
PRESIDENTIAL ADDRESS

Beyond Bipolar Conceptualizations and Measures: The Case of Attitudes and Evaluative Space

John T. Cacioppo

*Department of Psychology
Ohio State University, Columbus*

Wendi L. Gardner

*Department of Psychology
Northwestern University*

Gary G. Berntson

*Department of Psychology
Ohio State University, Columbus*

All organisms must be capable of differentiating hostile from hospitable stimuli to survive. Typically, this evaluative discrimination is conceptualized as being bipolar (hostile–hospitable). This conceptualization is certainly evident in the area of attitudes, where the ubiquitous bipolar attitude measure, by gauging the net affective predisposition toward a stimulus, treats positive and negative evaluative processes as equivalent, reciprocally activated, and interchangeable. Contrary to conceptualizations of this evaluative process as bipolar, recent evidence suggests that distinguishable motivational systems underlie assessments of the positive and negative significance of a stimulus. Thus, a stimulus may vary in terms of the strength of positive evaluative activation and the strength of negative evaluative activation it evokes. Low activation of positive and negative evaluative processes by a stimulus reflects attitude neutrality or indifference, whereas high activation of positive and negative evaluative processes reflects attitude ambivalence. As such, attitudes can be represented more completely within a bivariate space than along a bipolar continuum. Evidence is reviewed showing that the positive and negative evaluative processes underlying many attitudes are distinguishable (stochastically and functionally independent), are characterized by distinct activation functions (positivity offset and negativity bias principles), are related differentially to attitude ambivalence (corollary of ambivalence asymmetries), have distinguishable antecedents (heteroscedacity principle), and tend to gravitate from a bivariate toward a bipolar structure when the underlying beliefs are the target of deliberation or a guide for behavior (principle of motivational certainty). The implications for societal phenomena such as political elections and democratic structures are discussed.

This article is based on John T. Cacioppo's presidential address to the Society for Personality and Social Psychology, New York, August 12, 1995.

This research was supported by National Science Foundation Grants SBR-9512459 and DBS-9211483, by National Institute of Mental Health Grant P50MH52384-01A1, and by funding from the John D. and Catherine T. MacArthur Foundation Program on Mind-Body Interactions.

We thank David Messick, Jon Krosnick, John Bargh, Marilynn B. Brewer, William von Hippel, Timothy C. Brock, Sarah T. Boysen, Stephen L. Crites, Mary A. Snyder-Smith, Allyson Ankerbrand, Penny Visser, Barbara L. Andersen, and Bert N. Uchino for their comments on or contributions to various aspects of this research.

Requests for reprints should be sent to John T. Cacioppo, Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, OH 43210-1222. E-mail: cacioppo.1@osu.edu.

The causes and consequences of social behavior are readily observable and appear to be self-evident. From inquisitive journalists to power-seeking politicians, from well-intentioned physicians to chance seat partners on airplanes, many seem to believe they already know how families work, why people live in poverty, what fosters violence, and why Johnny can't read. Some have therefore reasoned that all but reductionistic scientific analyses of social behavior are superfluous. In recent years, genetic markers of alcoholism, obesity, and impulsivity have been heralded as major scientific breakthroughs in the media whereas demonstrations of situational and social determinants have been largely ignored.

The problem of obviousness is pervasive in the social sciences (Lazarsfeld, 1949; Nisbett & Wilson, 1977). It may appear self-evident that the opposite of good is bad, and equally intuitive that the same stimulus can arouse feelings of both good and bad. It appears obvious that opposites attract and that birds of a feather flock together; it is less apparent that both of these intuitions cannot generally be true. According to Lazarsfeld (1949):

It is hard to find a form of behavior that has not already been observed somewhere. Consequently, if a study reports a prevailing regularity, many people respond to it by thinking, "Of course that is the way things are." (p. 379)

Among the tasks confronting social scientists are sorting through axioms, developing theories that specify the conditions under which each holds, and devising measures that illuminate the psychological operations underlying the obvious. All three of these tasks are at least as important as the pursuit of reductionism (Cacioppo & Berntson, 1992). Early theoreticians and contemporary researchers have noted that positivity and negativity may be separable, for instance, yet bipolar conceptualizations and self-report measures of attitudes and affect have continued to dominate the field (D. P. Green, Goldman, & Salovey, 1993; cf. Cacioppo & Berntson, 1994).

The problem of obviousness may be exacerbated by the reliance on verbal reports but may also be attributable in part to the relative youth of the social sciences. To anyone who has ever participated in or watched a contact sport, Newton's second law of motion—force = mass * acceleration—is hardly surprising or counter-intuitive. However, it also appeared obvious that the sun circled the earth. As personally compelling and occasionally accurate as intuition can be, older fields of science have sufficient histories with the foibles of intuition and are less biased to accept commonsense

axioms without the same rigorous empiricism to which nonintuitive hypotheses are subjected.

People's perceptions, interpretations, and construals of stimuli and events in their world are clearly powerful determinants of behavior. It is worth remembering, however, that these mental states are neither universally accurate nor the only determinants of social behavior (e.g., Bargh, 1994; Greenwald & Banaji, 1995). Intuitions and reportable mental states are important phenomena for social scientists to quantify and understand, just as it was important for physicists and astronomers to understand why it was that the sun appeared to be circling the earth. However, just as the personal experience that the earth is motionless obstructed rather than illuminated the underlying processes that produced this percept, intuitions can subtly hinder the understanding of social phenomena. This is particularly the case when reportable mental states are assumed to reflect explanations (processes) rather than psychological endpoints or states (Nisbett & Wilson, 1977).

Attitudes and Other Bipolar Constructs

In the present article, we consider the case of attitudes as a psychological phenomenon and the ubiquitous bipolar attitude scale as the measurement instrument. The concept of attitude remains one of the most important in personality and social psychology. Among the reasons the concept of attitudes has occupied a central role in psychology are that attitudes (a) are a pervasive aspect of mental and social life, (b) influence a wide range of decisions and behaviors, (c) act as a summary statistic for knowledge about and reactions to stimuli, and (d) reduce the effort or stress of decision making.

The dominant conceptualization of attitudes as a unidimensional evaluative categorization or response dates back to the time of Thurstone (1928, 1931):

Several years ago when I was teaching conventional psychophysics, it seemed to me that psychophysics was really a very dull subject in spite of the fact that it did offer the satisfaction of clean and quantitative logic. ... One way in which to retain the satisfactions that can be found in the logic of this subject is to change its content. ... Instead of asking a person, "Which of these two little cylinders is the heavier?" ... we might as well ask him something interesting, such as, "Which of these two nationalities do you in general prefer ...?" or, "Which of these two pictures or colored designs do you like better?" (Thurstone, 1931, pp. 249–250)

Thurstone (1928, 1931) used psychophysical phenomena such as brightness and temperature as a model for his bipolar conceptualization and measure of attitudes.

The opposite of dim, for instance, is bright (i.e., bright and dim are reciprocally related), and bright and dim can be treated as polar opposites and measured accordingly. The positive and negative processes underlying attitudes were similarly treated as being polar opposites and have traditionally been measured on bipolar scales. Indeed, the bipolarity of attitudes and attitude measures has only occasionally been questioned (cf. Cacioppo & Berntson, 1994; R. F. Green & Goldfried, 1965).

Laypersons might ask what could possibly be wrong with the conceptualization of attitudes as bipolar or the measurement of attitude using bipolar scales. After all, is the dichotomy between positive and negative not obvious, stable, expected, and psychologically harmonious? Is it not self-evident that the opposite of harmful is beneficial, optimism is pessimism, wise is foolish, and good is bad? Is not the bipolarity of attitudes one of the few generally accepted principles across the social sciences?

The bipolar conceptualization of attitudes implies that the positive and negative evaluative processes underlying attitudes are reciprocally activated and interchangeable; as an individual's negative regard for a stimulus increases, the net effect—in terms of the attitude—is that the positive regard for the stimulus decreases. The endorsement of positive beliefs about an attitude object is treated as equivalent to the rejection of negative beliefs about the object. Accordingly, the evaluative processes elicited by an object are thought to range along a single continuum between two endpoints: maximally positive (and minimally negative) to maximally negative (and minimally positive). Attitudes have thus been conceptualized (and measured) as being analogous to the position of a balance knob on an audio stereo, with very negative and very positive substituted for left speaker and right speaker.

The assumption that positive and negative evaluative processes are reciprocally controlled and interchangeable was an important starting point that simplified the measurement of attitudes: The endpoints could be placed on a single (bipolar) continuum, and attitudes could be measured along this continuum. Furthermore, the bipolar conceptualization underscored the principle that positivity and negativity generally had antagonistic effects on attitudes and behavior (e.g., Lewin, 1935). This conceptualization and measurement strategy fit well with the behavioral dispositions of approach-withdrawal, and some early attitude theorists working during the heyday of behaviorism defined attitudes as nothing more than a bivalent action disposition toward a stimulus (e.g., Doob, 1947). With the advent of the cognitive paradigm in psychology, attitude theorists began distinguishing evaluative categorizations of stimuli from bivalent action dispositions toward stimuli (e.g., Berntson, Boysen, & Cacioppo, 1993; Cacioppo,

Crites, Berntson, & Coles, 1993; Crites, Cacioppo, Gardner, & Berntson, 1995; Zanna & Rempel, 1988). Nevertheless, the notion that the evaluative processes underlying attitudes are reciprocally activated continued largely unchallenged.

In the present article, we propose a conceptualization of attitudes that emphasizes their motivational substrates and, in particular, the distinctions between the motivational processes underlying positive and negative evaluative processes. These motivational substrates are conceptualized as two fundamental, evolutionarily adaptive, information-organizing, goal-oriented, action-guiding systems that are instantiated in distinctive brain mechanisms (Cacioppo & Berntson, 1994). These motivational substrates are further assumed to underlie primitive affective as well as attitudinal processes. The term *evaluative processes* is used here as a general term to refer to any means by which organisms distinguish hostile from hospitable stimuli in their environment. Evidence in the literature and from our laboratory is reviewed indicating that the positive and negative evaluative processes underlying many attitudes (a) are distinguishable (stochastically and functionally independent), (b) are characterized by distinct activation functions (positivity offset and negativity bias), (c) are related differentially to attitude ambivalence (corollary of ambivalence asymmetries), (d) have distinguishable antecedents (heteroscedacity), and (e) tend to gravitate from a bivariate toward a bipolar structure when the underlying beliefs are the target of deliberation or a guide for behavior (desire for motivational certainty).

Much has been written in recent years about the structure and measurement of positive and negative moods (e.g., Diener & Iran-Nejad, 1986; D. P. Green et al., 1993; Lang, 1995; Larsen & Diener, 1992; Russell, Lewicka, & Niit, 1989; Tellegen, 1985; Watson, 1988). We are concerned here, however, with evaluative processes and the resulting attitudes, not moods or people's conceptual organization of emotions. As should become clear, there may be important differences in the processes operating to discriminate hostile from hospitable stimuli and those underlying people's moods and conceptual organizations of attitudes and emotions.

We should also be clear from the outset that we are not suggesting that positive and negative evaluative processes are agonistic in their effects on attitudes and behavior but instead we, too, assume their effects are generally antagonistic. The set of assumptions justifying the measurement of attitude using bipolar scales might be expressed as follows (Cacioppo & Berntson, 1994):

1. An attitude is a joint function of positive (appetitive) and negative (aversive) affective/motivational reactions to a stimulus.

2. Positive and negative reactions to a stimulus have generally opposing effects on an attitude.
3. The positive and negative reactions that determine an attitude toward a stimulus are essentially reciprocally controlled.¹

The assumption that is being questioned is not the second—the antagonistic effects of positive and negative hedonic processes on attitudes—but the third: the uniformly bipolar (reciprocal) activation of the positive and negative evaluative processes underlying attitudes. We posit that reciprocal effects between positive and negative evaluative processes do exist but so do nonreciprocal effects.

Bipolar Versus Bivariate Conceptualizations

To be sure, there are in fact bipolarities and dichotomies in the world. The physical phenomena of hot and cold are bipolar concepts because cold represents literally the absence of heat. As cold increases, heat decreases. The transfer (i.e., activation) functions for hot and cold are essentially mirror images of one another. Reducing the physical heat of an object is equivalent to increasing the coolness of the object.

Although there are bipolarities in the world, there is also a natural tendency for people to simplify their worlds by introducing artificial bipolarities and dichotomies when assigning psychological meaning. People filter, level, and simplify information during encoding and decoding. Is it not likely that people simplify their affective reactions to stimuli as well? If so, then the treatment of the positivity and negativity aroused by an attitude object as falling along a bipolar continuum may conform well to common personal experience, but this experience—like the appearance of the sun circling the earth—may reflect a psychological or behavioral endpoint or outcome, not a process. This endpoint is important to explain but its existence does not imply that the underlying processes are organized similarly.

Even though physical constraints may restrict behavioral manifestations to bivalent actions (ap-

proach-withdrawal), early behavioral theorists recognized that approach and withdrawal were behavioral manifestations that could come from distinguishable motivational substrates (Miller, 1951, 1961). Conflict theory was enriched by conceptualizing approach and withdrawal separately, investigating their unique antecedents and consequences, and examining the psychological and physiological constraints that typically led to the reciprocal activation of approach and withdrawal tendencies (Berntson et al., 1993; Lang, Bradley, & Cuthbert, 1990).

Attitude theory and measurement may similarly benefit from expanding the principle of reciprocal evaluative activation to accommodate (a) the distinguishable activation of positive and negative evaluative processes, (b) the investigation of their unique antecedents and consequences, and (c) the examination of the psychological and physiological constraints that produce their reciprocal activation. We recently proposed such a model in which the activation of positivity and of negativity is conceived as representing the outcome of distinguishable motivational (and brain) processes with only partially overlapping antecedents and somewhat different transfer functions (Cacioppo & Berntson, 1994). Furthermore, nonreciprocal and independent (uncoupled) as well as reciprocal effects of positive and negative evaluative processes were posited to exist. One of the implications of this model is that a bipolar dimension and measure may be insufficient to portray comprehensively the evaluative processes underlying attitudes; instead, a two-dimensional representation of positive and negative evaluative activation may be minimally required.

The bivariate evaluative space depicted in Figure 1, for instance, accommodates all possible combinations of positive and negative evaluative activation: (a) the reciprocal mode of evaluative activation is represented as one diagonal vector that ranges from maximal positivity/minimal negativity to maximal negativity/minimal positivity, (b) the nonreciprocal mode of evaluative activation is represented as the alternate diagonal that ranges from minimal positivity and negativity to maximal positivity and negativity, and (c) the uncoupled modes of evaluative activation are represented as vectors lying along the axes. The family of vectors parallel to the aforementioned represent the general categories, or modes, of evaluative activation expressed from varying starting points within the two-dimensional plane depicted in evaluative space. The essential question in a bivariate framework is not whether positive and negative evaluative processes are reciprocally activated, but rather under what conditions are they reciprocally, nonreciprocally, or independently activated. Thus, this bivariate evaluative space does not reject the bipolar conceptualization but rather subsumes it as the reciprocal diagonal and the family of vectors parallel to this diagonal.

¹In an important theoretical development in the area of emotions for instance, Lang, Bradley, and Cuthbert (1990) reviewed evidence for the separability of appetitive and aversive motivational systems and for the antagonistic effects of each system. However, they also assumed a reciprocal relationship between these systems (cf. Lang, 1995). They stated, for instance, that "a reciprocal rule is also postulated: A mismatched reflex (e.g., a defensive reflex initiated during a pleasant state of affairs) will show relative inhibition or attenuation" (p. 377), and that "these systems were held to have reciprocal inhibitory connections that modulated learned responses and reactions to new, unconditioned input" (p. 381).

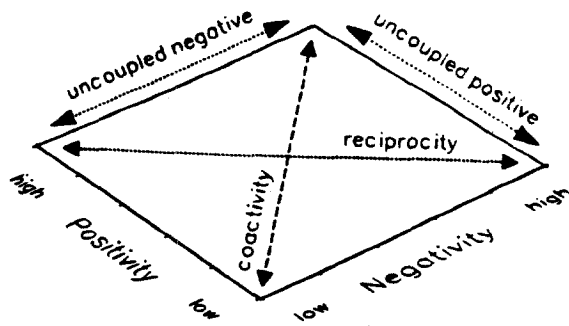


Figure 1. Bivariate evaluative space. The left axis represents the level of excitatory activation of positive evaluative processes (labeled *positivity*), and the right axis represents the level of excitatory activation of negative evaluative processes (labeled *negativity*). Along each axis, the level of activity increases with movement away from the front axis intersection. The dotted diagonal extending from the left to the right axis intersections represents the diagonal of reciprocal control (labeled *reciprocity*). The dashed diagonal extending from the back to the front axis intersections depicts the diagonal of nonreciprocal control (labeled *coactivity*). The arrows alongside the axes represent uncoupled changes in positive or in negative evaluative processing. These diagonals and axes, and vectors parallel to them, illustrate the major modes of evaluative activation. *Note.* From "Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of Positive and Negative Substrates," by J. T. Cacioppo and G. G. Berntson, 1994, *Psychological Bulletin*, 115, p. 402. Copyright © 1994 by the American Psychological Association. Reprinted with permission.

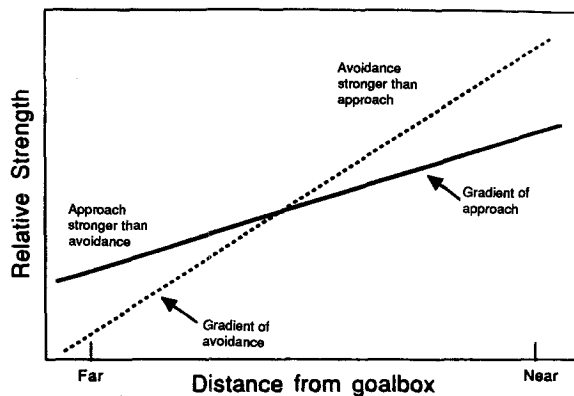


Figure 2. Graphic representation of the dynamics of an approach-avoidance conflict. *Note.* From "Liberalization of Basic S-R Concepts: Extensions to Conflict Behavior, Motivation and Social Learning," by N. E. Miller, in S. Koch (Ed.), *Psychology: A Study of a Science, Study 1* (p. 407), 1959, New York: McGraw-Hill. Copyright © 1959 by the McGraw-Hill Companies. Reprinted with permission of the McGraw-Hill Companies.

If the positive and negative hedonic processes underlying people's attitudes are not polar opposites, then evidence of uncoupled and nonreciprocal activation as well as reciprocal activation should exist. Indeed, a particularly powerful type of data for a bivariate rather than bipolar framework is evidence that both positive and negative valent forces can be activated by the same

manipulation. The classic research by Neal Miller (1948, 1959, 1961, 1966) is an interesting early case in point. Miller's (1959) research on conflict demonstrated that animals approached the goal when distant from it, withdrew from the goal when near it, and hesitated when at the distance from the goal at which the approach and avoidance gradients intersect. Furthermore, Miller and colleagues found that both the strength of approach and the strength of avoidance (or withdrawal) increased as the distance from the goal decreased (as depicted by the positive gradients in Figure 2). This result suggested that not only were positive and negative processes distinguishable in their influence on approach-withdrawal behavior, but the same spatial cue (e.g., distance from the goal box) could produce a coactivation of positive and negative motivational processes. These observations are in complete accord with the model of evaluative space, where the difference between the activation of positivity and the activation of negativity at any point within evaluative space reflects the strength and valence of the bipolar disposition arising as a result of the bivariate components.

Distinctions Between the Positivity and Negativity Underlying Attitudes

We have suggested that the motivational substrates underlying positive and negative evaluative processes and attitudes are two distinct, fundamental action-guiding systems that are instantiated in distinctive brain mechanisms, and that these motivational substrates can be reciprocally activated, individually activated, or nonreciprocally activated (i.e., coactivated, co-inhibited). Neurophysiological evidence across species dating back to the mid-1950s (Delgado, Roberts, & Miller, 1954; Olds, 1958; Olds & Milner, 1954) is consistent with the notion that positive and negative hedonic processes rely at least partially on distinct neural mechanisms. More recently, Gray (e.g., 1990, 1991) outlined separate motivational and neural substrates underlying approach and withdrawal behaviors based on his animal research, and LeDoux (1995) outlined neural circuitry involved in fear conditioning. The amygdala, for instance, has been implicated in goal-oriented withdrawal (e.g., fear) behaviors, whereas the nucleus accumbens—ventral tegmental area, lateral septal nucleus, and tenia tecta (in the prefrontal cortex)—are among the areas that may be involved in goal-oriented approach behaviors.

Brain imaging research in humans is also consistent with the notion that different cerebral structures are involved in positive and negative hedonic processes. In a recent study by George and his colleagues (George et al., 1995), regional cerebral blood flow was measured using positron emission tomography (PET) in 11

healthy women who induced happiness, sadness, or neutral emotions by recalling affect-laden or neutral memories while looking at congruent happy, sad, or neutral human faces. Comparisons between the sadness-minus-neutral and the happy-minus-neutral conditions revealed that, rather than reciprocal changes in blood flow to the same brain regions, a change from sad to happy affective state produced increased cerebral blood flow to entirely different brain regions. These results supply suggestive human evidence that the neural mechanisms involved in the experience of positive and negative states differ.²

There are psychological as well as physiological data that call into question the sufficiency of a bipolar conceptualization and measure of attitudes. With true bipolar constructs such as hot and cold, the value of one pole cannot be altered without there being an opposite and essentially equal effect on the other. When heat is removed from an element, the element becomes colder. Cold is the absence of heat. Events can, however, affect positivity without altering negativity, and vice versa. For instance, the efficacy of systematic desensitization in diminishing negative feelings—such as lessening the feelings of disgust upon seeing a cockroach—can be thought of as lowering the activation of negative motivational forces acting on an individual. This reduction in disgust upon seeing a cockroach through systematic desensitization need not increase positive motivational forces. Unlike cold and heat, a lessening of negativity does not necessarily produce concomitant increases in positivity, and vice versa.

This point is illustrated by Goldstein and Strube's (1994) study of students' reactions to academic success and failure. Goldstein and Strube (1994) administered Watson, Clark, and Tellegen's (1988) Positive and Negative Affect Schedule (PANAS) to students at the beginning and end of a class session on three successive class meetings. These class meetings were selected such that the participants completed the PANAS on the first and third day of the study and students received performance feedback on a class examination on the second day. The students were instructed on the first and third days to indicate on the PANAS how they felt "right now (that is, at the present moment)" (Goldstein & Strube, 1994, p. 59).

²These data should be interpreted cautiously. It is likely, for instance, that some of the structures that are involved in evaluative reactions are activated so transiently that they may not be identified by 1-min scanning procedures (Robinson, 1995). In addition, an association between changes in regional brain blood flow and transient emotional state does not mean they are causally linked (Sarter, Berntson, & Cacioppo, 1996). It is noteworthy, however, that research on patients with focal brain lesions, although preliminary, has provided converging evidence for the separability of the neural mechanisms underlying positive and negative evaluative (affective) processes. Robinson (1995), for instance, observed that a patient with a basal ganglia lesion did not report experiencing happy emotion but did experience fearful emotion.

Analyses of the responses to the PANAS indicated that students who scored above the mean (i.e., who received success feedback) showed elevations in positive affect but no change in negative affect, whereas students who scored below the mean (i.e., who received failure feedback) showed elevations in negative affect but no change in positive affect (see Figure 3). Goldstein and Strube (1994) further found nonsignificant correlations within each measurement period between the measures of positivity and negativity. Thus, not only were the momentary feelings of positivity and negativity independent, but the activation of positive and negative evaluative processes were uncoupled—with different events (success and failure feedback, respectively) associated with the uncoupled activation of positive versus negative evaluative processes (see also Matthews, Wilkinson, & Robbins, 1996).

The Goldstein and Strube (1994) study demonstrates that positive and negative affective reactions can be uncoupled, but it bears only indirectly on the independence of the positive and negative evaluative processes underlying attitudes. To address this issue, we (Cacioppo, Crites, Snyder-Smith, & Gardner, 1996) conducted a study in collaboration with Steve Crites and Mary Snyder-Smith to investigate the evaluative underpinnings of the attitudes of undergraduate women toward their college roommates. The participants in the study—all of whom lived in dormitories on campus—were asked to indicate how they felt about their roommate using the PANAS, to indicate the extent to

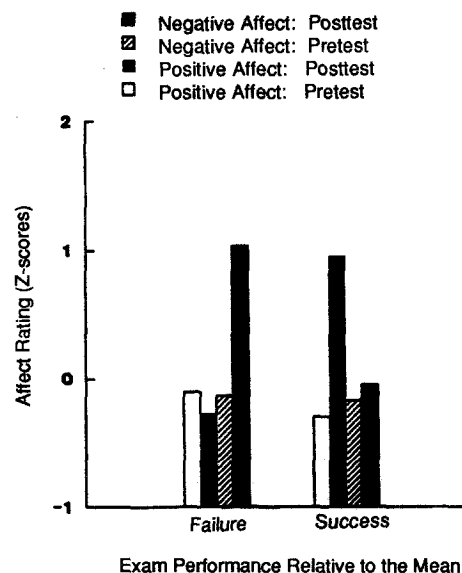


Figure 3. Standardized affect ratings as a function of exam performance, time of measurement, and type of affect. Note. From "Independence Revisited: The Relation Between Positive and Negative Affect in a Naturalistic Setting," by M. D. Goldstein and M. J. Strube, 1994, *Personality and Social Psychology Bulletin*, 20, p. 61. Copyright © 1994 by Sage Publications. Reprinted by permission of Sage Publications.

which they considered their roommate a friend and the amount of time they spent with their roommate, and to characterize their general mood using the PANAS. The study was conducted during the spring quarter of their first year in college, and all participants had lived with their roommates for more than two quarters.

Cronbach's alpha coefficients, which are depicted along the diagonal of Table 1, indicated that the internal consistencies of the scales measuring positivity and negativity were comparable and high. Correlational analyses summarized in Table 1 further suggested that the positivity and negativity elicited by the roommate were stochastically independent. That is, participants who felt very positively about their roommates were no more or less likely to feel negatively about the roommate than participants who felt only somewhat positively about their roommate. We (Cacioppo et al., 1996) further found evidence for functional independence between the positive and negative affects elicited by the roommate, as only the level of positivity predicted the extent of the friendship and the amount of time spent with the roommate (for similar evidence in the political arena, see Abelson, Kinder, Peters, & Fiske, 1982).

More recently, we examined the evaluative substrates of people's attitudes toward blood donation (Gardner & Cacioppo, 1995, 1996; see also Cacioppo & Gardner, 1993). In Study 1 of Gardner and Cacioppo (1996), 56 participants listed thoughts about blood donation. The 9 most frequently listed positive thoughts (e.g., "Donating blood would give me a good feeling afterward") and the 9 most frequently listed negative thoughts (e.g., "Donating blood would leave me weak afterward") about blood donation were used as items in Study 2.

In Study 2, 100 participants expressed their evaluations of blood donation by rating this topic on each of the 18 items generated in Study 1 using 5-point (-2 to +2) Likert-like scales. Reliability analyses resulted in the retention of 12 items. Factor analyses and scree tests revealed a two-factor solution accounting for 51.4% of the variance (6 Positivity items, Cronbach's $\alpha = .88$; 6 Negativity items, Cronbach's $\alpha = .70$). The Positivity and Negativity factors were uncorrelated ($r = -.08$, *ns*).

In Study 3, 317 participants completed the 12-item scale for assessing evaluations of blood donation that

was developed in Study 2. The Positivity and Negativity factors were again uncorrelated ($r = -.07$, *ns*), replicating the results of Study 2. Confirmatory factor analyses testing one-, two-, and four-factor solutions confirmed that the two-factor (Positivity $\times$ Negativity) solution was optimal (Cronbach's α s = .87 and .80, respectively). These data suggest that the positive and negative beliefs underlying people's attitudes toward blood donation are distinguishable (i.e., stochastically independent) and cohere within valence.

Upon completion of Study 3, the participants also completed a behavioral measure, used by the Red Cross, in which they expressed their willingness to donate blood (Gardner & Cacioppo, 1996). Four distinct groups of participants were identified based on median splits:

1. Participants for whom blood donation evoked low positivity and high negativity (negative attitudes).
2. Participants for whom blood donation evoked low positivity and low negativity (attitude neutrality or indifference).
3. Participants for whom blood donation evoked high positivity and high negativity (attitude ambivalence).
4. Participants for whom blood donation evoked high positivity and low negativity (positive attitudes).

Analyses of the measure of behavioral intentions indicated not only that participants with positive attitudes were more willing to donate blood than participants with negative attitudes, but that participants with neutral attitudes toward blood donation were less willing to donate than participants who felt ambivalent about donating blood. This result would not have been evident had we relied solely on bipolar attitude measures. More important, these data imply that attitudes that fall differentially within a bivariate (Positivity $\times$ Negativity) evaluative space, but would have to be placed comparably along the midpoint of a bipolar attitude scale, may

Table 1. *College Roommates: The Bivariate Evaluative Organization Underlying Attitudes and Behaviors*

	Roommate		General Mood	
	Positivity	Negativity	Positivity	Negativity
Positivity: Roommate	(.90)			
Negativity: Roommate	-.09	(.77)		
Positivity: Moods	.25	.02	(.85)	
Negativity: Moods	.01	.54*	.02	(.83)
Friendship	.45*	-.27	-.16	-.15
Time Spent With Roommate	.40*	-.04	-.23	-.15

* $p < .05$.

have different motivational substrates and behavioral consequences.

The evaluative substrates of social processes such as in-group favoritism and intergroup discrimination (Brewer, in press; Mummendey, 1995) and critical social attitudes such as racial prejudice also appear to adhere more to a bivariate than a bipolar conceptualization. Patchen, Hofman, and Davidson (1976), for instance, conducted a factor-analytic study of the racial attitudes of more than 4,000 desegregated Black and White high school students. Their results indicated that the positive and negative racial attitudes were not merely opposite ends of a continuum. Rather, one factor was characterized as a general positive evaluation of Blacks or Whites, and a second factor was characterized as a general negative evaluation of Blacks or Whites. Some individuals felt neither positively nor negatively about African Americans, whereas others felt both very positively and very negatively about African Americans. If the racial attitudes of these groups were meas-

ured on a bipolar scale, the attitudes of each would fall along the midpoint of the scale even though the motivational forces underlying these attitudes differ dramatically. Consistent with the results of Patchen, Davidson, Hofman, and Brown (1977), a sizeable literature has emerged on the distinctions between positive and negative racial attitudes (e.g., Hass, Katz, Rizzo, Bailey, & Moore, 1992; Katz & Hass, 1988; Katz, Wackenhut, & Hass, 1986; Patchen et al., 1977; Pettigrew & Meertens, 1995).

The results of these studies suggest that a two-dimensional representation of positive and negative evaluative activation may provide a more comprehensive formulation for understanding attitudes than the traditional bipolar model. One such representation is the bivariate evaluative plane depicted in Figure 1, which accommodates all possible combinations of positive and negative evaluative activation.

In the next section, we offer a model by which these evaluative processes map into attitudinal expressions. The theory described in the next section is designed to depict, at least in part:

1. How people respond to bipolar attitude scales (depicted on the vertical axis in Figure 4).
2. The deeper composition (motivational substrates) of points on bipolar attitude scales (depicted as loci within the Positivity $\times$ Negativity plane in Figure 4).
3. How these motivational subsystems are combined to produce an attitudinal expression (depicted in the computational model by which points on the evaluative plane map into the overlying attitudinal surface; e.g., see Equation 1).
4. The various modes of evaluative activation (reciprocal, uncoupled, and nonreciprocal modes of evaluative activation).

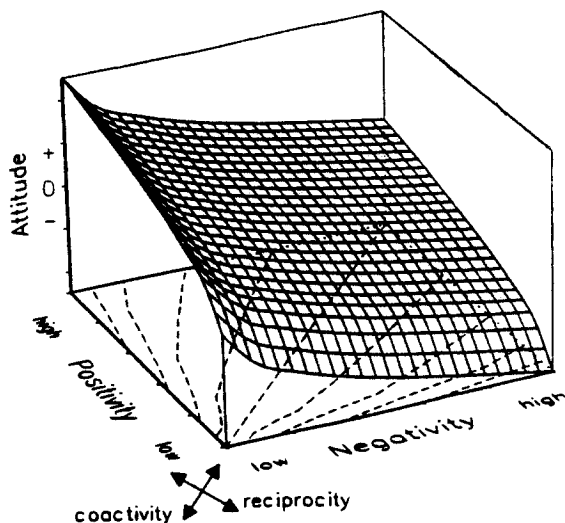


Figure 4. Illustrative bivariate evaluative space and its associated attitudinal surface. This surface represents the attitudinal disposition of an individual toward (+) or away from (-) the target stimulus. The attitudinal disposition is expressed in relative units and the axis dimensions are in relative units of activation. The point on the surface overlying the left axis intersection represents a maximally positive attitude, and the point on the surface overlying the right axis intersection represents a maximally negative attitude. Each of the points overlying the dashed diagonal extending from the back to the front axis intersections depict the same midpoint attitude. Thus, the nonreciprocal diagonal on the evaluative plane—which represents different evaluative processes (e.g., neutral to ambivalent)—yields the same middling expression on the attitudinal surface. Dashed lines (including the coactivity diagonal) represent isoattitudinal contours on the evaluative plane. These isoattitudinal contours are illustrative rather than exhaustive. *Note.* From "Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of Positive and Negative Substrates," by J. T. Cacioppo and G. G. Berntson, 1994, *Psychological Bulletin*, 115, p. 412. Copyright © 1994 by the American Psychological Association. Reprinted with permission.

The Evaluative Space Model of Attitudes: By the Numbers

The Positivity $\times$ Negativity plane depicted in Figures 1 and 4 represent the status in psychological (evaluative) space, whereas the resulting behavioral predisposition to approach or withdraw can be represented in terms of an overlying surface. That is, it is possible to derive an attitudinal surface for all combinations of positivity and negativity. The exact shape of the attitudinal surface is dependent on the form of the activation functions, the relative value of the weighting coefficients, and the scaling of the axes. For purposes of an initial illustration, the weights in Figure 4 were assumed to be equal, the activation functions for positivity and for negativity were assumed to have an exponent of 0.5 (e.g., power functions representative of psychophysical functions), and the interaction and error terms were assumed to be zero. The resulting attitudinal

surface overlying the evaluative plane is that depicted in Figure 4.

There are a number of features and properties of this evaluative plane and attitudinal surface that are worth noting here because they depend only on the positivity and the negativity functions being monotonic. First, it distinguishes between traditional attitude expressions (and the general approach-withdrawal tendencies) that are depicted by loci on the attitudinal surface and the underlying motivational substrates that are depicted by loci on the evaluative plane. J. D. Edwards and Ostrom (1971), for instance, investigated attitude ratings of "neutral attitudes" that derived from (a) all neutral experiences with an attitude object or (b) a wide range of positive, neutral, and negative experiences. In the bivariate model of evaluative space, the former experiences are conceptualized as producing a neutral attitude and a middling rating on a bipolar attitude scale because neither positive nor negative evaluative processes were activated, whereas the latter more varied experiences are conceptualized as producing a more ambivalent attitude and a middling rating on a bipolar attitude scale due to the coactivation of positive and negative evaluative processes. Edwards and Ostrom (1971) found that although these different substrates of neutral attitudes were apparent in their belief protocols, these attitude differences were not evident using Fishbein and Raven's (1962) five semantic differential scales (*bad-good*, *harmful-beneficial*, *wise-foolish*, *sick-healthy*, and *dirty-clean*) or a self-rating scale that ranged from 1 (*highly unfavorable*) to 8 (*highly favorable*; Anderson, 1965). Inspection of the vertical axis associated with loci on the attitudinal surface in Figure 4 conforms to this empirical finding in that changes in the evaluative processes underlying an attitude (i.e., changes in values on the evaluative plane) may not result in measurable changes using bipolar scales (i.e., values on the attitudinal surface).

The relationship between loci on the evaluative plane and points on the attitudinal surface is generally characterized by a many-to-one mapping. That is, each coordinate on the evaluative plane dictates a specific value on the attitudinal surface, but it is not the case that each value on the attitudinal surface translates unequivocally into a specific coordinate on the evaluative plane. It is for this reason that the evaluative (psychological) processes underlying an attitude may change without an apparent change on a bipolar attitude scale.

This is most obvious along (although not limited to) the coactivity diagonal, which extends from the front to the back corners of the evaluative plane. For any level of negative activation, the surface rises as positivity increases; for any level of positive activation, the surface falls as negativity increases. Therefore, as one moves from indifference to ambivalence, the surface rises in the direction of positivity and falls in the direction of negativity so the net changes are relatively small

or negligible. Thus, as is apparent from inspecting Figure 4, all of the points along the coactivity diagonal map into the same value on the attitudinal surface. At the front corner of the evaluative plane, the activation level of both positivity and negativity equals zero, the resulting positivity and negativity equal zero, and the combination of these outputs is zero. The eliciting stimulus in this case evokes neutrality—that is, no affectivity (as reflected by the locus on the evaluative plane)—and the attitudinal response would be a middling position on a bipolar attitude scale (as reflected by the locus on the attitudinal surface). At the far corner, in contrast, the activation level of both positivity and negativity is maximal and their outputs are asymptotic, but the combination of these outputs still equals zero. The eliciting stimulus in this case evokes a great deal of affectivity and ambivalence (as reflected by the locus on the evaluative plane), but the attitudinal response on a bipolar scale is again a middling position (as reflected by the locus on the attitudinal surface).

The coactivity diagonal can be thought of as an isoattitudinal expression contour—a set of distinct loci on the evaluative plane that map into the same attitudinal expression (i.e., many-to-one mapping). It is important to note that many isoattitudinal contours exist on the evaluative plane (e.g., see the dashed lines in Figure 4). Furthermore, this form of many-to-one mapping may render ambiguous the psychological significance of an attitude expression. That is, measurements obtained from the attitudinal surface that fall above an isoattitudinal contour do not allow unambiguous inferences to be drawn about a person's attitudinal disposition using a bipolar index because any such measurement has multiple potential origins in the evaluative plane, each of which has quite different psychological implications (cf. Cacioppo & Tassinary, 1990). Thus, isoattitudinal contours represent regions of indeterminism in the mapping from the attitudinal surface (e.g., bipolar attitude scale responses) onto the underlying evaluative plane. Indeed, the studies just outlined suggest that the behavioral implications for neutral and ambivalent attitudes may differ in important respects (see also Lewin, 1935).

Attitude theory and research are among the richest and most developed in the social sciences. We do not mean to imply that measurements of the attitudinal surface have no bearing on their underlying origins as represented in the bivariate evaluative plane. Note, for instance, that as the locus on the evaluative plane moves toward either the left or the right corner in Figure 4, the region of indeterminism becomes smaller until, at the extreme left and right corners, one-to-one mappings are achieved between the points on the evaluative plane and the attitudinal surface. The diagonal connecting these two corners, of course, reflects the reciprocal modes of evaluative activation—the implicit assumption underlying the ubiquitous bipolar scale (and, consequently,

many theories) in the area of attitudes. Given that the only reasonably good mapping between surface measures and their underlying psychological origin is found in regions near the corners of the reciprocal diagonal, it perhaps should not be surprising that the past 60 years of research on attitudes can be interpreted as falling along this diagonal. Factors that produce nonreciprocal movements along the evaluative plane (e.g., the coactivation of positivity and negativity) produce attitudinal manifestations that are largely indistinguishable using bipolar scales. In the absence of distinguishable manifestations, these factors appear to have little or no effect on attitudes and would be assigned little or no significance in attitude theories. Thus, bipolar attitude conceptualizations and scales may have shaped and constrained attitude theories in more fundamental and substantial ways than generally recognized. Other regions of the evaluative plane may be at least as important to understand, but the interpretive restrictions of traditional attitude conceptualizations and measures may have limited theory and investigation of these regions.

The model presented in Figure 4 depicts the transfer functions for positivity and negativity as being identical, and we believe this model is an oversimplification. We and numerous other investigators have suggested that there are differential activation functions for positivity and negativity (e.g., Cacioppo & Berntson, 1994; Gilbert, 1993; Miller, 1959; Mummendey, 1995; Peeters & Czapinski, 1990; Skowronski & Carlston, 1989). We review representative evidence in the next section, and we present a more complete computational model of evaluative space and attitudes that takes these differential activation functions into account.

Differential Activation Functions for Positivity and Negativity

Recall that Neal Miller and colleagues demonstrated that the manipulation of positive and negative environmental contingencies in the same situation consistently had opposing effects on approach-withdrawal behaviors. Miller (1959) also posited that when positive and negative motivational forces were coactivated, the stronger dominated (although vacillation was predicted and observed; see Cacioppo & Berntson, 1994, Table 1). The research by Miller and colleagues points to yet another potential shortcoming in the bipolar conceptualization of attitudes. The transfer functions for truly bipolar constructs, such as for hot and cold, are essentially identical in form.³ For instance, the rate at which an element cools is essentially the same rate at which heat dissipates. Hot and cold are endpoints on a bipolar

continuum, and the transfer functions for these endpoints are equal and reciprocal. In contrast, the transfer functions of positivity and negativity appear to differ in several noteworthy respects.

Positivity Offset

Research on approach-withdrawal conflict provided early evidence for what we have termed a *positivity offset*.⁴ As is illustrated in Figure 2, the positivity offset is the tendency for there to be a weak positive (approach) motivational output at zero input (i.e., intercept difference). As a consequence of the positivity offset, the motivation to approach is stronger than the motivation to avoid at distances far from the goal (i.e., at low levels of evaluative activation).

What might be the possible evolutionary significance of the positivity offset? Without a positivity offset, an organism in a neutral environment may be unmotivated to approach novel objects, stimuli, or contexts. This is especially true given the neophobia of most species. Neophobia likely evolved in many species because it helps protect organisms against potential predators. In terms of the evaluative space model, foreign stimuli are conceived as activating negativity much as might an aversive unconditioned stimulus. That is, in terms of Equation 1, these stimuli increase $N_g(j)$. Unlike most aversive unconditioned stimuli, however, the fear of a foreign or novel stimulus is conceived as habituating quickly in the absence of negative outcomes. Such organisms would learn little about novel or neutral-appearing environments and their potential reward (e.g., nutrient) value.

With a positivity offset, however, an organism facing neutral or unfamiliar (i.e., nonvalenced) stimuli would be weakly motivated to approach and—with the quick habituation of the initial fear response (i.e., the diminution of the evoked negativity by foreign stimuli or startling events)—to engage in exploratory behavior. Such a pairing of tendencies may have important survival value, at least at the level of a species. The neophobic response to foreign stimuli permits an initial period of observation. With no negative outcomes, this exposure allows the initial neophobic response to habituate, thereby allowing exploratory behavior to mani-

³A *transfer function* is the form of the output of a system given linear input across the dynamic range of the system.

⁴In this article, differences in *offset* in the transfer functions for positivity and negativity refers to differences in intercept (i.e., output at zero input or activation), whereas differences in *bias* in the transfer functions refers to differences in slope (i.e., increase in output per quantum of input or unit of activation). The term *positivity offset* is preferred to *positivity bias* because it refers to the positive motivational output at zero input and it avoids the confusion that would ensue if positivity bias (rather than offset) referred to differences in the intercept, and negativity bias referred to differences in the slope of the transfer functions for positivity and negativity.

fest. Thus, the positivity offset reflects a difference in the transfer functions for positivity and negativity that would not be expected if positivity and negativity were simply the opposite ends of a single bipolar continuum.

The positivity offset may also bear on Zajonc's (1968) mere exposure effect, that is, the tendency for unfamiliar neutral stimuli to become liked as a function of repeated (i.e., unreinforced) exposure per se. One explanation for the mere exposure effect has been that repeated exposure lessens the response competition elicited by the unfamiliar stimulus. Yet this diminution of response competition invariably produces increased liking. The positivity offset is consistent with this resolution tending to end in increased liking rather than increased liking in some instances and increased disliking in others. In addition, a socially useful consequence of the positivity offset is that cohesion (i.e., positive attitudes) may be fostered among conspecifics who are simply exposed repeatedly to one another (Moreland & Zajonc, 1977; Swap, 1977).

Negativity Bias

The slope for the avoidance gradient was found to be steeper than the slope for the approach gradient in Miller's research on conflict behavior (see Figure 2). This difference illustrates a second difference in the transfer functions of positivity and negativity, the *negativity bias*: With each unit of activation, the change in negative motivational output is larger than the change in positive motivational output. Thus, the positivity offset can be conceived as relative activation of positivity even in the absence of any hedonic input, and the negativity bias refers to the finding that the increase in motivational output per quantum of activation tends to be greater for negativity than positivity.

Exploratory behavior can provide useful information about an organism's environment, but it can also place the organism in the proximity of hostile stimuli. Because it is more difficult to reverse the consequences of an injurious or fatal assault than an opportunity unpursued, the process of natural selection may have resulted in distinguishable motivational organizations for positivity and negativity with the propensity to react more strongly to negative than positive stimuli. The negativity bias—which underlies a stronger responsiveness to proximate negative than proximate positive or neutral events may therefore be a complementary, adaptive motivational organization. Species with both a positivity offset and a negativity bias enjoy the benefits of exploratory behavior as well as the self-preservative benefits of a predisposition to avoid or withdraw from threatening events.⁵

The positivity offset and negativity bias have been well documented in the person perception literature (e.g., Peeters, 1971; Peeters & Czapinski, 1990; Skowronski & Carlston, 1989). In addition, Taylor (1991) summarized a wide range of evidence showing that negative events in a context evoke stronger and more rapid physiological, cognitive, emotional, and social responses than neutral or positive events. Thus, there is considerable evidence for the differential transfer functions for the activation of positivity and negativity. In the next section, we incorporate these distinctions into the bivariate model of evaluative space.

Computational Model of Evaluative Space

Miller's (1959, 1961) theory concerns behavioral conflict originating from complex learning histories that included both positive and negative events (i.e., coactivation). Not surprisingly, therefore, it can be subsumed within the present conceptualization with the following revisions to the computational formula by which loci in evaluative space are mapped into loci on the attitudinal surface:

$$\text{Attitude} = Pf(i) + c - Ng(j) + I(ij) + e \quad (1)$$

where P and N are weighting coefficients; i represents the level of positivity activated by an attitude object; $f(i) = i^k$ (where $k < 1$) and is the activation function for positivity; c is a constant (equivalent to approximately 15% of the scale range for positivity) that contributes to a higher intercept for positivity than negativity (i.e., positivity offset); j represents the level of negativity activated by an attitude object; $g(j) = j^m$ (where $m < 1$) and is the activation function for negativity; $I(ij)$ represents nonadditive effects; and e represents an error term. Equation 1 is a first approximation, but inclusion of the interaction term indicates the assumption that the relationship between positive and negative evaluative processes is configural rather than strictly additive. For instance, as i and j increase (i.e., positivity and negativity are coactivated), pressures toward motivational disambiguity are assumed to generally work against yet additional coactivation. The negativity bias can be instantiated in one of two ways in Equation 1: (a) The weighting of negativity may be greater than the weighting of positivity (i.e., $P < N$) when forming an attitudinal judgment, or (b) the activation function for negativity itself may be steeper than the activation function for positivity (i.e., $k < m$).

The attitudinal surface depicted in Figure 5 was generated using the equation $.4(i)^5 + 1 + .6(j)^5$. The inset superimposes the attitudinal consequences of the activation of positivity and negativity as a function of

⁵Although we have focused here on general differences in the operation of positivity and negativity, the model also allows for

individual differences in parameters such as the strength of the positivity offset and negativity bias.

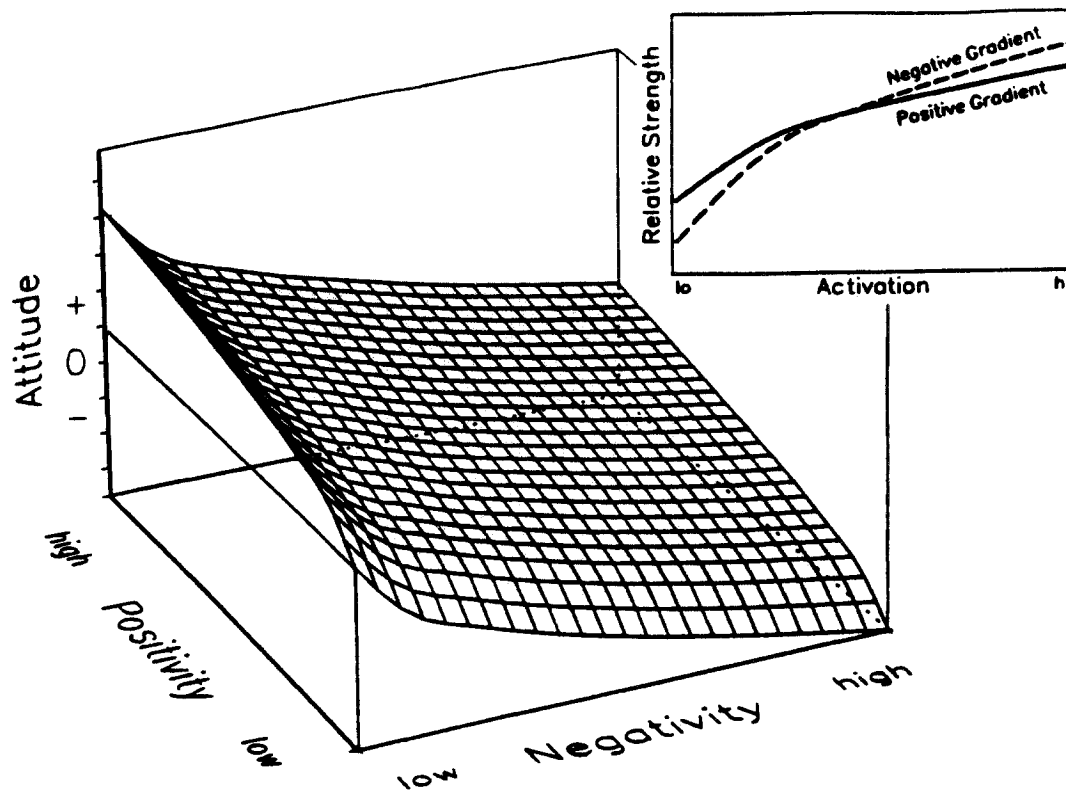


Figure 5. The bivariate evaluative plane and its associated attitudinal surface. The surface represents the attitudinal disposition of an individual toward (+) or away from (-) the target stimulus. The attitude is expressed in relative units. The point on the surface overlying the left axis intersection represents the maximally positive attitude evoked by the target stimulus, and the point on the surface overlying the right axis intersection represents the maximally negative attitude toward the target stimulus. The inset superimposes the strength of the activation of positive and negative valent forces as a function of movements along the coactivity diagonal. Note that the predictions depicted in this inset mirror those for approach-avoidance conflict in Miller's (1959) theory of conflict (see Figure 2). *Note.* From "Relationship Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of Positive and Negative Substrates," by J. T. Cacioppo and G. G. Berntson, 1994, *Psychological Bulletin*, 115, p. 415. Copyright © 1994 by the American Psychological Association. Reprinted with permission.

movements along the coactivity diagonal. Note that the predictions depicted in this inset mirror those for approach-avoidance conflict in Miller's (1959) theory of conflict.

Heteroscedacity of Antecedents

Tolstoy noted in *Anna Karenina* that all happy families are like one another, each unhappy family is unhappy in its own way. Tolstoy's observation may not be limited to families but rather may reveal yet another fundamental difference in positive and negative evaluative processes: Positive and negative hedonic processes may also be characterized by a heteroscedacity of antecedents (i.e., more conjunctions of factors that produce negativity than positivity) that is not found typically in truly bipolar concepts such as hot and cold. As an example, an individual may regard a piece of clothing negatively or as unfit for sundry reasons; it can be too long, too short, too narrow, too wide, unequal in length across the garment, and so on. If all these vari-

ables are in place, the garment fits; if any one or more of these variables is not appropriate, the garment does not fit. Thus, to say that a piece of clothing is unfit does not say much about the reason(s) it does not fit (i.e., the conjunction of factors responsible for the evaluation). To say that a piece of clothing fits, however, carries more information because there are fewer ways in which this outcome can be achieved (i.e., all dimensions have to match whereas only one needs to mismatch for it not to fit).

This does not mean that there are more antecedents for a specific negative reaction than for a specific positive reaction. Technological accidents, for instance, are usually the result of a conjunction of a number of unfortunate events—employees were not vigilant, technology failed, fail-safe systems failed, signals were misread, and so forth.⁶ Nevertheless, there may be many

⁶We thank David Messick for the point about technological accidents.

ways in which things can go wrong, but most if not all have to be proper to be right. There may be no more configurations of variables that elicit a smile than a frown or an expression of anger, fear, disgust, or contempt; if this is the case, there still would be more conjunctions of factors (antecedents) that arouse the negative than the positive motivational system. Moreover, there may be different action implications for the different negative outcomes, and these differing action implications may have fostered the evolution of multiple negative emotional expressions. Thus, it is possible that the existence of more negative than positive attributions of emotion to facial expressions is reflective of, if not derived from, this heteroscedacity and the associated differences in action implications (Ekman, 1972, 1994; Izard, 1994; Russell, 1994, 1995).

Asymmetries of Ambivalence

There are other reasons—related to the differences in the transfer functions of positivity and negativity—to conceptualize the processes underlying attitudes in terms of a bivariate evaluative space rather than in the traditional terms of a positive–negative bipolar continuum. For instance, despite the strong currents in the history of attitude theory and research that kept bipolar conceptualizations and measures in the mainstream, attitude neutrality, indifference, and ambivalence have long constituted troublesome eddies (e.g., A. I. Edwards, 1946; Kaplan, 1972; Scott, 1966).

In his review of attitude measurement and ambivalence, Scott (1968) presaged the present model of evaluative space:

The “directional” property raises, but does not automatically answer, the question as to how neutral attitudes are to be conceived. Are they midway between positive and negative attitudes? Should they be subdivided into indifferent and ambivalent attitudes? The conception of favorable and unfavorable as “opposites” implies that persons will not be found with attitudes simultaneously at both ends of the dimension. Yet an alternative formulation might treat degree of favorableness and degree of unfavorableness as conceptually distinct (although no doubt empirically correlated) components, on which persons may make, simultaneously, a variety of position combinations. In other words, it is only by convention that the direction of an attitude is conceptualized as a single bipolar attribute. (p. 206)

Various conceptualizations and measures of attitude ambivalence have been proposed since Scott’s (1968) observations (e.g., see Kaplan, 1972; Thompson, Zanna, & Griffin, 1995), but all predict that attitude ambivalence (a) increases as the distance from the

origin (zero evaluative activation) increases along the coactivity diagonal and (b) is symmetrically distributed about the coactivity diagonal. These similarities have made it difficult to distinguish empirically between the existing conceptualizations of attitude ambivalence. Although the evaluative space model shares the first feature with extant formulations of attitude ambivalence, the positivity offset and negativity bias lead to specific and unique predictions about asymmetries in attitude ambivalence.

One source of asymmetry stems from the constant, c , in Equation 1 (see Figure 6, left matrix). By virtue of this constant (i.e., the positivity offset), attitude ambivalence should be higher when only negativity is aroused than when only positivity is aroused by an attitude stimulus. When an attitude stimulus elicits only positivity, there is no negativity output and, therefore, no coactivation of affective outputs to arouse ambivalence. When an attitude stimulus elicits only negativity, however, there is still a weak level of positive output (by virtue of the positivity offset) and, consequently, attitude ambivalence can be aroused. Thus, attitude ambivalence should be more closely associated with the activation of negativity than positivity.

A second reason that attitude ambivalence is predicted to be asymmetrical about the coactivity diagonal is the differential weighting of positivity and negativity (i.e., the negativity bias; see Figure 6, right matrix). Ambivalence is predicted to increase as the distance from the origin (zero evaluative activation) increases along the coactivity diagonal. Because the increase in motivational output per unit of activation is predicted to be greater for negativity than positivity, however, ambivalence is greater on the negativity than positivity side of the coactivity diagonal. Thus, although the reason for the asymmetry is different, the nature of the asymmetry is similar: Attitude ambivalence should again be more closely associated with the activation of

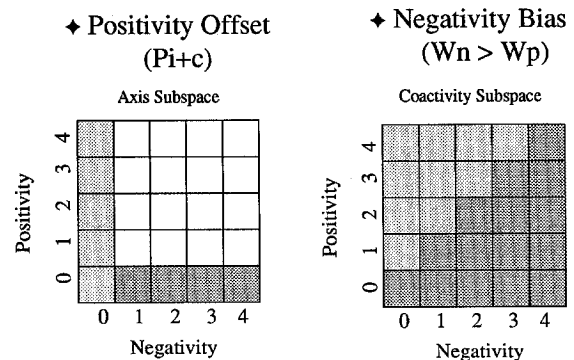


Figure 6. Subspaces within evaluative space highlighting the predicted asymmetries in ambivalence. Left panel: The positivity offset produces greater ambivalence when negativity alone is aroused than when positivity alone is aroused. Right panel: The negativity bias displaces feelings of maximal ambivalence during coactivation toward the negativity axis.

negativity than positivity. As discussed in the next section, our data on attitude ambivalence are consistent with this reasoning (see also Messick & Sentis, 1985).

Bivariate Evaluations and Ambivalence Measures (BEAMs)

We have investigated whether the transfer functions for positivity and negativity were the same or different in both traditional social psychological studies and in research using event-related brain potentials; we will discuss only traditional social psychological studies here. To investigate the model of evaluative space in traditional social psychological studies, we developed a set of unipolar self-report scales for measuring positivity, negativity, and ambivalence. We describe BEAMs in this section. In the next section, we review representative studies of the bivariate model using the BEAMs.

As noted, we initially used the PANAS to measure the strength of positivity and negativity. The PANAS, however, was designed to measure moods rather than the kind of rapid positive and negative valuation processes activated by attitude stimuli. In addition, the items on the Positivity and Negativity subscales of the PANAS are very different, varying in semantics and arousal as well as valence. Finally, the items on the PANAS may not be applicable to, or may underestimate the extremity of, attitudes that lack a strong emotional basis. Therefore, we developed the BEAMs for use in our studies of attitudes and evaluative space (Cacioppo, Crites, Snyder-Smith, & Gardner, 1996).

To identify potential items for our unipolar scales of Positivity and Negativity, we examined the items used in a number of extant scales (e.g., Osgood, Suci, & Tannenbaum, 1957; Russell, 1980; Watson et al., 1988). From these scales we selected a subset of positive and negative items that were equally applicable to any attitude object or event. Additional items were then identified by using a thesaurus to generate synonyms and antonyms of these initial items. Sixteen antonym pairs (e.g., *good–bad*, *positive–negative*) were selected for inclusion in the Positivity and Negativity unipolar scales, respectively. Two forms (A & B) of the Positivity and Negativity scales were created from these 16 antonym pairs to minimize the possibility that participants would treat these unipolar scales as bipolar scales (Table 2). Which positive items appeared on which form was determined randomly. To minimize carryover, the antonyms for the items used in the Positivity scale on Form A were used in the Negativity scale on Form B, and the antonyms for the items used in the Positivity scale on Form B were used in the Negativity scale on Form A (see Table 2).

Half of the participants completed Form A of the Positivity and Negativity scales and half completed

Table 2. *Bivariate Evaluations and Ambivalence Measures (BEAMs)*

Form A		Form B	
Positivity	Negativity	Positivity	Negativity
Favorable	Undesirable	Desirable	Unfavorable
Appealing	Negative	Positive	Unappealing
Pleasant	Unlikable	Likable	Unpleasant
Agreeable	Unhappy	Happy	Disagreeable
Approving	Opposing	Supporting	Disapproving
Rewarding	Bad	Good	Punishing
Delighted	Unattractive	Attractive	Distressed
Comfortable	Unsatisfying	Satisfying	Uncomfortable

Form B. Thus, although all participants rated attitude objects on equivalent scales, no participant rated an attitude object using an antonym pair. To further minimize carryover (see Thompson et al., 1995), participants completed all of the positivity (or negativity) ratings before they completed the negativity (or positivity) ratings, and administration of these scales was usually separated by a filler task. Participants indicated the extent to which each scale item reflected their reaction to each attitude object on a scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*). Not surprisingly, given the manner in which these scales were formulated, we have found Forms A and B to be statistically equivalent.

Items for a unipolar ambivalence scale were also generated. This resulted in six items that appeared to index inconsistent or contradictory information and would thus be descriptive of attitudes that included both positive and negative evaluative reactions (muddled, divided, tense, contradictory, jumbled, and conflicted), and three items—which were reverse scored—that appeared to index consistent information and would thus be descriptive of attitudes that consisted of evaluatively congruous information (consistent, uniform, and harmonious). As with the Positivity and Negativity scales, participants indicated the extent to which each item reflected their reaction to the attitude object on a scale ranging from 1 (*very slightly or not at all*) to 5 (*extremely*).

In the initial studies, participants were given positive and/or negative information about hypothetical roommates, coworkers, or jobs and used the BEAMs to rate their feelings toward these attitude objects. Factor analyses consistently yielded a three-factor solution, with negativity typically accounting for the most variance, followed by positivity, then ambivalence. Furthermore, the eight positive items used in Forms A and B were found to load highly on the same factor and did not load substantially on the other two factors; and the eight negative items used in Forms A and B were found to load highly on the same factor, albeit a different factor than the positive items, and did not load substan-

tially on the other two factors. Six of the nine ambivalence items were also characterized by good convergent and discriminant validity, but three ambivalence items (i.e., uniform, consistent, and harmonious) did not load highly on any factor. These three items, therefore, were dropped from the ambivalence scale. The finding that the resulting ambivalence scale was comprised of items that have a negative valence may reflect the strong tendency for ambivalence to be associated with conflict and tension—that is, dysphoric states. Furthermore, participants have shown little difficulty in grasping instructions on how to use the scales in Table 2, but we have found that extended instructions may be required for participants to understand what is being asked by the ambivalence items.

In a follow-up study, 107 male and female undergraduates used the BEAMs to rate their feelings about condoms. A factor analysis again suggested the existence of three internally consistent factors (e.g., see the scree plot in Figure 7). Forms A and B for measuring positivity and negativity were again found to be equivalent, so correlations among the total scale scores were calculated collapsing across forms. Results revealed a small negative correlation between the positivity and negativity aroused by condoms ($r = -.29, p < .05$) and a stronger correlation between negativity and ambivalence ($r = +.43, p < .05$) than between positivity and ambivalence ($r = -.08$). Together, these studies demonstrated that the BEAMs provided internally consistent and discriminable measures of the positive, negative, and ambivalent feelings aroused by hypothetical and real attitude stimuli.

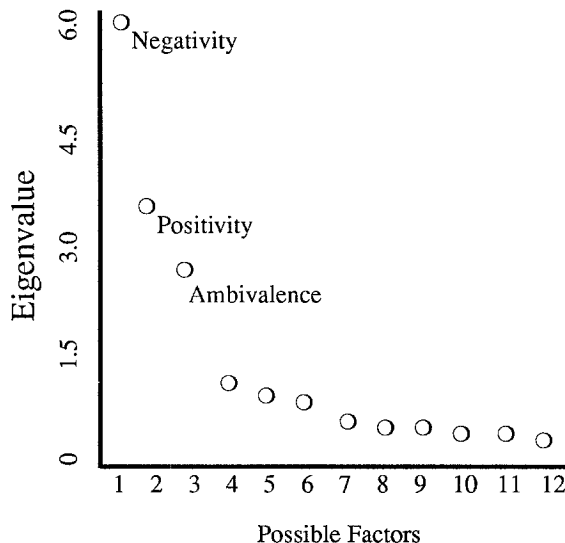


Figure 7. Scree plots from factor analyses of people's ratings of condoms using the BEAMs. A three-factor solution was obtained, with negativity items loading on the first factor, positivity items loading on the second factor, and ambivalence items loading on the third factor.

Further Tests of the Separability of Positivity and Negativity: The Positivity Offset and Negativity Bias in Attitude Formation

In the bipolar conceptualization, the positive and negative valuation processes underlying attitudes are treated as if reciprocally activated and interchangeable. The endorsement of positive beliefs about an attitude object is treated as equivalent to the rejection of negative beliefs about the object. Accordingly, the evaluative processes elicited by an object are thought to range along a single continuum—from maximally positive (and minimally negative) to maximally negative (and minimally positive)—with no distinctions made in the form of the activation functions for positivity and negativity. If attitudes are conceptualized as being bipolar—as being analogous to the position of a balance knob on an audio stereo, with very negative and very positive substituted for left speaker and right speaker—then the transfer functions for positivity and negativity should be equivalent.

As outlined in this article, however, the model of evaluative space holds that the transfer functions for the positive and negative evaluative processes underlying attitudes tend to differ in specific ways (e.g., positivity offset, negativity bias). To test this reasoning, we conducted a series of studies on attitude formation (Gardner, 1996). In an initial study, 80 undergraduate students served as participants in an experiment on impression formation. Participants read six neutral behaviors (e.g., "Sam weighed the vegetables before buying them") that were attributed to Sam and rated the extent to which each of the 9 positivity, 9 negativity, and 6 ambivalence items from the BEAMs reflected their evaluations of this person.

The results of these ratings, which are depicted in Figure 8, indicated that the neutral information about Sam aroused stronger feelings of positivity than nega-

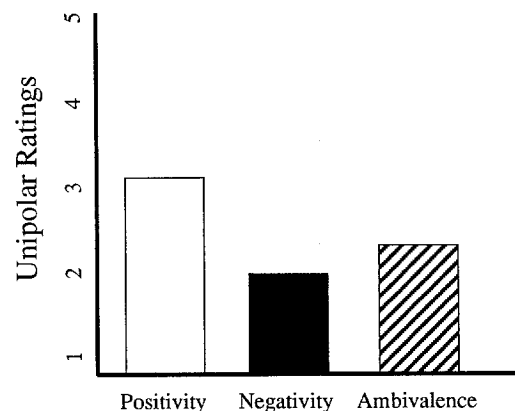


Figure 8. Mean ratings using the BEAMs of a hypothetical person, Sam, after exposure to neutral behavioral descriptions.

tivity or ambivalence. It is interesting that when another group of participants rated Sam before receiving any information about him, the positivity aroused by Sam was lower than when participants received the neutral and generally nondescript behavioral information. Thus, the results in Figure 8 are consistent with the operation of a positivity bias in evaluative activation rather than a response bias *per se*. Correlational analyses further confirmed that positivity and negativity were uncorrelated.

Participants in this study next read six additional behaviors attributed to Sam. Twenty participants read six additional neutral behaviors; 20 participants read six positive behaviors (e.g., "Sam gave vegetables from his garden to all of his neighbors"); 20 participants read six negative behaviors (e.g., "Sam stole vegetables from his neighbor's garden"); and 20 participants read six mixed behaviors (i.e., three positive and three negative, randomly ordered). The positive and negative behavioral descriptors that served as stimuli were matched in terms of extremity. Immediately afterwards, all participants completed the BEAMs to assess their evaluations of Sam.

The results of this second set of ratings, expressed as attitude change scores, are summarized in Figure 9. Several features of these results are noteworthy here. First, and as would be expected, participants who received yet additional neutral information about Sam showed only small additional attitude changes. Second, participants who read positive behavioral descriptions of Sam began to feel more positively about Sam, and participants who read negative behavioral descriptions began to feel more negatively about Sam. As inspection of Figure 9 further reveals, however, the negative infor-

mation had a much stronger impact on the ratings of Sam than did the positive information even though the individual positive and negative behavioral descriptors were matched on extremity. This result is, of course, consistent with the operation of a negativity bias. Finally, increases in negativity were associated with a corresponding reduction in positivity, suggesting some level of reciprocal activation in this impression formation task.

Results also suggested that there was an asymmetry in the effect of positive and negative information about Sam on ambivalence ratings. As predicted by the positivity offset, participants felt most ambivalent about Sam when the neutral information was followed by negative or mixed information and felt least ambivalent about Sam when the neutral information was followed by positive information.

One possible limitation for the evidence consistent with the proposed positivity offset is that participants were using neutral behavioral descriptors to rule out Sam's being a blatantly hideous individual. If Sam weighs vegetables before buying them, one might reason, he probably is not pathological and may even be conscientious and alert. To examine this issue, a replication and extension was conducted in which one group of participants read six neutral and nondiagnostic behaviors about Sam—that is, behaviors that all human beings perform (e.g., "Sam breathes oxygen"). Learning that Sam breathes oxygen or is subject to the laws of gravity provides no information that could be used to rule out the possibility that Sam is a blatantly hideous individual.

Ninety-five undergraduates served as participants in an experiment purportedly on artificial intelligence

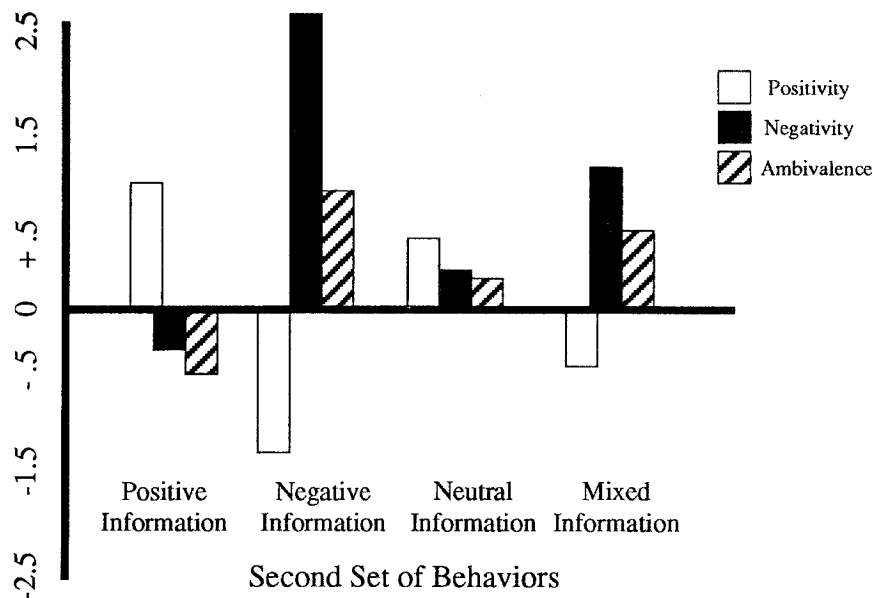


Figure 9. Mean ratings using the BEAMs of a hypothetical person, Sam, after exposure to six additional behavioral descriptors. Change in positivity, negativity, and ambivalence by type of information.

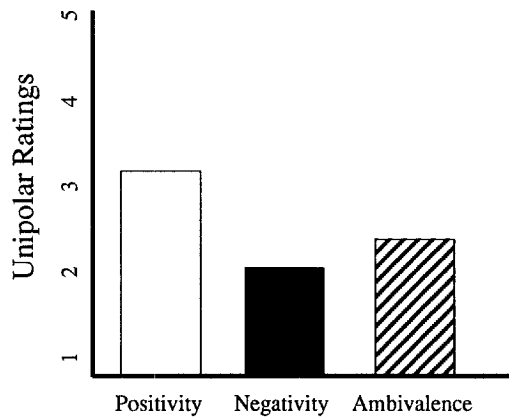


Figure 10. Mean ratings using the BEAMs of Sam after exposure to normal (standard) neutral behavioral descriptions.

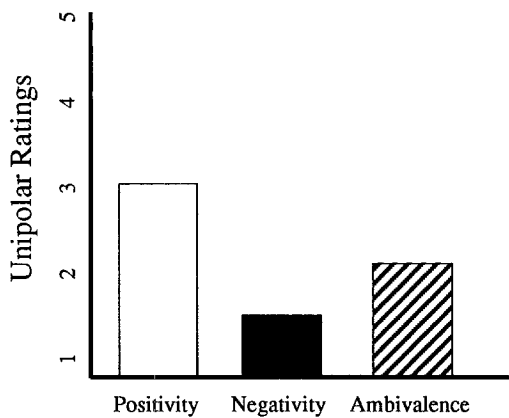


Figure 11. Mean ratings using the BEAMs of Sam after exposure to nondiagnostic neutral behavioral descriptions.

(Gardner, 1996). Participants were informed that the computer was programmed to simulate social intelligence and that their impressions, based on information about Sam accessed by the computer from its database, would serve as feedback to the computer. As in the prior studies, participants completed the BEAMs to assess their evaluations of Sam at two points in time. One group of participants began by reading the same six neutral behaviors as in our prior studies, whereas the remaining participants read six neutral and nondiagnostic behaviors about Sam—that is, behaviors that all human beings perform. Participants then rated Sam on the BEAMs.

As can be seen in Figures 10 and 11, the participants' ratings of Sam were similar for both types of neutral information: Participants were just as likely to show evidence of the positivity offset if they were exposed to completely nondiagnostic, neutral behavioral descriptions of Sam as when exposed to normal, neutral behavioral descriptions. Thus, the positivity offset does not appear to be the result of the diagnostic value of the neutral but normal behavioral descriptions. Correlational analyses again showed that the positive and negative evaluations of Sam were uncorrelated.

Participants were then exposed to either six additional normal neutral behaviors, six additional nondiagnostic neutral behaviors, six positive behaviors, six negative behaviors, or six mixed behaviors. Participants then completed the BEAMs to assess their evaluations of Sam (Gardner, 1996). The results are summarized in Figure 12. Not only were the results statistically equivalent for the normal and nondiagnostic neutral behav-

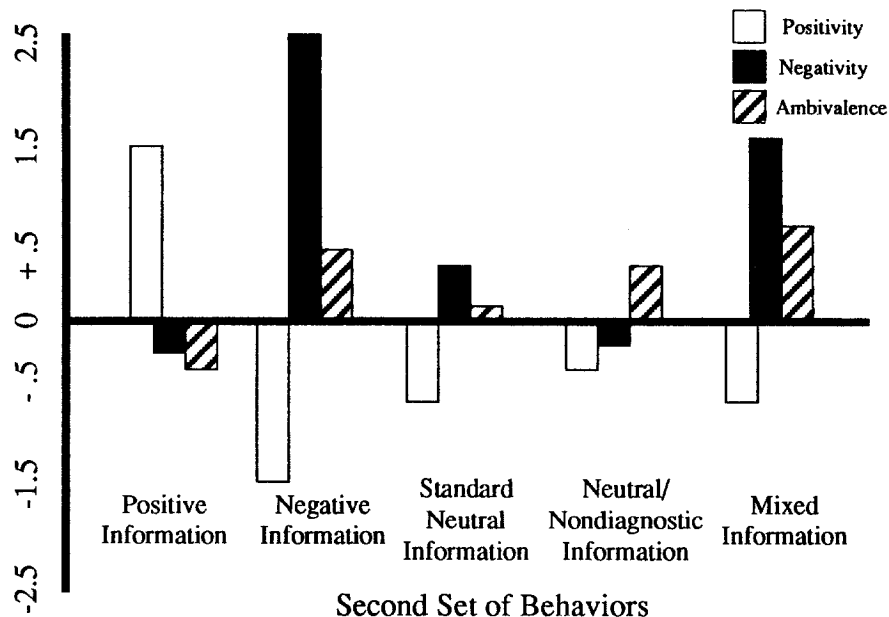


Figure 12. Mean ratings using the BEAMs of Sam after exposure to six additional behavioral descriptions. Change in positivity, negativity, and ambivalence by type of information.

ioral descriptions, but all of the major findings of the preceding study were replicated.

We next examined whether the results predicted by the positivity offset and the negativity bias were attributable to the fact that participants were judging another individual (Gardner, 1996). Forty-eight undergraduate students served as participants in a conceptual replication of the prior studies; however, the target of their impressions was the fictional aguaphore fish. As in the prior studies, participants completed the BEAMs to assess their feelings toward the fish at two points in time. All participants were first exposed to six neutral behaviors describing the fish (e.g., "The aguaphore rests at regular intervals"). After completing the BEAMs, participants were exposed to an additional six neutral, six positive (e.g., "The aguaphore will often

protect its mate and offspring"), six negative (e.g., "The aguaphore is useless commercially"), or six mixed behavioral descriptions of the fish and completed the BEAMs.

The evidence for the positivity offset was as clear for the hypothetical aguaphore fish as for Sam (see Figure 13), suggesting that the effect is not limited to judgments of people. Furthermore, correlational analyses again indicated that the positivity and negativity ratings were unrelated. The consistency of this result across cover stories, neutral descriptors, and attitude objects is notable. Indeed, we have observed the same effect for an attitude stimulus about which participants felt negatively initially—the imaginary six-legged entophore insect!

The changes in evaluations of and ambivalence toward the aguaphore fish also mirrored those observed in our studies of Sam (Gardner, 1996). As evident in Figure 14, participants who received yet additional neutral information about the aguaphore fish showed small additional attitude changes. Participants who received positive information about the aguaphore fish began to feel more positively, and participants who read negative behavioral descriptions began to feel more negatively and less positively about the aguaphore fish. More important, and as in the prior studies, the negative information had a much stronger impact on the ratings of the attitude stimulus than did the positive information even though the individual positive and negative behavioral descriptors were again matched on extremity. Indeed, the negative descriptors again had a stronger impact on all three rating dimensions (positivity, negativity, and ambivalence) than any other set of descriptors, and the increase in negativity was associated with

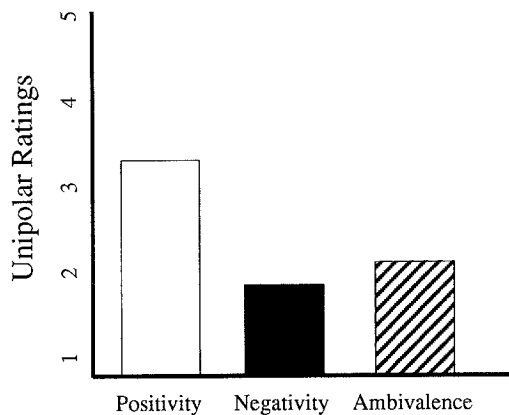


Figure 13. Mean ratings using the BEAMs of a hypothetical type of fish, the aguaphore, after exposure to standard neutral behavioral descriptions.

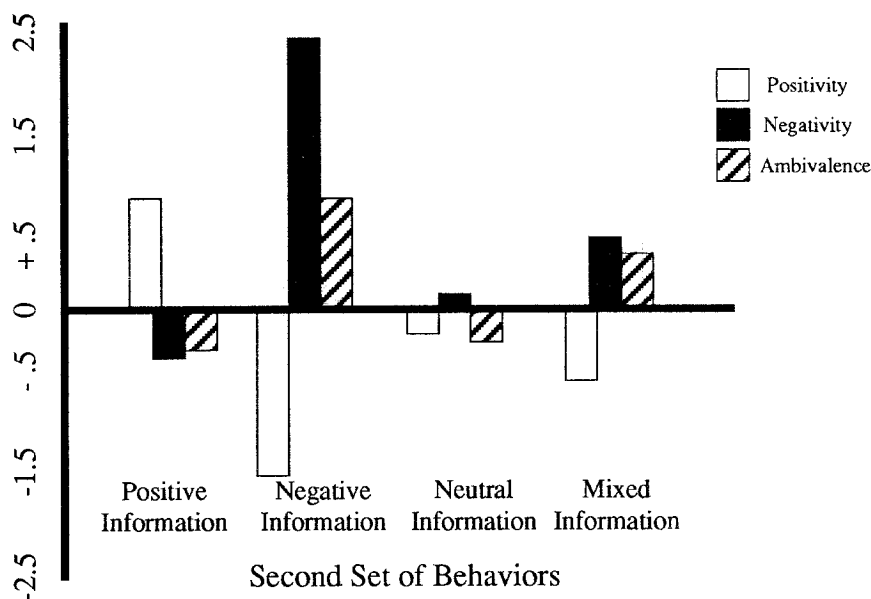


Figure 14. Mean ratings using the BEAMs of a hypothetical type of fish, the aguaphore, after exposure to six additional behavioral descriptors. Change in positivity, negativity, ambivalence.

a corresponding reduction of positivity (again suggesting some level of reciprocal evaluative activation in this impression formation task). Finally, the results were again consistent with an asymmetry in the effect of positive and negative information on ambivalence ratings. Participants felt most ambivalent about the aguaphore fish when the neutral information was followed by negative information and least ambivalent when the neutral information was followed by neutral or positive information.

In sum, these studies yielded evidence that the positive and negative evaluative processes underlying many attitudes are (a) distinguishable (stochastically and functionally independent), (b) characterized by distinct activation functions (positivity offset and negativity bias principles), and (c) related differentially to attitude ambivalence (corollary of ambivalence asymmetries). In the following section, we explore briefly some of the implications of these findings for political campaigns and voter behavior.

Positive and Negative Political Advertising: Equivalent and Interchangeable?

The notion of evaluative space originated in part from neuroscience research (Berntson, Cacioppo, & Quigley, 1991; Cacioppo & Berntson, 1992, 1994). Given this origin, it is interesting to speculate on the possible implications of the model for molar phenomena that have been of interest to social scientists. Consider recent voter behavior, in which the Republicans were voted out of the White House in 1992, the Democrats were voted out of the majority of Congress in 1994, and voters in 1994 were strikingly dour about the politicians they had just elected. Can the model of evaluative space explain any of these effects?

Campaign ads on television have revolutionized national elections. One of the pioneers of campaign ads on television was Democratic Senator William Benton of Connecticut (who was also the cofounder of a major ad agency). His 1950 television ads were characterized by action shots, patriotic music, spousal testimonials ("My husband has worked hard"), election promises ("I shall continue to fight for ... tax relief for small businesses"), and platitudes ("He's in there pitching for you"). What also characterized these ads (and the campaign ads that followed for decades) was their focus on the positive features of the candidate—what good the candidate could do if elected rather than what harm the candidate's opponent would do if elected. In the past decade, attack ads have become a staple.

The separability of positivity and negativity and the relative weight of negativity predicted by the evaluative space model may help explain the evolution of negative campaign ads as well as voter reactions following elections. Consider the predicted payoff matrix in an elec-

tion featuring two major candidates. If both candidates use positive advertising, the voters are exposed to the strengths and offerings of each candidate and, thus, are exposed to a choice between two positive alternatives. Although only one candidate can win, the electorate can be expected to feel relatively good—about their choices, the elected candidate, and the system of representative government—as both candidates are associated with approach motivational forces.

Extrapolating from our model, positive and negative campaign ads may not be equivalent and interchangeable in impact. Just as negative behavioral descriptions were found to have a stronger impact on participants' attitudes than equally extreme positive behavioral descriptions (Gardner, 1996), attack campaign ads may have a stronger effect on electoral evaluations and behavior than positive campaign ads, assuming all else is constant. Although speculative at this juncture, a candidate using negative campaign ads may have an advantage over an opponent using positive campaign ads. If this assumption holds true, it should not take long for political parties to learn that attack ads confer an advantage to their candidates and therefore adopt attack ads in close races. Thus, elections in which both major political parties use negative campaign ads might evolve fairly quickly once one begins using them.

The choices and perceptions of representative government, as well as the underlying motivations aroused in voters, are quite different when candidates feature negative advertising. Rather than voters learning strengths and offerings of each candidate, they learn of each candidate's failings, dishonesty, treachery, and infidelity. The voters, who must elect someone, are faced with a choice between two (or more) increasingly unattractive alternatives as they are bombarded by negative ads. Consequently, the electorate may feel disillusioned by their choices, the candidate they elect, and their system of representative government. The irony is that once negative campaign ads become a staple of the candidates' campaigns, the attack ads no longer provide a unique advantage to either candidate but only diminish the esteem, value, and effectiveness of the position to which the candidates aspire. It may not be so enigmatic, therefore, that the electorate has reacted as they have since the wide-scale introduction of attack ads (see also Marcus & Mackuen, 1993).

Conclusion

If the bivalent processes underlying some attitudes are not captured fully by bipolarity, the bivariate framework provides a better heuristic for representing these processes and for guiding research on their unique antecedents and consequences. The evaluative space model subsumes the bipolar model and captures any bivariate deviation from the more limited bipolar per-

spective. In light of the potential generative utility of a bivariate conceptualization of positive and negative evaluative processes, why might the notion of bipolarity have proven so durable? One reason is that the bipolar attitude scale has shaped data and theory in fundamental ways. Yet differences in the activation functions for positive and negative evaluative processes and in the modes of evaluative activation are difficult to detect using bipolar scales (J. D. Edwards & Ostrom, 1971). A second possible reason is that psychological and physiological processes in some conditions tend to constrain the full and enduring expression of the bivariate plane because they are incompatible or unstable. Examples include psychological states such as conflict, ambivalence, and inconsistency among beliefs about an attitude object—psychological states that tend to be unexpected, nonharmonious, and unstable (see Abelson et al., 1968; Festinger, 1957; Heider, 1946; Petty & Cacioppo, 1981).

In a classic experiment, Brehm (1956) demonstrated that choosing between two similarly attractive alternatives resulted in individuals exaggerating the positive features and discounting the negative features of the chosen alternative, while discounting the positive features and exaggerating the negative features of the unchosen alternative. In Brehm's study, the coactivation of positive and negative evaluative processes was evident but transient, and the participants' subsequent representations of the attitude stimuli were organized more along the lines of bipolarity than orthogonality.⁷ Thompson et al. (1995) have further demonstrated that ratings of positive and negative reactions to an attitude stimulus are more negatively correlated (i.e., organized in a bipolar fashion) when relevant questions are close together rather than separated within a questionnaire, presumably because placing the questions close together makes salient the behaviorally antagonistic motivational states that can be aroused by the stimulus and sets into motion motivational disambiguation (e.g., dissonance reduction). People's conceptual organization of affective states thus may tend toward a bipolar endpoint because of the operation of motives to maintain a simple and motivationally consistent representation of the world (principle of motivational certainty). Contemporary theory and research on cognitive dissonance emphasizes the importance of self-perception, self-consistency, and self-affirmation processes. Although contemporary research has demonstrated the importance of the self in cognitive dissonance, it remains the case that behaviorally ambiguous (e.g., coactivated positive and negative) motivational states may also be a determinant of cognitive dissonance (e.g., see Harmon-Jones, Brehm, Greenberg, Simon, & Nelson, 1996). As Har-

mon-Jones et al. (1996) suggested, the distal motivation to reduce cognitive dissonance may stem, at least in part, from the desire for predictability, direction, and stability. Psychologically speaking, the world is a simpler place when viewed in terms of dichotomies or bipolarities. This psychological endpoint is neither universal nor unambiguous regarding the hedonic processes that produced it, however.

Attitude ambivalence and cognitive consistency motives are familiar concepts to personality and social psychologists. The notions of the separability of positive and negative affective processes, the positivity offset, and the negativity bias can also be found in the extant literature. What, then, is different about the present formulation? First, these concepts are not treated as isolated empirical effects but rather are recognized as integral parts of a general evolutionary model of motivation and affect. Second, positive and negative motivational processes are not simply conceptualized as being distinguishable and negatively accelerating activation (i.e., marginal utility) functions are posited for each. Third, the prior literature has often cast the issue as whether positivity and negativity are bipolar or bivariate. In the evaluative space model, the more relevant question is under what conditions are positivity and negativity organized in a bipolar versus bivariate fashion. Fourth, we distinguish between (a) the early (primitive) motivational processes that differentiate between and marshal resources for dealing with hostile and hospitable stimuli, and (b) the conceptual organization of positive and negative beliefs and feelings about the stimuli. Furthermore, we suggest that the adaptive behavioral utility and, thus, the tendency toward bipolar organizations is greater for the latter than the former. Fifth, and relatedly, we posit the concept of modes of evaluative activation, which accommodates all possible combinations of positive and negative evaluative activation:

1. The reciprocal mode of evaluative activation is represented as one diagonal vector that ranges from maximal positivity–minimal negativity to maximal negativity–minimal positivity.
2. The nonreciprocal mode of evaluative activation is represented as the alternate diagonal that ranges from minimal positivity and negativity to maximal positivity and negativity.
3. The uncoupled modes of evaluative activation are represented as vectors lying along the axes.

Future research is needed to determine what conditions trigger these various modes of evaluative activation.

Sixth, the present conceptualization not only specifies that the behavioral effects of positivity and negativity are antagonistic, but posits a specific quantitative model and topographical representation for their effects. Seventh, the proposed model provides a quanti-

⁷It is these consequent conceptual organizations of affective and mood states that have been the focus of most of the research and debate to date (e.g., D. P. Green et al., 1993; Larson & Diener, 1992.)

tative heuristic for investigating individual differences in motivational orientation. Specifically, the parameters in the quantitative model specified in Equation 1 are estimates of central tendency. These values are expected to differ across individuals as a function of genetic and environmental factors.

In addition to these conceptual distinctions, several methodological issues are highlighted by the model. First, scientific instruments are shaped by a developer's implicit theory of the phenomena to be investigated, and once developed these scientific instruments shape data and theory in ways more fundamental than are often appreciated. Bipolar scales of hedonic processes and outcomes may be an important case in point. As outlined in this article, bipolar scales provide sensitive indices of reciprocal changes in positivity and negativity but are relatively insensitive to nonreciprocal changes (i.e., regions of indeterminism). Second, the notion that people find coactivated affective states (e.g., ambivalence) unpleasant, unexpected, and disharmonious and tend to avoid remaining in these states suggests that simply administering self-report unipolar scales may be insufficient to gauge accurately the initial levels of positivity and negativity that were activated by a stimulus (Thompson et al., 1995). If the unipolar scales are administered in close proximity to one another, for instance, one may garner more information on a psychological endpoint than the hedonic processes that produced this endpoint.

Finally, experimental conditions are sometimes established that allow one to contrast the effects of positive and negative states. For instance PET images of functional activation are generally based on paired image subtraction (Raichle, 1994). Two tasks (a control and a stimulated task) are developed that, theoretically, differ only in terms of the cognitive operation(s) of interest. The resulting images derived from the control task are subtracted from those obtained during the stimulated task to obtain an image that presumably reflects differences in activation between the two tasks. The resulting difference image is interpreted as revealing those areas of the brain concerned with the cognitive operation(s) unique to the stimulated task state (Fox, Mintun, Reiman, & Raichle, 1988). The general limitations of, caveats to, and benefits of the subtractive method in the analysis of physiological data have been discussed elsewhere (e.g., Cacioppo & Petty, 1986; Mintun, Fox, & Raichle, 1989; Posner, Petersen, Fox, & Raichle, 1988; Sarter et al., 1996). However, the subtraction of the images obtained during positive and negative states presumes these derive from the same neural substrates. If the activation of positivity versus negativity reflect the activity of distinguishable sets of neural substrates, stereogeometric functional brain imaging procedures such as PET will be blind to these distinctions unless each is contrasted with a neutral (i.e., no activation of positivity or negativity) state.

In sum, we began this article by noting the seductiveness of assuming that, because people's common-sense theories of social processes (e.g., the relationship between good and bad) can influence their social behavior, these intuitive theories also describe the essentials if not the entirety of the underlying psychological processes. We have suggested that people's intuitions about social psychological processes such as the bipolarity of good and bad may reflect a stable psychological endpoint but not the psychological means by which this state is achieved. One of the advantages of the bivariate evaluative space is that one can map the etiology of this psychological endpoint. In social psychology, as in other sciences, it may simply be that self-evident truths may not always be what they seem, and that common sense should not be awarded any privilege in theory and research on the governance of the social world.

References

- Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., & Tannenbaum, P. H. (1968). *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally.
- Abelson, R. P., Kinder, D. R., Peters, M. D., & Fiske, S. T. (1982). Affective and semantic components in political person perception. *Journal of Personality and Social Psychology*, 42, 619-630.
- Anderson, N. H. (1965). Primacy effects in personality impression formation using generalized order effects paradigm. *Journal of Personality and Social Psychology*, 2, 1-9.
- Bargh, J. A. (1994). The four horseman of automaticity: Varieties of automatic influence on social perception and cognition. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought: Causes and consequences for judgment, emotion, and behavior*. New York: Guilford.
- Berntson, G. G., Boysen, S. T., & Cacioppo, J. T. (1993). Neurobehavioral organization and the cardinal principle of evaluative bivalence. *Annals of the New York Academy of Sciences*, 702, 75-102.
- Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review*, 98, 459-487.
- Brehm, J. (1956). Post-decision changes in desirability of alternatives. *Journal of Abnormal and Social Psychology*, 52, 384-389.
- Brewer, M. B. (in press). In-group favoritism: The subtle side of intergroup discrimination. In D. Messick & A. Tenbrunsel (Eds.), *Behavioral research and business ethics*. New York: Russell Sage Foundation.
- Cacioppo, J. T., & Berntson, G. G. (1992). Social psychological contributions to the decade of the brain: The doctrine of multi-level analysis. *American Psychologist*, 47, 1019-1028.
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115, 401-423.
- Cacioppo, J. T., Crites, S. L., Jr., Berntson, G. G., & Coles, M. G. H. (1993). If attitudes affect how stimuli are processed, should not they affect the event-related brain potential? *Psychological Science*, 4, 108-112.
- Cacioppo, J. T., Crites, S. L., Jr., Snyder-Smith, M. A., & Gardner, W. L. (1996). *Bivariate evaluative and ambivalence measures (BEAMs)*. Manuscript in preparation.

- Cacioppo, J. T., & Gardner, W. L. (1993). What is underlying medical donor attitudes and behavior? *Health Psychology, 12*, 269–271.
- Cacioppo, J. T., & Petty, R. E. (1986). Social processes. In M. G. H. Coles, E. Donchin, & S. Porges (Eds.), *Psychophysiology: Systems, processes, and applications* (pp. 646–679). New York: Guilford.
- Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45*, 16–28.
- Crites, S. L., Jr., Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1995). Bioelectrical echoes from evaluative categorization: II. A late positive brain potential that varies as a function of attitude registration rather than attitude report. *Journal of Personality and Social Psychology, 68*, 997–1013.
- Delgado, J. M. R., Roberts, W. W., & Miller, N. E. (1954). Learning motivated by electrical stimulation of the brain. *American Journal of Physiology, 179*, 587–593.
- Diener, E., & Iran-Nejad, A. (1986). The relationship in experience between various types of affect. *Journal of Personality and Social Psychology, 50*, 1031–1038.
- Doob, L. W. (1947). The behavior of attitudes. *Psychological Review, 54*, 135–156.
- Edwards, A. I. (1946). A critique of “neutral” items in attitude scales constructed by the method of equal appearing intervals. *Psychological Review, 53*, 159–169.
- Edwards, J. D., & Ostrom, T. M. (1971). Cognitive structure and neutral attitudes. *Journal of Experimental Social Psychology, 7*, 36–47.
- Ekman, P. (1972). Universal and cultural differences in facial expressions of emotion. In J. K. Cole (Ed.), *Nebraska Symposium on Motivation, 1971* (pp. 207–283). Lincoln: University of Nebraska Press.
- Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell’s mistaken critique. *Psychological Bulletin, 115*, 286–187.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Fishbein, M., & Raven, B. H. (1962). The AB scales. *Human relations, 15*, 35–43.
- Fox, P. T., Mintun, M. A., Reiman, E. M., & Raichle, M. E. (1988). Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *Journal of Cerebral Blood Flow Metabolism, 8*, 642–653.
- Gardner, W. L. (1996). *Biases in impression formation: A demonstration of a bivariate model of evaluation*. Unpublished doctoral dissertation, Ohio State University, Columbus.
- Gardner, W. L., & Cacioppo, J. T. (1995). Multi-gallon blood donors: Why do they give? *Transfusion, 35*, 795–798.
- Gardner, W. L., & Cacioppo, J. T. (1996). *Attitudes toward blood donation: A demonstration of the bivariate nature of evaluation*. Manuscript under review.
- George, M. S., Ketter, T. A., Parekh, P. I., Horwitz, B., Herscovitch, P., & Post, R. M. (1995). Brain activity during transient sadness and happiness in healthy women. *American Journal of Psychiatry, 152*, 341–351.
- Gilbert, P. (1993). Defence and safety: Their function in social behaviour and psychopathology. *British Journal of Clinical Psychology, 32*, 131–153.
- Goldstein, M. D., & Strube, M. J. (1994). Independence revisited: The relation between positive and negative affect in a naturalistic setting. *Personality and Social Psychology Bulletin, 20*, 57–64.
- Gray, J. A. (1990). Brain systems that mediate both emotion and cognition. In J. A. Gray (Ed.), *Psychobiological aspects of relationships between emotion and cognition* (pp. 239–288). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Gray, J. A. (1991). Neural systems, emotion and personality. In J. Madden (Ed.), *Neurobiology of learning, emotion and affect* (pp. 273–306). New York: Raven.
- Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology, 64*, 1029–1041.
- Green, R. F., & Goldfried, M. R. (1965). On the bipolarity of semantic space. *Psychological Monographs, 79*(6, Whole No. 599).
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27.
- Harmon-Jones, E., Brehm, J. W., Greenberg, J., Simon, L., & Nelson, D. E. (1996). Evidence that the production of aversive consequences is not necessary to create cognitive dissonance. *Journal of Personality and Social Psychology, 70*, 1–12.
- Hass, R. G., Katz, I., Rizzo, N., Bailey, J., & Moore, L. (1992). When racial ambivalence evokes negative affect, using a disguised measure of mood. *Personality and Social Psychology Bulletin, 18*, 786–797.
- Heider, F. (1946). Attitudes and cognitive organization. *Journal of Psychology, 21*, 107–112.
- Izard, C. E. (1994). Innate and universal facial expressions: Evidence from developmental and cross-cultural research. *Psychological Bulletin, 115*, 288–299.
- Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin, 77*, 361–372.
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and American value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology, 55*, 893–905.
- Katz, I., Wackenhut, J., & Hass, R. G. (1986). Racial ambivalence, value duality and behavior. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 35–59). New York: Academic.
- Lang, P. J. (1995). The emotion probe. *American Psychologist, 50*, 372–385.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377–395.
- Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. *Review of Personality and Social Psychology, 13*, 25–59.
- Lazarsfeld, P. F. (1949). The American soldier—An expository review. *Public Opinion Quarterly, 13*, 377–404.
- LeDoux, J. E. (1995). In search of an emotional system in the brain: Leaping from fear to emotion and consciousness. In J. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1049–1062). Cambridge, MA: MIT Press.
- Lewin, K. (1935). *A dynamic theory of personality*. New York: McGraw-Hill.
- Marcus, G. E., & Mackuen, M. B. (1993). Anxiety, enthusiasm, and the vote: The emotional underpinnings of learning and involvement during presidential campaigns. *American Political Science Review, 87*, 672–685.
- Matthews, K., Wilkinson, L. S., & Robbins, T. W. (1996). Repeated maternal separation of preweanling rats attenuates behavioral responses to primary and conditioned incentives in adulthood. *Physiology and Behavior, 59*, 99–107.
- Messick, D. M., & Sentis, K. P. (1985). Estimating social and nonsocial utility functions from ordinal data. *European Journal of Social Psychology, 15*, 389–399.
- Miller, N. E. (1948). Theory and experiment relating psychoanalytic displacement to stimulus–response generalization. *Journal of*

- Abnormal and Social Psychology*, 43, 155–178.
- Miller, N. E. (1951). Comments on theoretical models illustrated by the development of a theory of conflict behavior. *Journal of Personality*, 20, 82–100.
- Miller, N. E. (1959). Liberalization of basic S–R concepts: Extensions to conflict behavior, motivation and social learning. In S. Koch (Eds.), *Psychology: A study of a science, Study I* (pp. 198–292). New York: McGraw-Hill.
- Miller, N. E. (1961). Some recent studies on conflict behavior and drugs. *American Psychologist*, 16, 12–24.
- Miller, N. E. (1966). Some animal experiments pertinent to the problem of combining psychotherapy with drug therapy. *Comprehensive Psychiatry*, 7, 1–12.
- Mintun, M. A., Fox, P. T., & Raichle, M. E. (1989). A highly accurate method of localizing regions of neuronal activation in the human brain with positron emission tomography. *Journal of Cerebral Blood Flow and Metabolism*, 9, 96–103.
- Moreland, R. L., & Zajonc, R. B. (1977). Is stimulus recognition a necessary condition for the occurrence of exposure effects? *Journal of Personality and Social Psychology*, 35, 191–199.
- Mummendey, A. (1995). Positive distinctiveness and social discrimination: An old couple living in divorce. *European Journal of Social Psychology*, 25, 657–670.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84, 231–259.
- Olds, J. (1958). Self-stimulation in the brain. *Science*, 127, 315–324.
- Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions. *Journal of Comparative and Physiological Psychology*, 47, 419–427.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Patchen, M., Davidson, J. D., Hofman, G., & Brown, W. R. (1977). Determinants of students' interracial behavior and opinion change. *Sociology of Education*, 50, 55–75.
- Patchen, M., Hofman, G., & Davidson, J. D. (1976). Interracial perceptions among high school students. *Sociometry*, 39, 341–354.
- Peeters, G. (1971). The positive–negative asymmetry: On cognitive consistency and positivity bias. *European Journal of Social Psychology*, 1, 455–474.
- Peeters, G., & Czapinski, J. (1990). Positive–negative asymmetry in evaluations: The distinction between affective and informational negativity effects. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 1, pp. 33–60). Chichester, England: Wiley.
- Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology*, 25, 57–75.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion*. Dubuque, IA: Brown.
- Posner, M. I., Petersen, S. E., Fox, P. T., & Raichle, M. E. (1988). Localization of cognitive operations in the human brain. *Science*, 240, 1627–1631.
- Raichle, M. E. (1994). Images of the mind: Studies with modern imaging techniques. *Annual Review of Psychology*, 45, 333–356.
- Robinson, R. G. (1995). Mapping brain activity associated with emotion. *American Journal of Psychiatry*, 152, 327–329.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141.
- Russell, J. A. (1995). Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin*, 118, 379–391.
- Russell, J. A., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of Personality and Social Psychology*, 57, 848–856.
- Sarter, M., Berntson, G. G., & Cacioppo, J. T. (1996). Brain imaging and cognitive neuroscience: Towards strong inference in attributing function to structure. *American Psychologist*, 51, 13–21.
- Scott, W. A. (1966). Brief report: Measures of cognitive structure. *Multivariate Behavior Research*, 1, 391–395.
- Scott, W. A. (1968). Attitude measurement. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 204–273). Reading, MA: Addison-Wesley.
- Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, 105, 131–142.
- Swap, W. C. (1977). Interpersonal attraction and repeated exposure to rewarders and punishers. *Personality and Social Psychology Bulletin*, 3, 248–251.
- Taylor, S. E. (1991). Asymmetrical effects of positive and negative events: The mobilization–minimization hypothesis. *Psychological Bulletin*, 110, 67–85.
- Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma & J. D. Maser (Eds.), *Anxiety and the anxiety disorders* (pp. 681–706). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Thompson, M. M., Zanna, M. P., & Griffin, D. W. (1995). Let's not be indifferent about (attitudinal) ambivalence. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 361–386). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33, 529–554.
- Thurstone, L. L. (1931). The measurement of attitudes. *Journal of Abnormal Psychology*, 26, 249–269.
- Watson, D. (1988). The vicissitudes of mood management: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of Personality and Social Psychology*, 55, 128–141.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–1070.
- Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement*, 9, 1–27.
- Zanna, M. P., & Rempel, J. K. (1988). A new look at an old concept. In D. Bar-Tal & A. W. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 315–334). Cambridge, England: Cambridge University Press.

