Banaji, M. R. (2003). The opposite of a great truth is also true. In J. Jost, D. Prentice, & M. R. Banaji (Eds.) The yin and yang of progress in social psychology: Perspectivism at work. Washington, DC: American Psychological Association.

10

The Opposite of a Great Truth Is Also True: Homage to Koan #7

Mahzarin R. Banaji

Prologue

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 at a dollar apiece). Until then, I had flip-flopped between an arcane psychophysics, which was the core of training in general psychology, and Marxian sociology, which exhilarated the soul but disappointed the mind. The 1968 *Handbook* filled the orienting role of teacher. From the writings of McGuire, Zajonc, Janis, and Abelson, I learned not as much about the content of social psychology (that was too much to grasp with no background) as I did about a particular way of thinking about the relationship between mind and society. The handbook told me that somewhere far from South India, a tribe existed for whom ordinary aspects of social behavior, how people thought and felt, seemed to be respectable topics of study and investigation in much the way that physical entities were in other sciences—through experimental analysis. This seemed so remarkable and so right that I had to do it myself, even if it meant leaving an intellectually rich environment to testify at the American embassy in New Delhi that I was not now, nor had ever been, a member of a communist party.

The issues raised in this paper emerged from the work of students past and present: Irene Blair, Nilanjana Dasgupta, Curtis Hardin, Jason Mitchell, and Brian Nosek, whose view of the power of the immediate situation and the context in which thoughts and feelings are elicited challenge theory about the nature of implicit social cognition, and William Cunningham, whose belief that attitude unification in personality is an equally important element produced new tests and interpretations. 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 by 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. I am also grateful to Brian Nosek and Shari Stout for comments on an earlier draft.

At Ohio State, I acquired tools from scientists who took their trade seriously and who had a point of view about everything. Here, in the company of remarkable graduate students and faculty, I learned how to decipher what I found to be exciting and challenging, to separate it from the rest, and to be able to articulate the difference. At Yale, where I took my first job, the environment was one of tolerance for all points of view (it was, after all, the place where it had been possible to join learning theory and psychoanalysis). Such tolerance, accompanied by a benign neglect of junior faculty, often regarded as an unpleasant aspect of life at Yale, 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. The learning in social (with McGuire, Abelson, and Salovey) and cognitive psychology (with Bob Crowder) that continued at Yale was juxtaposed with different ways of thinking on the part of colleagues in history and women's studies. It was in their joint company that the ideas about separate memory systems and of a social world marked by hierarchy and inequality suggested a melding. The idea was to assume a parallel between unconscious memory and unconscious attitudes and beliefs.

Doing it my way, as Bill McGuire would urge, has involved taking some risks—to continue to work with one's mentor was regarded as an act of suicide. If that were not enough, to regularly visit the women's studies program and even take on administrative duties there was considered a surefire way to turn oneself into cannon fodder in psychology. But learning to have a perspective and articulating it, being prepared to meld dramatically different viewpoints, and being slow to choose a problem to settle on were all possible because I had the privilege of being in great environments. Among the features of great environments such as Ohio State and Yale is that they provide the opportunity to develop one's preferences in the company of superb models, as well as the opportunity to learn to articulate those preferences to those who do not necessarily share them. Both environments did that for me, and in the discourse they facilitated, I was surely 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

[1]Koan, derived from the Chinese *kung-an*, is defined as 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)

speak of the creative early stages of research, of ways to find pattern in chaos, of observing people (not data), and 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, because 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!" and 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, silent and kind benefactor, and trusted confidante on the darkest of days. From "Yin and Yang," I select the seventh 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 with 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 in "Yin and Yang" for recovery from malaise 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 include assumptions about how one can know or understand, and the perspectivist approach poses a challenge to logical empiricism (exemplars being Carnap, Hempel, Feigl, and Popper), 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 that the scientist 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

Because I was raised on the ideas of contextualism and perspectivism, 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 demonstrate, my 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 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 explicitly followed the perspectivist exercise, I would, more naturally, have predicted these outcomes as well. I undertake the self-imposed "outing" of this lapse publicly, while memory still serves. In each case, I was brought to see the fuller picture by the work of students past and present, and hence 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 onto other behavioral indicators of the same feelings and thoughts, the following options immediately arise as explanations. First and not interesting, it is possible that one

[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 in which oppositional thinking is advocated and explicitly rewarded must be created by all advisors, not just excellent ones.

of the two measures does not represent a faithful rendition of the construct or that the two are mismatched in one of a variety of ways, such as their specificity. 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 does not cause worry; rather the experimenter actually takes delight in specifying it a priori as a test of discriminant validity. This is 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.

When the research using various implicit measures of social cognition began in my lab in the late 1980s (judgment tasks, priming, Implicit Association Test [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, we 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 of social groups.

In both cases of attitudes and stereotypes, when the group averages for conscious and unconscious measures were placed side-by-side using a common metric, wide divergences were observed, as was 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 components could be offered (see Banaji, 2001). Explicit attitudes presumably reflected feeling states on which the conscious mind could reflect and report, in a 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 explicit attitude toward one political candidate over another ought to predict support for that candidate on other measures such as voting.

On the contrary, implicit attitudes presumably reflect feelings that may be equally influential but manage to escape the glare of the conscious eye. These feelings are relatively inaccessible to conscious thought, but their existence can nevertheless be tapped by means other than introspection and deliberate reflection. From their different evolution to their different modes of elicitation, implicit and explicit attitudes are not expected to fall in 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 in the 1980s, showed just how much the conditions present at learning and testing led to stark dissociations in memory. If the meaning of a word is attended to, tasks that engage semantic meaning, such as generation tasks, rather than tasks that do not (e.g., a reading task), produce superior memory on conscious measures such as recognition but not on priming. On the other hand, tasks that create traces of the physical features of a word (such as reading but not generation) produce superior memory on measures such as priming but not on recognition (Jacoby, 1983). Just as memory may depend on the type of manipulation performed at the moment of engagement with the material (at encoding) and the type of measure used to draw out the material (at 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 enough to be evaluatively opposed. This is indeed one side of the truth about the relationship between implicit and explicit attitudes.

Other possibilities about the relationship between these two types of attitudes were simply not under consideration 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 show differences between the two conditions of the implicit measure (faster pairings of Group A+good/Group B+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, an individual who scores relatively more negatively on the implicit measure may also score relatively more negatively 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 enough to change my mind about the dissociation or separateness of the two types of attitude being the only story.

Two pieces of evidence led me to a somewhat different place in understanding this relationship. First, data from a Web site 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 (Nosek & Banaji, 2002). In fact, new analyses had produced a sizable enough increase in the already high

explicit–implicit correlation for the Bush–Gore attitude to elevate it to almost .80. This discovery and others like it led to Brian Nosek's dissertation research in which he explored 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 the 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 or science (low demand, high elaboration) and political candidates (Bush–Gore; low demand, high elaboration preceding the election). A preliminary analysis gave support to the idea that both these factors may be operating.

First and foremost, relationships between implicit and explicit attitudes were expected to be about zero, and standard laboratory research with small *n*s had confirmed that result. The occasional large correlation could be written off as a Type I error. It was a series of Web-based data collections with large *n*s across many different tasks that gave the pattern credibility and demanded to be taken seriously—that is, to first accept that they were meaningful reflections of a positive relationship between implicit and explicit attitudes. In no case was the relationship negative or even zero: It was always positive, ranging from small to large.

Coming to the question with a different orientation, Wil Cunningham examined this issue as well (see Cunningham, Nezlek, & Banaji, 2003; Cunningham, Preacher, & Banaji, 2001). If social situations provoke similar experiences over time, individual differences in orientation toward social groups that are relatively stable modes of responding should develop. Cunningham favored a style of research that used (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 required 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) Cunningham showed, as did the early researchers interested in prejudiced personality, that there is indeed an underlying latent factor of ethnocentrism when examining explicit attitudes toward multiple social groups. In other words, those whose attitudes are negative toward the outgroup (also more socially disadvantaged), also tend to have negative attitudes toward other such groups.

Two new findings emerged from Cunningham's work. First, an underlying latent factor, labeled "implicit ethnocentrism," in parallel, 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 (~ .45) with the latent factor that captures implicit ethnocentrism. This relationship is weak on examination of the single

group correlations between implicit–explicit measures, but it is robust when measurement error is removed and the measure of association is computed on the 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 provides the best fit. In other words, implicit and explicit ethnocentrism are indeed unique constructs, but the opposite of this proposition is also true: They are also strongly related. 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 reveals, depending on the type of analysis, evidence for two seemingly contradictory findings that are not actually so.

From these data, no simplistic conclusion can be reached—that the two families of conscious and unconscious attitudes are fully independent or that they are identical constructs. 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 of implicit and explicit attitude can be seen as unrelated (if the sharp divergence in the valence of the group means is the focus), and this most successfully tells the story of practiced feelings: that practiced values of egalitarianism, fairness in judging individuals and groups, lead to genuine shifts in explicit attitudes. Yet a learning system like that of humans also carries information about the world (accurate and inaccurate) that is acquired within a culture and mediated through personal experience. The end product of these processes 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 or outgroups are explicitly corrected at the conscious level at which control is possible.

The correlational data indicate 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 (e.g, the .80 Bush–Gore IAT-explicit correlation). Both answers are true: There is a discrepancy between conscious and unconscious attitudes toward social groups (with the unconscious family of attitudes being more negative). There is wide agreement at the individual level such that there is positive covariation between implicit and explicit attitudes.

Attitudes Are Not Things But Construals of the Moment

As early as the first explorations on this topic, my collaborators and I were nudged to think about the malleability of implicit attitudes. Jim Sherman and his colleagues first reported that an internal state produced changes in implicit attitude (Sherman, Presson, Chassin, Rose, & Koch, 2003). Heavy smokers showed negative associations to smoking, as did 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. We were not able to replicate this finding with a larger group even though Gregg heroically tried various states of food and water deprivation and various locations such as gymnasiums in which the test was conducted (Gregg & Banaji, 1999).

Thus the view that implicit attitudes were not susceptible to intervention grew, and theory supported that view. Explicit attitudes are part of a system that is susceptible to conscious control and hence is capable of changing on demand. Implicit attitudes, in 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. A related assumption that incorrectly led me to view implicit attitudes as invariant across social situations came to be associated with this belief. In particular, I was unprepared for data that showed that the influence that minor variations in social situations, such as the presence or absence of a person, 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 one of these or a control task 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, 2002a). In the admired Black condition, subjects showed a significantly weaker race bias than in the control and opposite prime conditions (see Dasgupta & Greenwald, 2001).

Curtis Hardin and colleagues reached a similar conclusion by a more natural manipulation that varied the race of the experimenter. They showed that the mere presence of a Black experimenter reduced the automatic negative evaluation of that category (Lowery, Hardin, & Sinclair, 2001). Irene Blair (2002) used yet another manipulation, that of mental imagery, to show a similar effect on the gender–strength association. These studies showed, for the first time using the IAT as a dependent variable, that ordinary experiences such as the presence of a person or imagery could change implicit attitude and stereotypes that were not responsive to the dictates of a more effortful conscious desire to invoke change. This was intriguing because by their very nature, measures of implicit evaluation were expected to be tapping accumulated cognitions and therefore thought to be insensitive to trivial mental reorderings.

Knowledge of these findings long before they appeared in print did not persuade me but kept me sufficiently prepared to understand an outcome obtained in my own collaboration with Brian Nosek. For several years, we have conducted numerous studies of implicit academic orientations, in particular attitudes toward math and science as a function of implicit gender identity and gender stereotypes of a natural association between male (rather than female) with math and science. As we demonstrated, there is a strong gender difference in automatic attitudes toward math and science, with women showing more negative attitudes than men (Nosek, Banaji, & Greenwald, 2002b). This effect, obtained many times over, was not one that we were prepared to see disappear, but it did in a recent experiment, a finding that proved baffling. A finer-grained analysis of the data showed that the standard pattern of a gender difference was obtained but only when a male experimenter conducted the study. An opposite effect, with female participants 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 predicted and detected similar effects on math attitudes and math performance. Blair (2002) provided a meta-analysis of 50 experiments that show evidence of such malleability.

Experiments such as these challenge the naïve view, one that I may have harbored, 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 (Kim & Greenwald, 1998). Why then should something as mundane as consideration of positive African American exemplars (and negative European 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, and the implication of this may be more far-reaching than is possible to recognize at this time. The first error was 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, however, we came to believe that if uncontrollable via conscious will, the attitude is also unlikely to be influenced through other, less willful efforts. Yet, as the work of Bargh and colleagues has demonstrated so significantly (see Bargh & Chartrand, 1999, for a review), influences that come from outside of oneself can produce direct effects on behavior. The surprising finding common to the studies discussed here is 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 vary as a function of the elasticity of the attitude

object. A social group or a person (or any social object for that matter) has many facets, some of which are positive and others negative. 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 habitual evaluation of the group. Social groups, however, do not have a single evaluation attached to them: Multiple features are available, and each of them has the potential to influence the momentary representation that is formed and that constitutes the basis of the automatic evaluation. To ask one's attitude toward Chinese Americans will produce quite differing outcomes depending on the task but also on 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 is a plethora of ways to define 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, although that idea has intuitive appeal.

Because these are recent studies, they do not yet provide an interpretation of the nature of what they reveal. 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 feels as if the particular attitude has "changed." My assessment at this time 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 in this chapter reveal the importance of situational construal of attitude. Mitchell, Nosek, and Banaji (2002) 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 we psychologists can muster the imagination to think this way about attitudes, then such shifts need not suggest that "change" in the traditional sense has taken place. Rather—as Asch (1940) 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 by the features that exist in a given elicitation condition. The experiments by Blair, Dasgupta, Hardin, Nosek, and Mitchell indicate a renewed and even more rigorous emphasis on the acute properties of social situations. They also ought to 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 which can be accompanied by the intuition of a change from time 1 to time 2.

Mitchell, Nosek, and Banaji (2002) showed that attitudes toward exactly the same exemplars can be dramatically different as a function of the frame for processing or as a function of 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, Nosek, and Banaji 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 they 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 merely simplistic. It also 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: that even implicit attitudes can be situationally created. The experiments reviewed indicate a different mode for nondominant attitudes to be created: repeated, pervasive, and even minor interventions. The findings support those who have claimed that the shape and content of the immediate environment—the gender of the calculus teacher, the accent of the defense lawyer, the ethnicity of the janitor, the nationality of the scientist—matter. To extrapolate, what sits on our screensavers, the pictures that hang in our hallways, the advertisements that fill almost every social vacuum, and so on have the potential to influence not just our consciously framed ideas and feelings but perhaps especially our automatic and unintentional ones. 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 potential for 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. To do that will require understanding that environments constantly signal who can be what, and it appears that our minds are surprisingly sensitive to detecting such information and yielding to it.

Conclusion

This chapter is concerned with two discoveries that later revealed their opposites as well. If implicit and explicit attitudes were disjointed, it seemed 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. It now appears, however, that each of these original discoveries and their opposites are true. Conscious and unconscious

forms of evaluation are both independent and associated. Implicit attitudes may not bend to the instruction of conscious will, but they seem to be elastic in their response to even subtle suggestions in the environment. Theories of implicit social cognition must 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 to be cultivated—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 gain from their simplicity, ease, and please-all quality. Yet, as McGuire proposed 30 years ago, appreciating the virtues of paradoxical propositions, even fully contradictory ones, must become part of the daily activities of science (McGuire, 1973). The difficulty of practicing oppositional thinking in spite of consciously approving of its benefits, of reaching for the skill but never quite mastering it, is perhaps itself a process that fits into this evolving, never quite complete, view of doing science.

I have paid homage to the seventh koan but not to its author. To do so, I move a few thousand miles west to a different culture from which McGuire began the "Yin and Yang" paper in the Far East—the Indian epic of unparalleled length and complexity, the *Mahabharata,* which revolves 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, 1951). Without hesitation Arjuna picks the unarmed Krishna, leaving the evil Duryodhana happy at the prospect of gaining a full army. Those who know which side won also know a whole lot more: about Krishna and Arjuna's dialogues that 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. If I had Arjuna's choice, 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," and 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—give us the same gift as Krishna gave to Arjuna: words of a particular sort at moments of crisis, not answers but deeper questions, not instructions but revelations, not dismay but hope, not about his way to do it but about the discovery of yours.

References

Asch, S. E. (1940). Studies in the principles of judgments and attitudes: II. Determination of judgments by group and by ego standards. *Journal of Social Psychology, 12,* 433–465.

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

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54,* 462–479.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6,* 242–261.

Cleary, T. (1994). *Instant Zen—Waking up in the present.* Berkeley, CA: North Atlantic Books.

Cunningham, W. A., Nezlek, J., & Banaji, M. R. (2002). *Conscious and unconscious ethnocentrism: Revising the ideologies of prejudice.* Unpublished manuscript, Yale University, New Haven, CT.

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). Exposure to admired group members reduces automatic intergroup bias. *Journal of Personality and Social Psychology, 81,* 800–814.

Gregg, A. & Banaji, M. R. (1999). Malleability of attitudes as a function of biological states. Unpublished data.

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.

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

Lowery, B. S., Hardin, C., & 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., & Banaji, M. R. (2002). Contextual variations in implicit evaluation. Unpublished manuscript, Harvard University, Cambridge, MA.

Nosek, B. A., & Banaji, M. R. (in press). (At least) two factors mediate 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., & 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., & Greenwald, A. G. (2002b). Math = Male, Me = Female, therefore Math ≠ Me. *Journal of Personality and Social Psychology, 83,* 44–59.

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

Sherman, S. J., Presson, C. C., Chassin, L., Rose, J. S., & Koch, K. (2003). Implicit and explicit attitudes toward cigarette smoking: The effects of context and motivation. *Journal of Social and Clinical Psychology, 22,* 13–37.