The opposite of a great truth is also true

Homage to Koan #7

Mahzarin R. Banaji

Harvard University

Author's Acknowledgments: This essay is written in honor of William James McGuire, on the occasion of his Festschrift. The issues in this paper arise from the work of six students, past and present: Irene Blair, Nilanjana Dasgupta, Curtis Hardin, Jason Mitchell and Brian Nosek whose view of the power of the immediate situation challenge theory about the nature of implicit social cognition; William Cunningham whose belief that personality is the detector of patterns across social situations produced new tests and changed the interpretation of existing evidence. The research and writing were supported by grants from the National Science Foundation (SBR 9709924) and the National Institute of Mental Health (MH 57672) as well as grants from Yale University and Harvard University. A rare instance of genuine oppositional thinking in my work is a collaboration with John Jost in which he advanced the idea that the extreme strength of ingroup favoritism also suggests the presence of its opposite, outgroup favoritism. I am grateful to him for comments on this manuscript.

If social psychology in the 20th century has revealed anything stunning about human nature it is this: That individuals are shaped by social circumstances far more than they or their observers are able to recognize. Mainly in retrospect, I see my career as a textbook case of how fortuitous circumstances and responsive bystanders eased the path for my growth.

The Handbook of Social Psychology, the 1968 edition, was my portal from South India into the foreign land of experimental social psychology. I had purchased the five-volume set out of mild curiosity in the content, but mainly because it seemed like a lot of book for the money (they were being offered for a dollar a piece). Until then I had flip-flopped between an arcane psychophysics that was the core of training in general psychology and Marxian sociology, which exhilarated my soul but disappointed my mind. The 1968 Handbook filled the orienting role of teacher. From the writings of McGuire, Aronson, Janis, and Abelson I learned not so much about the content of social psychology (that would be too much to grasp with no technical background) as about a possible way of thinking about the association between mind and society. The handbook told me that somewhere far away from South India, a tribe existed for whom ordinary aspects of social behavior, how people thought and even felt, seemed to be respectable topics of study, and subject to investigation in much the way that physical entities typically were -- through experimental tests. This seemed so remarkable and so right that I had to do it myself, even if it meant leaving an intellectually rich environment to complete forms at the American Embassy in New Delhi to testify that I was not now, nor had ever been a member of a communist party.

I arrived intact in Columbus Ohio, with all five handbooks in tow, to be willingly taught many valuable lessons by Tony Greenwald, Tom Ostrom, Tim Brock, and Gifford Weary. At Ohio State I learned not only the tools of the trade but something far more important: how to decipher what I found to be exciting and aesthetic, to separate it from the rest, and to be able to articulate the difference. My thinking was spent equally in social and cognitive psychology with both thesis and dissertation work on memory – using memory to see what could be understood about the self and about emotion. Even now, as I look back, I see that I had an interest in a certain set of phenomena that showed a dissociation between two systems. The mere exposure effect struck me as it did everybody else as among the most fundamental of principles, but what appealed especially was the dissociation between conscious recollection and unconscious assessment. Likewise with the findings just beginning to be reported in those years with amnesic patients who showed little or no collection but intact implicit memory. A year spent with two free spirits, Claude Steele and Beth Loftus, at University of Washington resuscitated me from the happy but exhausting years at Ohio State and stretched the possibilities from which I could draw my preferences – ones I had not found in the entire midwest. I arrived at Yale as Irv Janis was departing and where my next-door neighbors were Abelson and McGuire – the very individuals whose writings in the handbook had served as my entry into the field. But also, there was Robert Crowder, who with his fastidious taste and candid commentary taught me a great deal more about how to sharpen my preferences and to defend them.

Having dabbled in social science in one form or another since the age of sixteen, one would think that I would have immediately found the problem on which to focus and settled down. But instead, I continued to blithely try out different research problems for their tractability and interest. There was no question that I was having a great deal of fun, but I was far from planning a career that would ensure research security for the future. The benign neglect of junior faculty is often regarded as an unpleasant aspect of life at Yale, but for me it turned into a much needed freedom to select problems and methods without the burden of worrying about the fluctuating opinions of senior colleagues or about tenure, for Yale does not hire assistant professors into a tenure track. The richness of the Yale Department and its explicit disrespect of area boundaries allowed me to wander back and forth without having to say who I was and what exactly I did. The learning in cognitive and social psychology that continued at Yale was juxtaposed with radically different ways of thinking on the part of colleagues in women's studies. The women I met there were among the most brilliant individuals I had encountered and it was in their company that the ideas I was reading about in the Journal of Experimental Psychology on memory and the issues concerning hierarchy and inequality suggested a melding. The idea was to assume a parallel between unconscious and conscious thought and apply it to the study of social cognition – attitudes and beliefs about groups. Much of the work on implicit social cognition has been conducted with Tony Greenwald and this collaboration is among the most satisfying to me, partly because of the comfort of a deeply shared view of social cognition and the research enterprise more generally, and partly because in his presence new learning never ceases.

Doing it my way, as Bill McGuire would urge, has involved taking risks – to continue to work with one's mentor was regarded to be an act of suicide and if that were not enough, to hang out in the women's studies program and even take on administrative duties there was considered to be a sure-fire way to turn oneself into cannon fodder (in case the will needed for suicide was failing). But all these preferences, including taking the time to come to settle on a research question, were possible because I had the luxury of being in great environments. Among the features of great environments like Ohio State and Yale is that they provide the opportunity to develop one's preferences in the company of superb models, and to learn to articulate those preferences to those who do not necessarily share them. Both environments did that for me, and in the discourse that was created, I have surely been the beneficiary.

In 1973, William J. McGuire produced a gem. Entitled *The Yin and Yang of Progress in Social Psychology: Seven Koan,* the paper was based on an address given at the Nineteenth Congress of the International Union of Scientific Psychology in Tokyo and written to stimulate hope in the face of growing pessimism about the state of social psychology and its future. The location of the meeting in the Far East perhaps suggested the use of a form that defied conventional notions of logic -- the koan.[1] Using seven of them to speak of the creative early stages of research, of ways to find pattern in chaos, of observing people (not data), of unexpected opportunities inherent in constraints, McGuire captured the possibilities that arise from permitting more than one way of thinking about all aspects of doing research. Some years later, in complex configuration, these ideas developed into his treatises on contextualism and perspectivism.

In the early 1980s, I was in graduate school at Ohio State when Bill McGuire gave a lecture on his work-in-progress on contextualism, a lecture that made my hair stand on end so exciting were the ideas and so lyrical the delivery. McGuire, the senior Irish Catholic American, speaking a strangely familiar eastern language of paradoxes, of 49 ways to generate hypotheses and turn them on their head, inspired this junior Zoroastrian Indian whose colonial education had dulled all appreciation of the delicacy of oppositional thinking. We talked over Chinese food, that is to say, he talked and I attended to every word and nuance, for it was obvious to me that an encounter with a mind such as his was rare. Some years later, during my job interview at Yale, he recognized me with the greeting "You ate all the Kung-Pao chicken!" His memory for our previous encounter put me immediately at ease. After describing the many ways in which I would be surprised by the poverty of Yale compared to the excellent research

support at Ohio State and University of Washington, he gently prepared me for meetings with his colleagues through sketches of their personalities that were sufficiently accurate as to be unrepeatable. From that day on, and for the past 16 years, Bill McGuire has been my outrageously brilliant colleague, kind and silent benefactor, and trusted confidante on the darkest of days. From *Yin and Yang* I select the 7th koan to pay homage.

The importance of oppositional thinking

McGuire captured the seventh and final koan in *Yin and Yang* not with the words of a Zen Buddhist but rather those of a quantum physicist. From Niels Bohr he drew the idea that: *There are trivial truths and great truths. The opposite of a trivial truth is plainly false. The opposite of a great truth is also true* (McGuire, 1973). McGuire used Bohr's popular statement to acknowledge that the multiple paths he had suggested for recovery from malaise in *Yin and Yang* may themselves be internally contradictory. Over the next 15 years, his profound and practical guides to conducting research, captured by the term *perspectivism* came to contain a complex set of guidelines about doing research (McGuire 1983, 1986, 1989). In brief, perspectivism is an approach to doing science, and it arrives, historically, as the newest in a line of major epistemological orientations. Dogmatism, rationalism, positivism, and logical empiricism each make assumptions about how one can know or understand, and perspectivism challenges its more immediate predecessor of logical empiricism (exemplars being Carnap, Hempel, Feigl, Popper among others) by advocating a system of greater flexibility that acknowledges the complexity of what is to be discovered and the reality of the practice of science.

Most centrally, perspectivism explicitly requires that any *a priori* hypothesis must be accounted for by multiple theories and the scientist should generate a contradictory

and opposing hypothesis that should itself be derived from multiple theories. More radically, perspectivism assumes that because "every proposition is generally wrong so also is its contradictory, and therefore every proposition is occasionally true, at least in certain contexts viewed from certain perspectives … This postulate, that all knowledge formulations are true, is perspectivism's *pons asinorum*, its hardest-to-accept principle. Perspectivism maintains that the task of science, in its *a posteriori* as well as *a priori* aspects is not the dull and easy job of showing that a fixed hypothesis is right or wrong in a given context. Such a modest project is suggested by Popperian inversion of the null hypothesis and his inadequate understanding that the task of current science is to account for covariance rather than, as in antiquity, to establish category membership. Science has the more exciting task of discovering in what senses the hypothesis and its theoretical explanations are true and in what senses false … Perspectivism assigns a higher purpose to the empirical confrontation, that it continue the discovery process, creating new knowledge by revealing, not whether one's fixed *a priori* hypothesis is correct or not, but what that hypothesis means, namely, the pattern of contexts (constituting interacting variables) in which it does and does not obtain, and the mix of reasons for which it obtains in any one context." (McGuire, 1999, p. 407)

The elusiveness of oppositional thinking: A confession

Raised on the milk of contextualism and perspectivism as a pup, I have self-consciously relied on these principles in my teaching and, I had assumed, in the practice of my research as well. The myth that continues and needs to be rectified, I knew, is that there is a fixed *a priori* hypothesis and that experiments are conducted to reveal whether the hypothesis is supported or refuted. Yet, as the following two examples will

demonstrate, my own thinking reveals a dissociation between endorsing the nuances of perspectivism and following its principles in the practice of daily research. My surprise at two unexpected results is the evidence that I had fallen short of constructing the opposite of an expected pattern of results because of strong assumptions about the nature of implicit social cognition. Had I followed the perspectivist exercise, I would have been able to more naturally predict these outcomes as well. I undertake this self-imposed "outing" as a public statement, while memory still serves, of my genuine difficulty in thinking in oppositional terms. Because in each case, I was brought to see the fuller picture by the work of students past and present, the moral of the story also includes a message about the benefit of paying attention to voices of dissent especially from those who are intellectually closest and themselves deeply engaged in the work.[2]

Association or dissociation between explicit-implicit social cognition?

When verbal behaviors of feeling and thinking do not map on to other behavioral indicators of the same feelings and thoughts, a couple of options arise immediately as explanations. First and uninterestingly for our purpose, it is possible that one of the two measures does not represent a faithful rendition of the feeling/attitude or thought/belief, or that the two are mismatched in one of a variety of ways such as their specificity. Those then can account for the lack of correspondence between measures. Alternatively, the lack of correspondence is expected *a priori* because the two measures are expected to tap theoretically distinct constructs. Each measure captures a particular state of affairs and each has, as William James said, its field of application and adaptation, but what it applies to or predicts is not the same. Here, the lack of correspondence between measures not only causes no worry, the experimenter takes delight in specifying it *a*

priori as a test of discriminant validity. This was how I approached the relationship between conscious and unconscious social cognition. I assumed that they were theoretically distinct constructs, and so a lack of correspondence between them was expected, especially when considering certain attitude objects: (a) when the object naturally elicits strong social demand leading the explicit attitude to be in line with what one consciously aspires to feel or think and (b) when the attitude is not elaborated and the opportunities for linkages between conscious and unconscious feeling and thoughts are not present (see Nosek's dissertation, 2002 for a more detailed description of these conditions).

Since the first research using various implicit measures of social cognition began to be conducted in our lab in the late 1980s (judgment tasks, priming, IAT), measures of explicit cognition were usually included and correlations between implicit and explicit measures were routinely assessed. Across dozens of studies using different attitude and belief objects but focusing on social groups, we, like others, discovered strong dissociations between conscious and unconscious attitudes and beliefs. For at least a large subset of attitude objects, neutral to positive attitudes toward socially disadvantaged outgroups were obtained using self-report measures of conscious attitude. An opposing and quite strong negative attitude was obtained on measures that bypassed conscious awareness or control. The same held for measures of beliefs (or stereotypes).

In both cases of attitudes and stereotypes, when the group average for conscious and unconscious measures were placed side-by-side using a common metric, wide divergences were observed as expected. In other words, the theoretical framework within which these data were analyzed supported such a disparity. Just as with other mental

constructs, most obviously memory, it appeared that a useful distinction between conscious and unconscious systems could be offered (see Banaji, 2001). Explicit attitudes presumably reflected feeling states on which the conscious mind could reflect and report in complex response to private and public standards of who one is, who one ought to be, or who one desires to be. These attitudes, I believed, would have their application in circumstances that appropriately elicited them. For example, there should be a high expectation that a relatively positive attitude toward one political candidate over another as expressed consciously ought to predict support for that candidate on other measures such as voting.

To the contrary, implicit attitudes presumably reflect feelings that may be equally influential but manage to escape the glaring look of the conscious eye. They are relatively inaccessible to conscious thought but their existence can nevertheless be tapped by means other than conscious reflection. From their fundamentally different nature to their fundamentally different modes of elicitation, implicit and explicit attitudes are not expected to fall into line with each other. Thus, implicit and explicit attitudes were not only expected to be unrelated, they were, under particular conditions, expected to be opposed to each other. Larry Jacoby, in an elegant series of experiments almost two decades ago, showed just how much the conditions present at learning and testing lead to stark dissociations in memory. If the meaning of a word is attended to, tasks that provoke attention to meaning (e.g., generation) rather than tasks that do not (e.g., reading) produce superior memory on conscious measures such as recognition but not priming. On the other hand, tasks that create traces of the physical features of a word (such as reading rather than generation) produce superior memory on automatic measures such as priming

but not recognition. Just as memory may depend on the type of manipulation performed at the moment of engagement with the material, i.e., at encoding and the type of measure used when that material was required to be drawn out, i.e., retrieval, so also it may be with attitudes. Depending on which aspect of the attitude is in question, the traces of past experience that are detected may vary, sometimes sharply. And this is indeed one truth about the nature of implicit versus explicit attitudes.

But other possibilities about the relationship between these two forms were not under as much speculation as perspectivist thinking would have advised. Instead, the assumption of attitudes as being fundamentally separate or divergent based on their status in consciousness came to be acceptable because it fit with an *a priori* hypothesis about their distinct nature, the different paths of their development, and the unique manner in which they were expressed or elicited. This explained why when the mean valence of attitudes toward disadvantaged outgroups pointed in opposite directions (neutral or positive on explicit measures, negative on implicit measures); there was no rush to seek out conditions in which that would not be the case. This view was further strengthened by the use of low powered, small *n* designs. In part because small *n* designs were sufficient to see differences between the two conditions of the implicit measure (faster pairings of groupA+good/groupB+bad than vice versa) and because the analogous computation on the explicit measure did not reveal the same effect, another measure of association, the correlation between the two, took second place. The data at the level of group means may show divergence, and yet there could easily be a relationship at the individual level – that is, a person who scores relatively more negatively on the implicit measure may also score relatively more negatively (even if the overall group mean is

neutral) on the explicit measure. Correlations were routinely conducted to test the association between the conscious and unconscious measure and they were often small and insignificant or small and significant – nothing striking that led me to change my mind about the dissociation or separateness of the attitude as being the only story.

Two pieces of evidence led to me to a somewhat different place in understanding this relationship. First, data from a website had the advantage of large *n*'s and standardized tasks. For each task, a simple explicit question was asked about the relative liking for two groups that could be correlated with the implicit measure. The analyses across tasks showed two clear findings: there was a good deal of variability in the correlation between explicit and implicit attitude across tasks but there were sizable correlations between the two types of measures on a large subset of tasks (see Table 1 from Nosek & Banaji, in press). In fact, new analyses had produced a sizable increase in the already high explicit-implicit correlation for the Bush-Gore attitude to elevate it to almost .80. This discovery led to Brian Nosek's dissertation research in which he is exploring two factors that may affect the strength of implicit-explicit attitude correlations: the degree of social demand created by the nature of the attitude object (lower relationships for higher demand attitude objects) and degree of elaboration (lower relationships for less well-elaborated attitudes). Thus, relations between implicit and explicit attitudes should be relatively low for racial attitudes (high demand) and for insect-flower (low elaboration) but higher for attitudes toward math/science (low demand, high elaboration) and political candidates (Bush-Gore; low demand, high elaboration preceding the election). A preliminary analysis of the tasks in Table 1 are reported in Nosek and Banaji (in press) and give some suggestion that both of these

factors are likely to be operating, although for the present purposes, it is the data in the first column only that are relevant. First and foremost then, relationships between implicit and explicit attitudes were expected to be around zero, and standard laboratory research with small *n*'s confirmed that result. The occasional large correlation could be written off as a Type 1 error. It was a series of web-based data collections with large *n*'s across many different tasks that gave the pattern credibility and moved me in the direction of taking such correlations seriously – that is, to first accept that they were meaningful reflections of the relationship between the two systems. Had the World Wide Web been around in the early 1970's, the *Yin and Yang* paper would surely have had something to say about its use in the excavation of social psychological goodies.

Coming to the question with a different orientation, Wil Cunningham has examined this issue in a different way (see Cunningham, Nezlek, & Banaji, 2002). If social situations provoke similar experiences over time, individual differences in orientation toward social groups should develop that are stable and that can be captured through the appropriate analyses. He has favored a style of research that uses (a) multiple measures of an attitude object, (b) multiple attitude objects to develop a strong measure of the underlying factor, (c) large *n*'s and (d) covariance structural modeling analysis as the statistical technique. The logic of Cunningham's experiments demand that a single individual provide data on more than one occasion, using more than one measure in each of the implicit and explicit categories, and about multiple attitude objects.

Using a cluster of five social groups (American-Foreign, Black-White, Gay-Straight, Jewish-Christian, Rich-Poor) he has replicated previous work on ethnocentrism to show that there is indeed an underlying latent factor of ethnocentrism when viewing

explicit attitudes. In other words, those whose attitudes are negative toward one of these groups also tend to have negative attitudes towards others of these groups. Two new findings also emerge from this research. First, an underlying latent factor, labelled "implicit ethnocentrism", in parallel, also emerges. In other words, speed to respond to combined pairings of group+positive and group+negative shows a consistency in favoring all advantaged groups over relatively disadvantaged groups. For the issue under consideration, there is another more important finding. The latent factor capturing explicit ethnocentrism is highly correlated (approx. .45) with the latent factor that captures implicit ethnocentrism. This relationship is weak when examining the single group correlations between implicit-explicit measures, but it is robust when measurement error is removed and the measure of association compares data that make up the two latent factors of implicit and explicit ethnocentrism.

Remarkably, the factor structure is such that a single factor solution (one that assumes no uniqueness between implicit and explicit constructs) does not fit the data. A two-factor solution that maintains the distinction between the two constructs provides the best fit. In other words, these are indeed unique constructs, but the opposite of this proposition is also true – they are indeed related constructs. In this case, it is not a matter of finding separate experimental conditions under which the two alternatives are each true. Rather, the same data set, when appropriate focus is brought to bear on it, reveals evidence for both.

From these data we draw a conclusion that does not have the benefit of a simple solution (i.e., that the two systems of conscious and unconscious attitudes are unrelated). Rather, these data are beginning to more faithfully reflect what may be the actual state of

affairs regarding their relationship, at least as it pertains to analyses of the social groups under study. First, the two constructs can be seen as unrelated (as seen in the sharp divergence in the valence of the group means), and this most successfully tells the story of culture – that deeply held values of egalitarianism, fairness, and moral responsibility lead to genuine shifts in explicit attitude toward social groups that are known to be historically disadvantaged. Yet the human learning system also carries information based on "cold" knowledge presented about the state of the world as well as personal experience that reflects the cultural pairing of positive and negative attributes with each social group. These need not be consistent and, in fact, one measure of the evolution of a society may indeed be the degree of separation between conscious and unconscious attitudes – that is, the degree to which primitive implicit evaluations that disfavor certain social groups are compensated for or corrected by explicit statements of consciously held attitudes of tolerance.

The picture obtained in the early days of our research that promoted me to stress the divergence between explicit and implicit attitudes is indeed true. But this view depends on the type of data. If the focus is brought to bear on individual differences, it appears that the conclusion regarding divergence cannot be said to operate at the individual level. In other words, it would no longer be correct to assume that a given individual's standing on explicit attitude measures is always unrelated to that individual's standing on implicit attitude measures. Yes, there are the usual constraints of type of attitude object (elaborated, high in demand, etc.) but the fact is that a deep association between the two systems is possible – the Bush-Gore IAT can show up to a .80 correlation with the explicit measure. This association is one that was ignored and it was

brought to our attention by Nosek's analyses of web data and especially by Cunningham's experiments that sought a personality interpretation by focusing on ethnocentrism.

Attitudes are not things but construals of the moment

As early as the first grant proposal we wrote, the feedback nudged us to think about the malleability of implicit attitudes. Jim Sherman and his colleagues first reported that an internal state produced changes in implicit attitude. Heavy smokers showed negative associations to smoking, as do nonsmokers, but not when they had abstained from smoking. This early report led me to test myself by varying states of water deprivation – no liquid consumption for 24-36 hours. Aiden Gregg created the tests of evaluation for water versus food, and I showed, as the Sherman data indicated, greater implicit preference for water over food in this state of water deprivation. But we were not able to replicate this finding with a larger group even though Gregg heroically tried various states of food/water deprivation and various locations such as gymnasiums. In addition, there was again, as in the previous case, the baggage of theory.

Explicit attitudes are part of a system that is susceptible to conscious control and hence capable of changing on demand. Implicit attitudes, by contrast, are disengaged from conscious thought and are unlikely to shift in response to the willful call for change. This assumption about the difference between the nature of the two constructs still holds, almost by definition, because tasks are constructed to vary in exactly this way. But a related assumption came to be associated with this belief that incorrectly led me to view implicit attitudes as inflexible. In particular, I was unprepared for data that showed the influence that minor variations in social situations can play in defining the attitude object

itself – the different construals possible of seemingly the same attitude object. Prior to the studies I now describe, it would be accurate to say that I would not only have failed to predict their outcomes, I would have advised against putting in the effort to test such an effect. Fortunately, I was not consulted.

The most helpful presentation here may be to summarize the findings from three laboratories which issued full reports in the November 2001 issue of the *Journal of Personality and Social Psychology*. Buju Dasgupta conducted studies in which participants performed a matching task of descriptions to pictures. The pictures and descriptions to be matched contained either admired Black individuals (Martin Luther King Jr.) and unadmired White individuals (Timothy McVeigh) or vice versa. After completing a control task or one of these in the experimental conditions, participants were given a standard race IAT using faces of unknown individuals representing the two groups. The task has been widely used and is known to reliably and robustly produce positive evaluation of White Americans relative to Black Americans (see Nosek, Banaji, & Greenwald, 2002). In the admired Black condition, subjects showed a significantly weaker race bias than in the control and opposite prime conditions (see Dasgupta and Greenwald, 2001).

Curtis Hardin and his colleagues reached a similar conclusion by varying the race of the experimenter. They showed, using paper-and-pencil measures of automatic race attitude that the mere presence of a Black experimenter reduced the negative evaluation of that category (Lowery, Hardin, Sinclair, 2001). Irene Blair used yet another manipulation, that of mental imagery, to show a similar effect on gender. Imagining what a strong woman is like (what is she capable of doing? what are her likes and

activities?) reduced the otherwise robust implicit stereotype of female+weak and male+strong.

Dividing up this universe into a 2 X 2 that represents whether the source of influence was within awareness and controllable or not, and whether the behavior that was measured was controllable or not, the focus so far even in the most impressive of studies had been on the degree of awareness or controllability of the prime or influencing agent. These studies showed for the first time that a relatively trivial experience such as the presence of a person or a mental exercise could change evaluation and other associations that were not responsive to conscious control. This was intriguing, because by their very nature, measures of implicit evaluation were expected to be tapping something that had accumulated over a long period of time and therefore presumably insensitive to small scale experiences in the recent past.

Knowledge of these findings, long before they appeared in print, kept me in a disbelieving but prepared mode to understand an outcome obtained in my own collaboration with Brian Nosek. For the past five years we have conducted numerous studies to examine implicit evaluation of academic orientations, in particular attitudes toward math and science as a function of implicit group identity (attachment of self with male or female) and the stereotype of a natural association between male (rather than female) and math/science. As Nosek, Banaji, and Greenwald (2002b) demonstrate, there is a strong gender difference in automatic attitudes toward math and science, with women showing more negative attitudes than men. This effect, obtained many times over, was not one that we were prepared to see disappear, but it did in a recent experiment, and that proved baffling. A finer-grained analysis of the data led to an observation that continued

to be replicated: the standard pattern of a gender difference was obtained but only when a male experimenter conducted the study. An opposite effect, with females showing positive implicit attitudes toward math was obtained when a female experimenter performed the study. Since this serendipitous discovery, we have found that other laboratories had predicted and detected similar effects on math attitudes and math performance.

Experiments like these challenge the naïve view, one that I may have harbored myself, that measures of implicit social cognition would not be sensitive to such interventions. This view had a reasonable basis in the position that implicit social cognition reflects routinized expressions of a slow learning system, one whose function is to reflect the output of long-term experience. Given this assumption, it was expected that putting aside uninteresting variations that reflect measurement error, measured implicit attitudes ought to be relatively impervious to situational demand. After all, these evaluations had developed over long periods of time and were resistant to simple attempts at faking. Why then should something as mundane as consideration of positive African American exemplars (and negative Euro American ones) produce weaker negative implicit attitudes toward African Americans as a group? Why should such a brief event lead to an automatic evaluation that is more positive than the typical one (obtained in the control condition)? The mistake may have been to assume that a representation that is not amenable to the dictates of conscious will is impervious to other inputs as well. The implication of this may be wider than immediately recognized. The first error we tend to make is to assume that more of our thoughts and feelings are within conscious control than may be the case. Having demonstrated the presence of automatic

and unintended thought, we come to believe that if relatively uncontrollable, unintended thought cannot be controlled through the mechanism assumed to be the Supreme Controller, conscious will, that it is unlikely to be controlled in other ways. But as the work of Bargh and colleagues has demonstrated so significantly (see Bargh & Chartrand, 19xx for a review), influences that come from outside of oneself can produce direct effects on behavior. The surprising finding (one that I had not anticipated) common to studies by Dasgupta and Greewald, Lowery et al, Blair et al., and Nosek et al. was that mild intervention could influence a behavior that is itself assumed to be automatic – implicit preferences for or against a given social group.

When are such effects likely to occur? It seems worth suggesting that effects of malleability may be greater when the evaluative boundaries of an attitude object are not restrictive. A social group or a person (or almost any social object for that matter) has a potentially infinite number of ways in which it can be brought to mind. Groups that are typically viewed negatively can be represented in positive ways because of the evaluative elasticity inherent in attitude objects, and the opposite is true as well. It is the case that implicit attitudes toward African Americans compared with European Americans are more negative, but that is because the typical circumstance under which the attitude is elicited pulls for the default or dominant evaluation of the group. But social groups do not have a single evaluation attached to them – multiple features are present and each of them has the potential to create a temporary representation, for a single evaluative episode, that can vary. To ask what one's attitude toward Chinese Americans is will produce quite differing outcomes depending on the task but also the particular features that represent the group – names, faces, maps, personality traits, food, cultural practices.

The experiments that I described in this section reveal that there are a plethora of ways to define a single attitude object that is a social group and that the mental shaping of the evaluative and stereotypical aspects of the group can shape the attitude that follows. To ask the "real" attitude to "please stand up" would be to assume that there is both a real attitude and that there is one attitude. Neither of these assumptions is supported by the current evidence, even though they may ring true and be supported in theory.

Because these recent studies are the first to demonstrate that automatic attitudes are themselves sensitive to contextual input, they are not designed to take a position on the question of what it is they are revealing. When an implicit attitude or stereotype appears to shift (e.g, a typically negative attitude toward Black Americans is reduced in negativity or a typically high association between gender and strength is attenuated) it intuitively appears as if that particular attitude or stereotype has "changed". My own assessment at this time (in an essay on the ways in which I have been wrong it would be ironic to generate yet new ideas that will need retraction!), is that there is no reason to assume that attitude "change" has occurred in the traditional sense. Rather it seems most parsimonious to conclude that the situations present in the experiments described here reveal the importance of situational construal. Eliot Smith (19xx) most notably has suggested that assuming a connectionist stance in thinking about mental representation can lead to a quite different manner of thinking about social cognition. Instead of thinking of an attitude as a thing that sits on a mental shelf, attitudes can be thought of as patterns of activation that reveal the presence of repeated learning and their reconstruction in a particular environment. If indeed we can muster the imagination to think this way about attitudes, these attitudes do not suggest that change has taken place,

because change in that traditional sense is not a plausible construct. Rather -- as Asch pointed out in his simple demonstrations of how we come to think positively or negatively of politicians (by construing of them in response to two different types of exemplars) -- the very representation of an attitude is shaped powerfully by the props that exist in a given elicitation condition. The experiments by Blair, Dasgupta, Hardin, Nosek, and Mitchell suggest a renewed and even more rigorous emphasis on the acute properties of social situations. They ought to also suggest caution in succumbing to the intuitive appeal of "attitude change" in the sense in which we mean it colloquially, that is to say when we refer to conscious attitude change.

Mitchell, Nosek, and Banaji (2002) have shown that attitudes toward exactly the same exemplars can be dramatically different as a function of the frame for processing or the construal of the object. The previous studies have shown some change in valence – such as the lowering of the otherwise positive bias toward White Americans, or the lowering of the otherwise stereotypic association between male and strength. Mitchell et al. sought to create circumstances in which a more dramatic shift in automatic evaluation can be detected. They focus on retaining the same exemplars to denote race groups, but shift the manner in which those exemplars are viewed. For example, Black and White males and females can be viewed either through the lens of race or gender. The particular lens that is used can produce noticeable shifts (i.e., from negative to positive) in the evaluation of the attitude object. When viewed through the race lens, Black females receive negative evaluation (as do Black males). But when viewed through the gender lens, Black females are viewed positively because females as a group are viewed more positively (especially by females).

In other words, given the many dimensions inherent in any attitude object, the particular one that is drawn out by the forces of situation or personality can determine the outcome. To speak of Black Americans as a group as eliciting a single attitude is not only simplistic but leads to an incorrect assessment that such an attitude requires deep-rooted change. Getting away from the notion of change will get away from questions regarding how long the type of "change" seen in these studies will last. Instead, it will focus on a far more important consequence of this discovery -- the mode by which corrections for existing bias can be initiated.

The experiments reviewed here suggest a different mode for non-dominant attitudes to be created: repeated, pervasive, even (especially) minor interventions. They support those who have claimed that the contents of the environments in which we think and do our work matter – the gender of the teacher teaching calculus, the accent of the judge probing for an answer, the ethnicity of the janitor cleaning the office, the nationality of the scientist testifying before congress. These contexts tell us who can be what, and it appears that our minds are exquisitely sensitive to detecting and using such information. The actual presence of human beings we observe in the roles that make up social life matter; so do all representations that contribute to experience, even two dimensional representations: what sits on our screen-savers, the pictures that hang in our hallways, the advertisements that fill almost every social vacuum, these have the potential to influence not just our consciously framed ideas and feelings but our automatic and unintentional ones as well. When particular evaluations come to be repeatedly paired with an object in a given culture they create the appearance that a single attitude exists – by being so easily and "naturally" evoked. The fact that the experiments described here

demonstrate the ability of even mild interventions to produce shifts on measures of automatic attitudes suggests, optimistically, the potential to effect deviations from ingrained cultural learning of attitudes and stereotypes.

Conclusion

This chapter has largely been concerned with two issues that began with discoveries that later revealed their opposite as well. If implicit and explicit attitudes were disjointed, it seemed to us at first that they could not also be strongly associated. Likewise, if implicit attitudes were automatic and relatively uncontrollable, it seemed that environmental probes could not shape or shift them. Yet it now appears that each of these original discoveries and their opposites are both true. Conscious and unconscious forms of evaluation are both related and independent. Implicit attitudes may not bend to the instruction of conscious will, but they are malleable in response to subtle features of the environment that can shape their expression, sometimes radically. Theories of implicit social cognition need to take these oppositional truths into account.

Oppositional thinking in science is not a dominant response. It is an acquired taste and as such it needs cultivation, that is to say, attention from deliberative thought and repeated practice. Even in the well-intentioned, it can easily fall to the side, as repeated tests of single hypotheses, from their simplicity, ease, and please-all quality, occupy all the rings of our various scientific circuses. Yet, as McGuire proposed almost 30 years ago, appreciating the virtues of paradoxical propositions, even fully contradictory ones, must become part of the daily activities of science. The difficulty of practicing oppositional thinking in spite of approving of its benefits at a more conscious intellectual

level, of reaching for the skill but never quite mastering it, is perhaps itself a process that would fit into this evolving, never quite complete, view of science.

I have paid homage to the seventh koan but not to the master himself. To do so, I move a few thousand miles west from the Far East, to a different culture. I pick up a piece of the Mahabharata, the Indian epic of unparalleled length and complexity revolving around the battle of Kurukshetra between the saintly Pandavas and their evil cousins, the Kauravas. In one scene, each side has sent its most prominent leader to their cousin Krishna, an avatar of god Vishnu, to persuade him to join their side in the war. Arjuna, the Pandava hero, is asked to choose first: the Pandavas can have Krishna's entire army "large and almost invincible" or Krishna himself who "shall wield no weapon and take no part in actual fighting" (Rajagopalachari, 19xx). Without hesitation Arjuna picks the unarmed Krishna, leaving the evil Duryodhana thrilled with the prospect of a full army. Those who know which side won, know a whole lot more. Krishna and Arjuna's dialogues address the primary duty of a human being, the distinction between self and group, questions of when war is moral, and the demands and limits of personal loyalty. Krishna's remarks in response to Arjuna's questions have come down to us as the Bhagavad Gita, Hinduism's enduring commentary on the human condition.

"I am a Vishnu man myself", Bill once said to me. I was surprised that an American colleague would know enough about the religion to have such an opinion and one based on apparent knowledge because it appeared in the conversationally perfect moment of being told of my relation, by marriage, to Shaivite Brahmins. So I use the story of Vishnu (Krishna) while confessing that I am a McGuire person myself. If I had a choice like Arjuna's, I would without hesitation pick Bill bearing no weapons in favor of

the most invincible army in psychology. Bill and his teachings -- in *Yin and Yang*, but also in other works addressing contextualism and perspectivism, and his reviews of attitude theory and research, not to mention his contributions to understanding thought systems -- take stances comparable to Vishnu's support to Arjuna: words of a particular sort. Not answers but more questions, not instructions but revelations, not dismay but hope, not about his way to do it but about the discovery of yours. As I sense it, all this from a quiet confidence stemming from his belief that the task is to strive, to seek, to find, and not to yield.[3]

References

Banaji, M. R. (2001) "Implicit Attitudes Can Be Measured." In Roediger, H. L. , Nairne, I.N., Suprenant, A. M. (Eds.). *The Nature of Remembering: Essays in honor of Robert G. Crowder.* (pp. 117 – 149). Washington, D.C.: American Psychological Association.

Bargh, J. A. and Chartrand, T. L. (2000). "Automatic assumptions of automaticity." *American Psychologist*, 55, 766 – 767.

Blair et al. (2001). "Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery." *Journal of Personality and Social Psychology*, 81, 828 – 841.

Cleary, Thomas. (1994). *Instant Zen - Waking Up in the Present*. Berkeley, Calif.: North Atlantic Books.

Cunningham, Nezlik and Banaji, M. R. (in progress). Conscious and Unconscious Ethnocentrism: Revising the Ideologies of Prejudice. *Journal of Personality and Social Psychology.*

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

Jacoby, L. L. (1983) "Remembering the data: Analyzing interactive processes in reading". *Journal of Verbal Learning and Verbal Behavior,* 22, 485-508.

Kapleau, R.P. (1989). *The Three Pillars of Zen.* New York: Doubleday.

Lowery, B. S. and Sinclair, S. (2001). "Social influence effects on automatic racial prejudice." *Journal of Personality and Social Psychology.* 81, 842 – 855.

McGuire, W.J. (1973) "The yin and yang of progress in social psychology: Seven koan". *Journal of Personality and Social Psychology,* 26, 446-456.

McGuire, W.J. (1983) "A contextualist theory of knowledge: Its implications for innovation and reform in psychological research". In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 16, pp. 1-47). New York : Academic Press.

McGuire, W.J. (1986) "A perspectivist looks at contextualism and the future of behavioral science". In R. Rosnow and M. Georgoudi (Eds.) *Contextualism and understanding in behavioral science: Implications for research and theory* (pp. 271-301). New York: Praeger.

McGuire, W.J. (1989) "A perspectivist approach to the strategic planning of programmatic scientific research." In B. Gholson, W.R. Shadish, Jr., R.A. Neimeyer, &

A.C. Houts (Eds.), *The psychology of science: Contributions to metascience* (pp. 214-245). New York: Cambridge University Press.

McGuire, W.J. (1999) *Constructing social psychology.* New York: Cambridge University Press.

Mitchell, J. P., Nosek, B. A. and Banaji, M. R. (in progress) "Multitudes of Attitudes: Automatic Evaluation of Multiply-Categorizable Targets."

Nosek & Banaji Polish Handbook.

Nosek, B. A., Banaji, M. R., and Greenwald, A. G. (2002a) "Harvesting Implicit Group Attitudes and Beliefs From a Demonstration Web Site." *Group Dynamics*: *Theory, Research, and Practice,* 6, 101 – 115.

Nosek, B.A., Banaji, M. R. and Greenwald, A. G. (2000b). "Math = Male, Me = Female, therefore Math ≠ Me." *Journal of Personality and Social Psychology*.

Smith, E. R., & Zarate, M. A. (1992). "Exemplar-based model of social judgment." *Psychological Review*, 99, 3 – 21.

Smith, E. R., Fazio, R. H., and Cejka, M. A. (1996). "Accessible attitudes influence categorization of multiply categorizable objects." *Journal of Personality and Social Psychology*, 71, 888- 898.

Sherman, S.J., Presson, C.C., Chassin, L., Rose, J.S., & Koch, K. (in press). "Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation". *Journal of Social and Clinical Psychology*.

Rajagopalachari, C. (1951) *Mahabharata.* Bombay: Bharatiya Vidya Bhavan.

Endnotes

[1] I rely entirely on experts to define and elucidate the concept of a koan. Koans are Zen challenges or riddles that defy conventional use of logic because they "cannot be grasped by a bifurcating intellect" (Kapleau, 1989). They are pithy sayings that contain "patterns, like blueprints, for various inner exercises in attention, mental posture, and higher perception, summarized in extremely brief vignettes enabling the individual to hold entire universes of thought in mind all at once, without running through doctrinal discourses or disrupting ordinary consciousness of everyday affairs." (Cleary, 1994). "To people who cherish the letter above the spirit, koans appear bewildering, for in their phrasing koans deliberately throw sand into the eyes of the intellect to force us to open our Mind's eye and see the world and everything in it undistorted by our concepts and judgments." (Kapleau, 1989).

[2] All students, not just fortunate ones, must feel safe and even elated when a result shows the opposite of what an advisor has predicted or is antagonistic to a theoretical position with which the laboratory is engaged. In order to do this, environments must be created by all advisors, not just excellent ones, in which oppositional thinking is advocated and explicitly rewarded.

[3] At Bill's Festschrift, his daughter Ann McGuire gave a hilarious talk about her father using family photos and ended her homage with a beautiful reading of Tennyson's Ulysses. I use her (his) lines to salute Bill and to remember that joyous day.

Endnotes

[1] I rely on experts to define and elucidate the concept of a koan. Koan, derived from the Chinese *kung-an*, meaning a public statement or proclamation. Koans are Zen challenges or riddles that defy conventional use of logic because they "cannot be grasped by a bifurcating intellect" (Kapleau, 1989). They are pithy sayings that contain "patterns, like blueprints, for various inner exercises in attention, mental posture, and higher perception, summarized in extremely brief vignettes enabling the individual to hold entire universes of thought in mind all at once, without running through doctrinal discourses or disrupting ordinary consciousness of everyday affairs." (Cleary, 1994). "To people who cherish the letter above the spirit, koans appear bewildering, for in their phrasing koans deliberately

throw sand into the eyes of the intellect to force us to open our Mind's eye and see the world and everything in it undistorted by our concepts and judgments." (Kapleau, 1989).

[2] All students, not just fortunate ones, must feel safe and even elated when a result shows the opposite of what an advisor has predicted or is antagonistic to a theoretical position with which the laboratory is engaged. In order to do this, environments must be created by all advisors, not just excellent ones, in which oppositional thinking is advocated and explicitly rewarded.

[3] At Bill's Festschrift, his daughter Ann McGuire gave a hilarious talk about her father using family photos and ended her homage with a beautiful reading of Tennyson's Ulysses. I use her (his) lines to salute Bill and to remember that joyous day.

Table 1. Summary of average implicit-explicit correlations, perceived elaboration, and perceived self-presentation demands for 15 attitude object pairs (from Nosek and Banaji, in press).

	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	1.3
Fruit / Bugs	.18	1.3	4.8
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