 72(2), 275-287.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the Effects of Accountability. Psychological Bulletin, 125(2), 255-275.

Logan, G.D. (1988). Toward an instance theory of automatization. Psychological Review, 95(4), 492-527.

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. Journal of Personality and Social Psychology, 81, 842-855.

Mitchell, J. A., Nosek, B. A., & Banaji, M. R. (2002). Contextual variations in implicit evaluation. Unpublished manuscript.

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. (in press). (At least) two factors moderate the relationship between implicit and explicit attitudes. In R.K. Ohme & M. Jarymowicz (Eds.), Natura Automatyzmow, Warszawa: WIP PAN & SWPS.

Nosek, B. A. & Banaji, M. R. (2002). The Power of the Immediate Situation: Gender Differences in Implicit Math Attitudes. Poster presented at the 2002 conference for the Society of Personality and Social Psychology. Savannah, GA. February 2, 2002.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. Social Cognition, 19(6), 625-666.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002a). eResearch: Ethics, security, design, and control in psychological research on the Internet. Journal of Social Issues, 58(1), 161-176.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002b). Harvesting intergroup implicit attitudes and beliefs from a demonstration website. Group Dynamics, 6(1), 101-115.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002c). Math = male, me = female, therefore math ≠ me. Journal of Personality and Social Psychology, 83(1), 44-59.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2002). Administering and interpreting the implicit association test: Lessons from some large data sets. Unpublished manuscript.

Paulhus, D. (1982). Individual differences, self-presentation, and cognitive dissonance: Their concurrent operation in forced compliance. Journal of Personality and Social Psychology, 43(4), 838-852.

Petty, R. E. & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 19, pp. 123-205). New York: Academic Press.

Petty, R. E. & Wegener, D. T. (1999). The elaboration likelihood model: Current status and controversies. In S. Chaiken & Y. Trope (Eds.), Dual-process Theories in Social Psychology (pp. 41-72). New York: Guilford Press.

Schwarz, N., Groves, R. M., & Schuman, H. (1998). Survey Methods. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), The Handbook of Social Psychology (pp. 143 – 179). New York: Oxford University Press.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? Journal of Personality and Social Psychology, 70, 893-912.

Smith, E. R., & DeCoster, J. (1999). Associative and rule-based processing. In S. Chaiken & Y. Trope, Eds., Dual-Process Theories in Social Psychology (pp. 323-360). New York: Guilford.

Tetlock, P. E. (1983). Accountability and the complexity of thought. Journal of Personality and Social Psychology, 45, 74-83.

Tetlock, P. E., Skitka, L., & Botteger, R. (1989). Social and cognitive strategies for coping with accountability: Conformity, complexity, and bolstering. Journal of Personality, 57, 632-640.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. Psychological Review, 107(1), 101-126.

Attachment A – How can implicit attitudes be malleable and slow to change at the same time?

One of the intriguing issues initiated by recent demonstrations of the malleability of implicit attitudes (e.g., Dasgupta et al., 2001; Lowery et al., 2001; Nosek & Banaji, 2002) is its apparent conflict with assertions that implicit attitudes are slow to change (Wilson et al., 2000). How can implicit attitudes be malleable and slow to change at the same time?

First, it is important to note that situational malleability is different than change as normally understood. A change in attitude requires that the shift is lasting and not dependent on the situational manipulation that elicited the shift. For example, exposure to a female experimenter produces less negativity in women's implicit math attitudes compared to having a male experimenter or (apparently) no experimenter at all (Nosek & Banaji, 2002). However, that situational shift would need to persist at a later time when the female experimenter is no longer present to properly be interpreted as a change (there is a great deal of ambiguity for the middle ground between malleability and change that will not be resolved in this brief exposition).

Second, it is conceivable that implicit attitudes could show a great deal of situational variability (malleability) and, at the same time, long-term consistency (slowness to change); whereas, explicit attitudes may show less situational variability and, at the same time, amenability to rapid and relatively permanent change. These possibilities are consistent with, and perhaps even anticipated by, a connectionist interpretation of attitudes.

Smith and DeCoster (1999) outlined a dual-process model of attitudes in which our use of the terms implicit and explicit was discussed as associative and rule-based respectively. In a connectionist framework, attitudes are constructed on-line through a network of nodes that are linked together with a variety of connection weights. The inputs to the connectionist network derive from environmental contingencies and those inputs traverse the nodal network to generate an output, which is the evaluative response. So, variation in the pattern of inputs can give rise to dramatic variation in the ultimate evaluation depending on the pattern of connections and the values of the connection weights between nodes (see Smith, 1996 for more detail). You might consider the inputs to represent the current situation and the connection weights to represent the person. Therefore, any evaluative response is the output of the interaction between the person and the situation. Sound familiar?

In the connectionist framework, the weights between nodes are slowly shaped by repeated activations that change the weights by strengthening connections that are repeatedly activated. Also, since the situation drives the pattern of inputs, which in turn affects the pattern of activation and the resulting evaluation, a situational manipulation can dramatically affect the resulting evaluation without dramatically changing connection weights in the nodal network. Significant change in the connection weights requires repeated activations over time. In this sense, an attitude can be both highly malleable (many possible resulting evaluations depending on the nature of the inputs) and slow to

change (connection weights that are stable and require many situational exposures before significant shifts will occur). This is an interpretation of how implicit attitudes might operate.

Smith and DeCoster (1999) also outlined a rule-based account, which can explain the operation of explicit attitudes – greater situational stability but easier to change. In a rule-based account, attitudes are stored as propositions, or rules (e.g., donuts = bad). A single sufficiently compelling exposure to contrary evidence (e.g., eating a Krispy Kreme donut for the first time) can be sufficient to alter the explicit attitude because all that is needed is an alteration of the proposition (i.e., donuts = good). At the same time, situational manipulations may have less of an impact on explicit attitudes because the person merely needs to recall the proposition, which can be stored and retrieved as a unit. Therefore, explicit attitudes can be both situationally stable (relatively speaking) and easy to change (relatively speaking).

Attachment B – Removing the assumption of homogeneity of variances from the Cohen's d estimate of effect size

Cohen's d, a popular standardized estimate of the magnitude of an effect, is defined by the following formula:

$$d = (\mu_1 - \mu_2)/\sigma$$

where μ_1 and μ_2 are estimated by M_x and M_y respectively. And, σ is estimated by $s_{x\text{-}y}$.

Dividing the difference of the group means by the standard deviation allows the magnitude of the difference between the two means to be compared to other effects regardless of the scaling characteristics of the original measures. This approach, however, assumes homogeneity of the variances that are used to estimate σ.

This may not make immediate sense because in typical scenarios that compare multiple effect sizes, different scaling parameters are used for each of the effects. This makes it impossible, even nonsensical, to attempt to test the homogeneity assumption. The current application for this paper, however, demonstrates how the homogeneity assumption is functioning when we estimate effect sizes.

Table 5 presents explicit attitudes for five attitude objects assessed using the same scale. They range from a high of .6 on that scale (preference for young over old) to a low of .31 on that scale (preference for White over Black). Comparing across these measures does not require standardization because they were all generated with the same scale. As such the 5 attitude objects are arranged in descending order from the largest preference difference to the smallest.

Table 5. Explicit attitudes for five attitude objects (data collected from the Internet)

	Explicit Mean	Explicit SD	Cohen's d	Overall SD	Effect size ρ
Old people / Young people	0.60	0.79	0.76	1.19	0.50
Math / Arts	0.45	1.44	0.31	1.19	0.38
George Bush / Al Gore	0.39	1.55	0.25	1.19	0.33
Others / Self	0.34	1.22	0.28	1.19	0.29
Black people / White people	0.31	0.70	0.45	1.19	0.26

The second column shows the standard deviations for the explicit attitudes. It is clear that they are not consistent with one another (differences are meaningful considering that each of these data sets numbers in the tens of thousands). Respondents are quite consistent in attitude reports for young over old and white over black, but are much more variable in attitude reports for arts over math and Al Gore over George Bush. This is not surprising. People have and willingly report a wide variety of attitudes toward academic domains and political candidates, but are much less extreme in attitude reports for age and ethnic social groups.

As a consequence, the estimates of effect size for each of these attitude objects, shown in column 3, do not follow the pattern of magnitudes reflected in the means of column 1. Based on the Cohen's d estimates, it would appear that explicit preferences for White over Black are actually stronger than preferences for arts over math, when the mean estimates show the opposite. The problem is clearly in the lack of homogeneity of the variances across attitude objects.

Because I would like to compare the magnitude of the explicit attitudes for these attitude objects with the magnitude of the implicit attitudes for the same objects, I need to have standardized estimates of the magnitude of the effects, but without holding on to the homogeneity assumption.

To do this, I adapted Cohen's d in this way. For all attitude objects that used the same measure (e.g., a 9-point semantic differential rating scale), I created a single index of variability by combining all of the attitude objects and calculating the standard deviation. This way, the estimate of variability represents the variability of responses to the family of attitude objects on that 9-point scale, not to any particular attitude object. This way, the variation in consistency of responding across attitude objects is ignored in favor of a simple standardized estimate of the magnitude of the effect. The standardization in this case includes all assessments that used the same scale for calculation. So, all attitude objects assessed with the same scale shared the same index of variability in the calculation of the magnitude of effect. You can see the overall standard deviation in column 4 of Table 5, and the resulting estimate of effect size in column 5.[27]

Therefore, the adjusted effect size calculation is:

$$\underline{\rho} = (M_x - M_y) / s_{overall}$$

One restriction that I added in the estimation of the overall standard deviation for each measurement scale was to assume the same sample size for each attitude object. If that weren't the case, then the overall standard deviation would be influenced more by those datasets with larger n's and underrepresent datasets with small n's.

Note:

The overall standard deviation can be calculated without access to all of the original data as long as the mean's, standard deviations, and sample sizes of all of the individual attitude objects are available. A shortened version of this proof is below:

[27] Note that this effect size calculation will generally produce lower effect size estimates (on average) than the standard Cohen's d approach because the overall variability will typically exceed the average of the individual variabilities.

Assume that we are trying to calculate the overall standard deviation for 3 attitude objects (X, Y, and Z) that were all rated on the same 9-point semantic differential scale. M is the overall mean across the three attitude objects, and M_x is the mean of X.

$$s_{overall} = [\Sigma(X - M)^2 + \Sigma(Y - M)^2 + \Sigma(Z - M)^2] / (n_x + n_y + n_z - 1)$$

$$= [\Sigma(X - M)^2] / (n_x + n_y + n_z - 1) + [\Sigma(Y - M)^2] / (n_x + n_y + n_z - 1) + [\Sigma(Z - M)^2] / (n_x + n_y + n_z - 1)$$

We set the above equation equal to the equation below and try to solve for the constants c, d, and e.

$$= c*s_x^2 + d*s_y^2 + e*s_z^2$$

which is equivalent to

$$= c*[\Sigma(X - M_x)^2] / (n_x - 1) + d*[\Sigma(Y - M_y)^2] / (n_x - 1) + e*[\Sigma(Z - M_z)^2] / (n_x - 1)$$

Because of the addition principle we can separately solve for c, d, and e.

After solving algebraically we get the following equation, which requires that we know only the variance of x (s_x^2), the mean of x (M_x), the sample sizes of x, y, and z ($n_x + n_y + n_z$) and the overall mean (M). We do not need to know anything about the original data points (X).

$$c*s_x^2 = [\ s_x^2*(n_x - 1) + M_x^2*n_x - 2*M*\ M_x*n_x + M^2*n_x\ ] / (n_x + n_y + n_z - 1)$$

Replacing the x elements of the equation with y elements and c with d will provide the estimate for the second part of the overall SD estimate, then the same procedure for the z and e elements. Summing $c*s_x^2$, $d*s_y^2$, and $e*s_z^2$ equals the overall standard deviation across attitude objects X, Y, and Z.

Appendix A – Stimuli for Studies 1 and 2

Category labels and stimuli used for Implicit Association Tests.

Study 1

Good. celebrating, pleasure, happy, friendly, joyful, loving, beautiful, smiling
Bad. horrible, angry, terrible, noxious, tragic, unpleasant, hate, destroy
Flowers. Daffodil, Daisy, Rose, Sunflower
Insects. Bugs, Caterpillar, Cricket, Fly
Gay. Gay, Homosexual
Straight. Straight, Heterosexual
Science. Science, Biology, Physics, Chemistry, Biochemistry, Engineering, Neuroscience, Astronomy
Humanities. Philosophy, Arts, Humanities, Spanish, English, History, Latin, Music
Democrats. Democrat, Liberal, Left-wing, Al Gore, Bill Bradley
Republicans. Republican, Conservative, Right-wing, John McCain, George Bush
African American. Faces of Black Americans (available at http://www.briannosek.com/)
White American. Faces of White Americans (available at http://www.briannosek.com/)

Study 2

Good: celebrating, pleasure, happy, joyful, loving, beautiful, smiling, glee, cheerful
Bad: horrible, angry, terrible, noxious, tragic, unpleasant, hate, destroy, brutal, nasty
Mainstreaming: Integrative Services, Same School, Integrated, All Together, Mixed Classes, Together, Mainstreaming
Special Education: Special Services, Unique Schools, Keep Separate, Separate Tracks, Distinct Tracks, Separate, Special Education
Conservation: Maintain Habitat, Conservation, Save Wildlife, Restrain Production, Protect Environment, Reduce Consumption, Environment
Development: Profit, Development, Provide Energy, Industry, Foster Wealth, Enhance Production, Promote Growth

Appendix B – Elaboration materials for Study 2

Mainstreaming versus Special Education reading materials

The following reading is a summary of two approaches for educating students with disabilities in America. Please read the following carefully and be prepared to answer questions about the reading as well as engage in a critical thinking task about the various advantages and disadvantages about the two different approaches.

Public education and health officials in the United States identified approximately 5.4 million infants, toddlers, children, and youth as disabled in 1994. Some of the most frequently reported disabilities are speech or language impairments; mental retardation and other developmental disorders; serious emotional disturbance; and specific learning disorders, such as memory disorders. Other disabilities include hearing, visual, or orthopedic impairments, autism, and traumatic brain injury. Also an increasing number of children in the United States are identified as having attention-deficit hyperactivity disorder (ADHD).

The America education system is embroiled in a debate about the appropriate approach for the education of children with disabilities. There are two distinct methods that have dominated the education of children with disabilities. The first is special education, which involves developing unique programs for children with disabilities to meet their particular needs in a safe and specially designed environment. This involves classes in which children with disabilities are separate from other children so that the instruction and structure of the classroom can be specifically tailored to meet their needs. The second approach is called mainstreaming, or inclusion. The mainstreaming approach integrates children with disabilities into standard classroom environments with children who do not have disabilities. This involves providing a variety of activities and experiences so that all students can participate in the regular classroom environment of their neighborhood school.

Specific programs for people with disabilities (either special education or mainstreaming) are relatively new. Historically, people with disabilities were placed in hospitals, asylums, or other institutions that provided little, if any, education. French physician and educator Jean Marc Gaspard Itard was one of the earliest teachers to argue that special teaching methods could be effective in educating disabled children. In 1848 French psychologist Edouard Seguin, who had studied with Itard, immigrated to the United States and developed several influential guidelines for educating children with special needs. Seguin's education programs stressed the importance of developing independence and self-reliance in disabled students by presenting them with a combination of physical and intellectual tasks.

During the 18th and 19th centuries in the United States, educators opened a variety of special schools for disabled students. Special education classes within regular school programs began at the beginning of the 20th century.

Special education in the United States has been most influenced by parent and professional advocacy groups, federal laws, national trends in conventional education, and the civil rights movement. Despite mandatory school attendance laws for all children, many U.S. schools excluded children with disabilities as recently as the 1960s.

In 1975 the Congress of the United States passed the Education of All Handicapped Children Act, which extended the right of equal educational opportunity to children with disabilities. This federal legislation was reenacted in 1990 as the Individuals with Disabilities Education Act (IDEA), and has undergone numerous amendments. This is the historical background that has now set the stage for our present day debate on how we will deal with the education of children with disabilities in the future.

Special Education

Today, special education services are delivered in many different settings and facilities. This may mean in some cases that a child would be sent to a separate school, hospital, or modified facility within their own home to receive education services from different teachers and specialists. Students in these programs do not interact with students without disabilities in the educational environment. Special educators plan and evaluate instruction in an individualized manner to accommodate each child's unique strengths and weaknesses. And special education services make use of an extraordinary array of instructional methods and settings that make it possible for all students to learn. This means that the child will have a program of instruction that is unique and targets his or her specific needs. Often a child in this environment will spend most of the day working one-on-one with a specific teacher or specialist. The focus of special education is to help disabled individuals with life-training skills. Depending on the student's abilities this might mean developing the skills to have a job, live alone, be able to read, or be able to navigate one's environment. The progress of the program is dependent on the individual student and the student is not subjected to the method or pace of a typical school setting.

Mainstreaming

Current mainstreaming approaches emphasize the creation of activities to ensure that every child, including children with disabilities, can participate and be successful in a single, integrated school setting. This involves bringing students with disabilities into a typical classroom so that they can work and play along side all varieties of children. In this form of education, teachers trained in special education collaborate with the teachers of the whole classroom to generate and design activities that can be inclusive of all students. Mainstreaming attempts to integrate children with disabilities into typical classroom environments to maximize the opportunity for the children to have a school experience similar to children without disabilities. The goal of mainstreaming is to normalize disabilities as just another feature of being human for both children with, and without, disabilities. Often, children with disabilities cannot maintain the same pace of learning and activity as the other students in the classroom. As a consequence, the focus of mainstreaming children with disabilities is to improve their social and interpersonal skills rather than necessarily provide the best possible educational environment.

The debate

While both approaches have some distinct advantages there are a variety of concerns on both sides. Some people feel that the fact that the special education approach separates children with disabilities from others constitutes unequal education. Opponents of special education contend that our educational system should try to integrate children of all types as much as possible. However, others contend that integrating children with disabilities into typical classroom environments does more harm than good. They argue that inequity is created by putting children with disabilities into an environment in which they cannot keep up.

Supporters of mainstreaming counter that inclusive settings offer expanded possibilities for educational, social, and emotional growth among all students. They worry that segregating students with special education courses will alienate them not just educationally but also publicly and communally in society. Special education supporters argue that students with disabilities may develop poor self-image and self-confidence if they are constantly comparing themselves to students without disabilities in an integrated environment. Further, children with disabilities may be ostracized by their classmates because of their differences and special challenges.

Special education advocates believe that placing students into independent settings will allow them to develop at their own pace and without concern for how their performance compares to other children. Further, special education supporters argue that the introduction of students with disabilities into regular classrooms maybe disruptive or require excess amounts of time. They worry that inclusion can slow educational progress for all and that children with special needs would be better suited in a separate setting where they can receive individualized attention. They feel that these children were born with obvious deficiencies and need specialized care that cannot be taken care of in a normal setting. Finally, special education advocates argue that including disabled children into regular classes also costs much more money a specialist or trained assistant is needed for every child with disabilities in each classroom. Supporters of mainstreaming reject these arguments saying that class disruptions and slow downs are minimal compared to the advantages for children with disabilities in having an opportunity to interact with all of their peers. In addition, mainstreaming advocates suggest that educating children with disabilities should not have an over-emphasis on knowledge and skill development. Instead, the emotional and social development is most important for these children.

Consensus has not been achieved for the best approach to educating America's children with disabilities. It is likely that any resolution to this debate is years away.

Conservation versus Development reading materials

The following reading is a summary of the debate concerning the competing needs of global conservation and development. Please read the following carefully and be prepared to answer questions about the reading as well as engage in a critical thinking task about the various advantages and disadvantages about the two different approaches.

The issue of environmental protection (conservation) versus growth and development (development) is a very important issue in modern America. The extraction of energy and natural materials to help society function can often put natural lands and wildlife at risk for damage or destruction. Some people would like to prevent development of many regions of United States in order to protect and preserve the natural habitats. However, the ever-present need for energy in the United States leads to a confusing and hotly debated topic. How should the priorities of conservation and development be weighed against one another? Under what conditions should each prevail?

From 1989 to 1999, according to the U. S. Energy Information Administration (a division of the Department of Energy) the amount of power generated in the U.S. rose 9.6 percent but the amount of power consumed in the U.S. increased 24.7 percent. Society's power needs far outstrip the current development of energy sources. The problems of energy shortages are quite evident from the recent crisis in California and are likely to appear more and more frequently in the coming years. A 'hot' debate right now between the relative priorities of conservation versus development is in regard to the coastal plain in the Arctic National Wildlife Refuge.

The "Coastal Plain" is a 1.5 million acre tract on the North Slope of Alaska. It is a designated area within the 19 million acre area known as the Arctic National Wildlife Refuge (ANWR). This region was designated by Congress in 1981 as requiring special study to determine its oil and gas potential along with an investigation of the effects of development on the local environment. In 1987, the Department of Interior recommended developing the region as an energy source but congressional and presidential authorization is required for the Coastal Plain to be open for development.

Some people argue that the answer to some of our nation's energy needs lie below the surface of ANWR. Those who are in favor of developing this region as an energy source have several arguments. First, the United States imports over 57% of the petroleum it uses. These oil imports cost more than 100 billion dollars a year. These figures are rising and could exceed 65% imports by the year 2020. Some speculate that ANWR production could replace Saudi oil imports for up to 30 years. If so, the financial benefits of developing ANWR as an energy source are profound.

Secondly, supporters of development point out that Prudhoe Bay, located 60 miles to the west of ANWR, has been operating for over 20 years and has produced in excess of 10 billion barrels of oil. However, the present output from Prudhoe Bay area fields has declined to 1.1 million barrels per day, and is continuing to decline. Development advocates argue that the Coastal Plain of ANWR is America's best bet for the discovery of another giant oil and gas region in North America. Between 250,000 and 700,000 jobs may be created if the Coastal Plain is developed. Federal revenues would be enhanced by billions of dollars from bonus bids, lease rentals, royalties, and taxes.

Finally, supporters of development say advancing technology has minimized the "footprint" of Arctic oil development. That means that development is not nearly as costly to the local environment as it is made to seem by supporters of conservation. As such, proponents argue that development can coexist with Alaska's Arctic wildlife. Because of technological advances the energy resources can be extracted with minimal effects on the surrounding habitat.

Proponents of conservation disagree that the development will not harm the Alaskan habitat. They argue that oil development of the Coastal Plain will require hundreds of miles of pipelines and roads, numerous drilling pads, production wells, waste pits, airstrips, and dorms. Whether or not any spills occur, that development will be disruptive enough. Conservation supporters say that oil development will disrupt the herd of Porcupine River Caribou, who migrate over 400 miles each year to the same place to give birth to their young. The Coastal Plain is also sacred ground to the Gwich'in Indians, a 20,000 year-old native culture whose subsistence lifestyle depends upon the harvest of caribou. They depend on the animal for food, clothing, medicine and their cultural lore. If the herd is disrupted it could cause a decline in caribou population and ultimately jeopardize their traditional way of life. The Coastal Plain is also home to unique and abundant wildlife: wolves, polar bear, musk ox and wolverines.

Finally, conservationists also state that there is no guarantee that the energy resources are so significant in this region. They argue that there may be only about a six-month supply given the current U.S. consumption rates. Further, because of development time it might take up to 10 years for any oil to begin to flow out of the region.

The information about the Coastal Plain of the Arctic National Wildlife Refuge is a microcosm of the larger debate that exists in our country today between conservation and development. How are the energy needs of the United States weighed against the desire to protect and conserve natural habitats? What is the appropriate approach for deciding the relative priority of these two ideas?