Journal of Personality and Social Psychology
1996, Vol. 71, No. 2, 364-374

0022-3514/96/$3.00

Mechanisms of Moral Disengagement in the Exercise of Moral Agency

Albert Bandura
Stanford University

Claudio Barbaranelli, Gian Vittorio Caprara, and Concetta Pastorelli
University of Rome, "La Sapienza"

This research examined the role of mechanisms of moral disengagement in the exercise of moral agency. Regulatory self-sanctions can be selectively disengaged from detrimental conduct by converting harmful acts to moral ones through linkage to worthy purposes, obscuring personal causal agency by diffusion and displacement of responsibility, misrepresenting or disregarding the injurious effects inflicted on others, and vilifying the recipients of maltreatment by blaming and dehumanizing them. The study examined the structure and impact of moral disengagement on detrimental conduct and the psychological processes through which it exerts its effects. Path analyses reveal that moral disengagement fosters detrimental conduct by reducing prosocialness and anticipatory self-censure and by promoting cognitive and affective reactions conducive to aggression. The structure of the paths of influence is very similar for interpersonal aggression and delinquent conduct. Although the various mechanisms of moral disengagement operate in concert, moral reconstruals of harmful conduct by linking it to worthy purposes and vilification of victims seem to contribute most heavily to engagement in detrimental activities.

Psychological theories of moral agency focus heavily on moral thought to the neglect of moral conduct. The limited attention to moral conduct reflects both the rationalistic bias of many theories of morality (Kohlberg, 1984) and the convenience of investigatory method. It is much easier to examine how people reason about hypothetical moral dilemmas than to study how they behave in difficult life predicaments. People suffer from the wrongs done to them, regardless of how perpetrators might justify their inhumane actions. The regulation of conduct involves much more than moral reasoning. A theory of morality must specify the mechanisms by which people come to live in accordance with moral standards. In social cognitive theory (Bandura, 1991), moral reasoning is translated into actions through self-regulatory mechanisms through which moral agency is exercised.

In the course of socialization, moral standards are constructed from information conveyed by direct tuition, evaluative social reactions to one's conduct, and exposure to the self-evaluative standards modeled by others. Once formed, such standards serve as guides and deterrents for action. People regulate their actions by the consequences they apply to themselves. They do things that give them satisfaction and a sense of self-worth. They refrain from behaving in ways that violate their moral standards, because such behavior will bring self-censure. In the face of situational inducements to behave in inhumane ways, people can choose to behave otherwise, by exerting counteracting self-influence. Anticipatory self-sanctions thus keep conduct in line with internal standards. It is through the ongoing exercise of self influence that moral conduct is motivated and regulated.

Social cognitive theory grounds moral agency in a self-regulatory system that operates through three major subfunctions. These include self-monitoring, judgmental, and self-reactive subfunctions. Self-monitoring of one's conduct is the first step toward exercising control over it. Action gives rise to self-reactions through a judgmental function in which conduct is evaluated against internal standards and situational circumstances. Moral judgment sets the occasion for self-reactive influence. People get themselves to behave in accordance with their moral standards through anticipatory positive and negative self-reactions for different courses of action.

Development of self-regulatory functions does not create an invariant control system within a person, as implied by theories of internalization that incorporate entities such as consciences, superegos, or moral principles as perpetual internal overseers of conduct. Self-reactive influences do not operate unless they are activated, and there are many psychosocial processes by which self-sanctions can be disengaged from inhumane conduct (Bandura, 1990, 1991). Selective activation and disengagement of internal control permits different types of conduct with the same moral standards. Figure 1 summarizes schematically the four major points in the self-regulatory system at which internal moral control can be disengaged from detrimental conduct. Self-sanctions can be disengaged by reconstruing the conduct, obscuring personal causal agency, misrepresenting or disregarding the injurious consequences of one's actions, and vilifying the recipients of maltreatment by blaming and devaluating them.

Albert Bandura, Department of Psychology, Stanford University; Claudio Barbaranelli, Gian Vittorio Caprara, and Concetta Pastorelli, University of Rome, "La Sapienza," Rome, Italy.

The research reported in this article was supported by grants from the Spencer Foundation to Albert Bandura and from the Johannn Jacobs Foundation to Gian Vittorio Caprara. We thank Delbert Elliott for his assistance in the early phases of the development of the scale of moral disengagement.

Correspondence concerning this article should be addressed to Albert Bandura, Department of Psychology, Stanford University, Stanford, California 94305-2130; or to Gian Vittorio Caprara, Dipartimento di Psicologia, Universita Degli Studi di Roma, "La Sapienza," Via dei Marsi, 78, 00185 Roma, Italy.

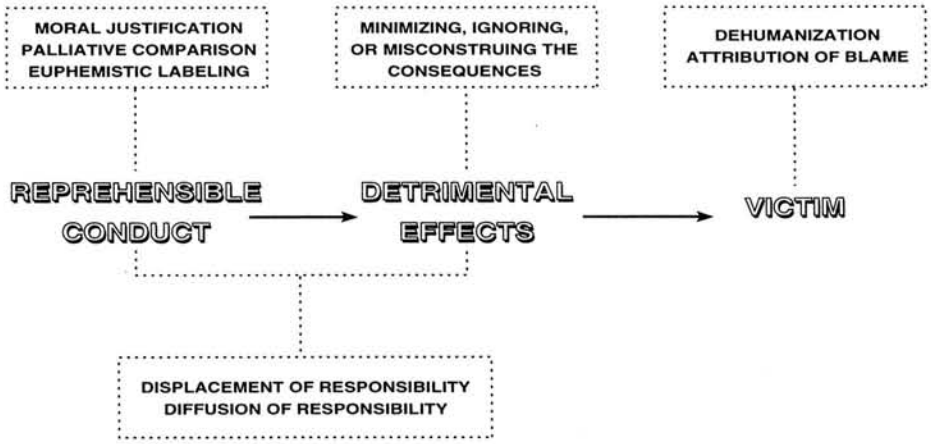

Figure 1. Mechanism through which moral self-sanctions are selectively activated and disengaged from detrimental behavior at different points in the self-regulatory process. From *Social Foundations of Thought and Action: A Social Cognitive Theory* (p. 376) by A. Bandura, 1986. Copyright 1986 by Prentice-Hall, Inc. Reprinted by permission of Prentice-Hall, Inc., Upper Saddle River, NJ.

A key set of disengagement practices operates on the construal of injurious behavior itself. People do not ordinarily engage in reprehensible conduct until they have justified to themselves the rightness of their actions. What is culpable can be made righteous through cognitive reconstrual. In this process of *moral justification*, detrimental conduct is made personally and socially acceptable by portraying it in the service of valued social or moral purposes (Kelman & Hamilton, 1989; Kramer, 1990; Sanford & Comstock, 1971). People then act on a social or moral imperative. In the transactions of everyday life, a lot of aggressive behavior gets justified in the name of protecting honor and reputation (Cohen & Nisbett, 1994).

Language shapes people's thought patterns on which they base many of their actions. Activities can take on markedly different appearances depending on what they are called. *Euphemistic language* thus provides a convenient tool for masking reprehensible activities or even conferring a respectable status upon them (Bolinger, 1982; Lutz, 1987). Through sanitized and convoluted verbiage, destructive conduct is made benign and those who engage in it are relieved of a sense of personal agency. Laboratory studies have revealed the disinhibitory power of euphemistic language (Diener, Dineen, Endresen, Beaman, & Fraser, 1975). People behave much more aggressively when assaulting a person is given a sanitized label than when it is called aggression.

Behavior can also assume very different qualities depending on what it is contrasted with. By exploiting *advantageous comparison* with more reprehensible activities, injurious conduct can be rendered benign or made to appear to be of little consequence. The more flagrant the contrasted activities, the more likely it is that one's own injurious conduct will appear trifling or even benevolent (Bandura, 1991). Cognitive transformation of harmful conduct into good conduct through moral justifications and palliative characterizations by euphemistic labeling and behavioral contrasts is the most effective psychological mechanism for disengagement of self-sanctions. This is because investing injurious means with high social or moral purpose not only eliminates self-deterrents but also engages self-approval in the service of harmful exploits. What was once morally censurable becomes a source of positive self-valuation.

Self-sanctions are activated most strongly when personal agency for detrimental effects is acknowledged. The second set of dissociative practices operates by obscuring or distorting the agentive relationship between actions and the effects they cause. Under *displacement of responsibility*, people view their actions as springing from the social pressures or dictates of others rather than as something for which they are personally responsible (Andrus, 1969). Because they are not the actual agents of their actions, they are spared self-censuring reactions. Hence, they are willing to behave in ways they normally repudiate if a legitimate authority accepts responsibility for the effects of their actions (Diener, 1977; Milgram, 1974).

The exercise of moral control is also weakened when personal agency is obscured by *diffusion of responsibility* for detrimental conduct. This is achieved in several ways. Responsibility can be diffused by division of labor for a venture with different members performing subdivided aspects that seem harmless in themselves but harmful in its totality (Kelman, 1973). Group decision making is another common practice, one that enables otherwise considerate people to behave inhumanely. When everyone is responsible, no one really feels responsible. Group action is still another expedient for weakening moral control. Any harm done by a group can always be attributed largely to the behavior of others. People behave more cruelly under group responsibility than when they hold themselves personally accountable for their actions (Bandura, Underwood, & Fromson, 1975; Diener, 1977; Zimbardo, 1969, 1995).

Additional ways of weakening self-deterring reactions operate by *disregarding or distorting the consequences* of action. When people pursue activities harmful to others for personal gain, or because of social inducements, they avoid facing the harm they cause, or they minimize it. They readily recall prior information given them about the potential benefits of the behavior but are less able to remember its harmful effects (Brock & Buss,

1962, 1964). In addition to selective inattention and cognitive distortion of effects, the misrepresentation may involve active efforts to discredit evidence of the harm they cause. As long as the detrimental results of one's conduct are ignored, minimized, distorted, or disbelieved, there is little reason for self-censure to be activated. In his studies of commanded aggression, Milgram (1974) obtained diminishing obedience as victims' pain became more evident and personalized.

The final set of disengagement practices operates on the recipients of detrimental acts. The strength of moral self-sanctions depends partly on how perpetrators view the people they mistreat. To perceive another as human activates empathetic and vicarious emotional reactions through perceived similarity (Bandura, 1992; McHugo, Smith, & Lanzetta, 1982). The joys and suffering of those with whom one identifies are more vicariously arousing than are those of strangers, out-group members, or those who have been divested of human qualities. It is, therefore, difficult to mistreat humanized persons without risking personal distress and self-censure. Subhumans are regarded as not only lacking sensitivities but also as being influenceable only by harsh means. In research on the dynamics of victimization, Perry, Williard, and Perry (1990) reported that aggressive children exhibit little sympathetic concern over hurting devalued peers. Habitual aggressors care less about inflicting suffering on victimized classmates than on those who are not cast in the devalued-victim role.

Self-censure for injurious conduct can be disengaged or blunted by *dehumanization* that divests people of human qualities or attributes bestial qualities to them. Once dehumanized, they are no longer viewed as persons with feelings, hopes, and concerns but as subhuman objects (Haritos-Fatouros, 1988; Keen, 1986; Kelman, 1973). In experimental studies in which otherwise considerate people are given punitive power, they treat dehumanized individuals much more harshly than humanized ones (Bandura et al., 1975). Dehumanization fosters different patterns of thought. People enlist moral justifications for punitive conduct directed toward individuals who have been deprived of humanness, but they disavow punitive actions and condemn them on moral grounds toward individuals depicted in humanized terms.

Blaming one's adversaries or circumstances is still another expedient that can serve self-exonerative purposes. In moral disengagement by *attribution of blame*, people view themselves as faultless victims driven to injurious conduct by forcible provocation. Punitive conduct thus becomes a justifiable defensive reaction to instigations. Victims get blamed for bringing suffering on themselves (Ferguson & Rule, 1983). Self-exoneration is also achievable by viewing one's harmful conduct as forced by compelling circumstances rather than as a personal decision. Even very young children are quite skilled in using mitigating factors to excuse harmdoing (Darley, Klosson, & Zanna, 1978). Children who are ready aggressors are quick to ascribe hostile intent to others, which provides justification for preemptive retaliatory acts (Crick & Dodge, 1994). By fixing the blame on others or on circumstances, not only are one's own injurious actions excusable but one can even feel self-righteous in the process.

The disinhibitory effects of the various forms of moral disengagement have been extensively documented in the perpetration of large-scale inhumanities (Andrus, 1969; Keen, 1986; Kelman & Hamilton, 1989; Rapoport & Alexander, 1982; Reich, 1990). Laboratory studies have further verified that conditions conducive to disengagement of moral self-sanctions heighten punitive behavior (Bandura et al., 1975; Diener, 1977; Diener et al., 1975; Milgram, 1974; Tilker, 1970; Zimbardo, 1969). However, because of the lack of measures of moral disengagement, the mediation of the effect of the manipulated conditions on punitive behavior through self-exonerative processes has been presumed rather than assessed.

The present research addressed several key issues concerning the exercise of moral agency. Research in this area has been seriously hampered by the lack of measures of moral disengagement. The instrument devised and tested in this program of research is grounded in a sociocognitive theory of moral agency that specifies the different loci in the self-regulatory system where moral self-sanctions can be effectively disengaged. Both the naturalistic and laboratory investigations usually examine only a single or a subset of disengagement mechanisms. The present study investigated how the full set of moral disengagement mechanisms operate in concert on socially injurious and antisocial conduct under naturally occurring conditions.

Moral disenagement can affect detrimental behavior both directly and by its impact on other theoretically relevant determinants. Therefore, this research also tested a conceptual model of the paths of influence through which moral disengagement produces its behavioral effects. The directional paths are specified both by theory (Bandura, 1991) and by empirical tests of particular links in the model (Caprara & Pastorelli, 1993). In the proposed causal structure of the model, which is presented in Figure 2, moral disengagement influences detrimental behavior both directly and through its effects on the following mediating factors. People have little reason to be troubled by guilt or to feel any need to make amends for inhumane conduct if they reconstrue it as serving worthy purposes or if they disown personal agency for it. We therefore predicted that high moral disengagement would be accompanied by low guilt, thus weakening anticipatory self-restraints against engagement in detrimental behavior. We further predicted that self-exoneration for harmful conduct and self-protective dehumanization of others and treating them as blameworthy would spawn a low prosocial orientation. Low prosocialness would, in turn, contribute to detrimental conduct in two ways. Having little sympathy for others would both remove the restraining influence of empathetic considerateness to the mistreatment of others and would activate little anticipatory guilt over such behavior. Effective moral disengagement creates a sense of social rectitude and self-righteousness that breeds ruminative hostility and retaliatory thoughts for perceived grievances. People often ruminate hostilely but do not act on their feelings. However, freed from the restraint of moral self-sanctions, they are more likely to act out their resentments. In this mediated link, moral disengagement fosters aggression proneness indexed by irascibility and hostile rumination which, in turn, heighten the likelihood of aggressive and transgressive behavior. Thus, in the sociocognitive conceptual model, moral disengagement affects aggressive and transgressive conduct both directly and through its influence on anticipatory guilt reactions, prosocial orientation, and cognitive and affective reactions that are conducive to aggression.

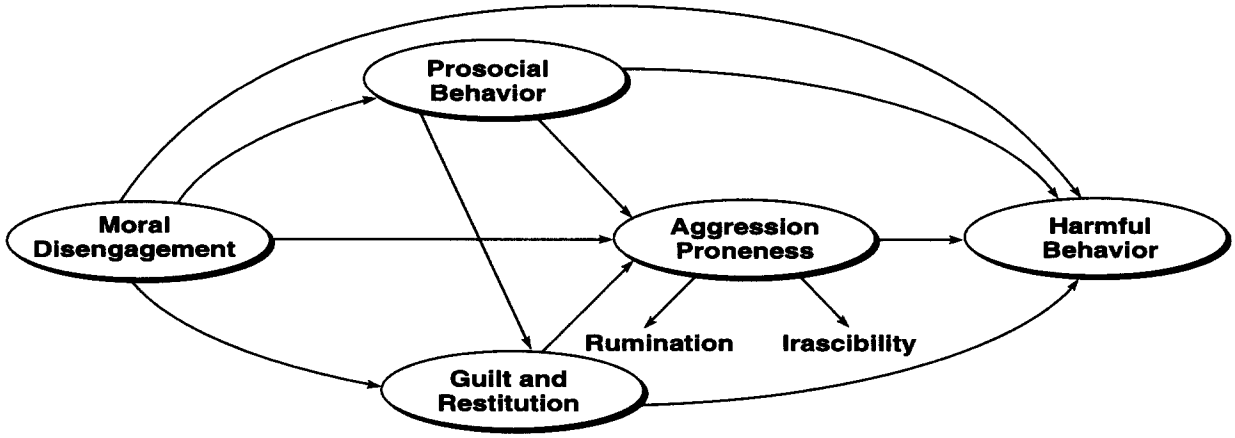

Figure 2. Proposed causal structure of the paths of influence through which moral disengagement affects detrimental conduct.

Method

Participants

The participants in this study were 124 children in the last year of elementary school and 675 junior high school students in Grades 6–8. They ranged in age from 10 to 15 years with a mean age of 11.8 years. There were 438 males and 361 females.

The students were drawn from four public schools: two elementary and two junior high schools in a residential community located near Rome, Italy. This community adheres to a stringent consent procedure for the conduct of research in the schools. A research proposal must gain approval from a school council composed of parent and teacher representatives and student representatives, as well, at the junior high school level. In addition, parents must give consent, and children are free to decline to take part if they choose. The parents not only consented to the study, but the mothers participated in the project themselves. All of the children enrolled in these grades participated in the study unless they happened to be absent from school when the measures were administered. The study was described to the parents and children as a project conducted through the University of Rome to gain better understanding of how children develop.

This community represents a socioeconomic microcosm of the larger society, containing families of skilled workers, farmers, professionals, and local merchants and their service staffs. Socioeconomic status of the family was assessed by father's occupation. Fourteen percent were in professional or managerial ranks, 25% were merchants or operators of other businesses, 31% were skilled workers, 29% were unskilled workers, and 1% were retired. The socioeconomic heterogeneity of the sample adds to the generalizability of the findings.

Children were administered the sets of scales measuring the variables of theoretical interest in their classrooms by two female experimenters. The various measures were administered over a period of several days. To add to the significance of obtained relationships, data for the variables of interest were collected by different methods from parents, teachers, and peers, as well as from the children themselves. The scales were administered individually to the teachers and parents.

Moral Disengagement

Extensive prior psychometric analyses were conducted in the development of the measure of moral disengagement. A large pool of items tapping the different disengagement mechanisms was constructed on the basis of the guiding conceptual scheme. They were then pretested on 251 elementary school children, 249 junior high students, and 315 high school students. Items were rewritten to remove ambiguities and eliminated if they were not internally consistent within each of the mechanisms. This pilot project revealed that proneness to moral disengagement was positively related to aggressive behavior and negatively related to prosocial behavior across the three age groups regardless of whether the children's behavior was measured by self-ratings, teacher ratings, or sociometric peer ratings.

The final form of the multifaceted scale used in the present study assessed proneness to moral disengagement of different forms of detrimental conduct in diverse contexts and interpersonal relationships (Bandura, 1995). Each of the eight mechanisms of moral disengagement was represented by a subset of 4 items. The full set of 32 items is presented in the Appendix. The items tapped children's readiness to resort to moral justification, euphemistic labeling, advantageous comparison, displacement and diffusion of responsibility, distortion of consequences, dehumanization, and attribution of blame for different forms of transgressive conduct. The transgressive activities involved physically injurious and destructive conduct, verbal abuse, deceptions, and thefts. The social contexts encompassed educational, familial, community, and peer relations. For each of the items, children rated on a 3-point Likert-type scale their degree of acceptance of moral exonerations for such conduct on an agree–disagree continuum. A principal-components factor analysis with varimax orthogonal rotation revealed a single factor structure. It accounted for 16.2% of the variance. Because no subfactors emerged, we summed the responses to the set of items to provide the composite measure of moral disengagement. The alpha reliability coefficient for this measure is .82.

Aggressive and Prosocial Behavior

We obtained data regarding children's prosocial, aggressive, and transgressive behavior from different sources, using diverse methods of assessment. The sources included the children themselves, their parents, teachers, and peers. The methods of measurement included personality questionnaires and peer sociometric ratings. To avoid possible response biases, several control items were included in each of the questionnaires.

The children completed two scales developed by Caprara and his colleagues to measure prosocial behavior and interpersonal aggression (Caprara, Barbaranelli, Pastorelli, & Perugini, 1994). A 3-point re-

sponse format was used throughout. The measure of *physical and verbal aggression,* containing 15 items, assessed the frequency with which children fought with others, hurt them, and verbally disparaged them. *Prosocial behavior* was assessed by 10 items in terms of helpfulness, sharing, kindness, and cooperativeness.

Teachers rated the children in their classes for physical and verbal aggression and prosocial behavior using the scales administered to the children but shortened to six items each and cast in a third-person format. The mothers also rated the frequency with which their children exhibited prosocial and aggressive forms of behavior, using the same set of scales as administered to the children. We computed internal consistency reliabilities using Cronbach's alpha. For the 16 sets of scores, except for an alpha of .61 for peer ratings of prosocial behavior, the reliability coefficients for the four sources of data (self, parents, teachers, peers) across the two educational levels for aggressiveness and prosocialness were virtually all in the .80s and .90s. The concurrent validity of these measures has been corroborated in studies relating children's ratings of their behavior to level of prosocialness and aggressiveness as rated by parents and teachers and by peers' sociometric nominations (Caprara & Pastorelli, 1993).

Sociometric peer nominations served as another source of assessment of prosocial and aggressive behavior. Children were presented with a booklet containing the names of children in their class along with 10 items, 3 of which measured aggressive and prosocial behavior and 4 of which measured peer popularity and rejection. Specifically, in the aggression domain peers circled the names of three classmates who fight a lot, insult other children, and often hurt them. For prosocial behavior, the peers circled the names of three classmates who help others, share things, and try to make sad people happier.

We measured *peer popularity* by having peers select the three classmates with whom they would like to play and to study. To distinguish between children who were disliked by their peers and those who were simply ignored or regarded indifferently, peers selected three classmates with whom they would neither want to play nor study as a measure of *peer rejection*. The two aspects of peer popularity were positively correlated ($r = .68$, $p < .001$), as were the two aspects of peer rejection ($r = .83$, $p < .001$). Popularity and rejection were negatively related ($r = -.42$, $p < .001$). The multidimensional assessment of both positive and negative status regarding both social and academic activities provided a good basis for gauging the impact of moral disengagement on quality of peer relations.

Affective and Cognitive Aspects of Aggression

Students at the junior high school level were administered two additional scales that measured the affective and cognitive aspects of aggressive and transgressive conduct for matters especially relevant for older children. The *hostile rumination* measure assessed with 15 items the level of preoccupation with personal grievances and retaliatory action. The *irascibility* measure, comprising 14 items, tapped petulance in social transactions and weak restraint over anger even to slight provocations. The alpha reliability coefficients were .86 for hostile rumination and .84 for irascibility. The predictive validity of these measures has been corroborated experimentally under simulated conditions in which participants can inflict shocks of varying intensity on a provocateur. Individuals who have a low threshold for anger arousal and are prone to hostile rumination behave more punitively than those who are slower to anger and disinclined to dwell on grievances and possible retaliations (Caprara, Coluzzi, Mazzotti, Renzi, & Zelli, 1985; Caprara, Renzi, Alcini, D'Imperio, & Travaglia, 1983; Caprara, Renzi, Amolini, D'Imperio, & Travaglia, 1984; Caprara et al., 1986).

Self-Sanctions for Transgressive Behavior

The scale measuring *guilt and restitution,* which included 15 items, dealt with the self-regulation of transgressive conduct by anticipatory self-sanctions. It assessed the degree of guilt, remorsefulness, and self-criticism anticipated for transgressive conduct and the need to make restitution if it were carried out. Factor analysis of the items revealed a single factor. The alpha reliability coefficient for this scale was .79.

Delinquent Behavior

Delinquent behavior was measured by the relevant items from the Child Behavior Checklist developed by Achenbach and Edelbrock (1978). Both the reliability and predictive validity of this measure of problem behavior are well established (Achenbach, McConaughy, & Howell, 1987). The Delinquency subscale, comprising 22 items for males and 19 items for females, covers a wide range of transgressive behaviors, including theft, cheating, lying, destructiveness, truancy, and use of alcohol and drugs. The Parental scale includes 12 items, which overlap with the children's version. Both the mothers and the children themselves recorded whether they engage in such antisocial activities and, if they do, whether they do so only occasionally or often. The reliability coefficients were .77 for parents, .77 for females, and .85 for males.

Results

Children's proneness to moral disengagement was unrelated to familial socioeconomic status, and it did not differ as a function of age. However, males exhibited higher moral disengagement than did females, $F = 22.17$, $p < .0001$. The major sources of this difference were the males' greater readiness to provide moral justifications for detrimental conduct, $F = 45.81$, $p < .0001$; to mask it in euphemistic language, $F = 33.81$, $p < .0001$; to minimize its injurious effects, $F = 6.14$, $p < .025$; and to dehumanize victims, $F = 26.60$, $p < .0001$, and attribute blame to them, $F = 9.92$, $p < .002$. The degrees of freedom for these analyses are $df = 1, 789$.

Although the analyses indicate that the various mechanisms of moral disengagement operate in concert in the self-regulatory process, they varied somewhat in degree of enlistment. Construing injurious behavior as serving righteous purposes, disowning responsibility for harmful effects, and devaluing those who are maltreated were the most widely used modes of exonerative disengagement of self-sanctions. Masquerading censurable activities in palliative language or rendering them benign by advantageous comparison, both of which require dexterous cognitive skills, were used less often.

Pattern of Relationships

Table 1 presents the relationships between moral disengagement and prosocial and detrimental conduct. The correlations are highly consistent across different sources of data, and the correlates did not differ significantly on any of the measures across the two educational levels. Compared to individuals who maintain a high level of moral agency, those who are highly prone to moral disengagement tend to be more irascible, ruminate about perceived grievances, and are neither much troubled by guilt nor feel the need to make amends for harmful conduct. They also engage in a higher level of interpersonal aggression and delinquent behavior.

Moral disengagement is related to prosocial behavior as well as to transgressive activities. High moral disengagers are less prosocially oriented and more likely to be rejected by peers.

Table 1
Relationship of Moral Disengagement to Prosocial, Aggressive, and Delinquent Behavior

Social and transgressive behavior	Moral disengagement	
	Elementary	Junior high
Self-ratings		
Physical and verbal aggression	.52***	.36***
Prosocial behavior	−.40***	−.28***
Irascibility	—	.36***
Hostile rumination	—	.38***
Guilt and restitution	—	−.18***
Delinquency	—	.45***
Teacher ratings		
Physical and verbal aggression	.22**	.06***
Prosocial behavior	−.30***	−.18***
Peer ratings		
Physical and verbal aggression	.46***	.29***
Prosocial behavior	−.16*	−.17***
Popularity	−.04	−.07*
Rejection	.23**	.22***
Parent ratings		
Physical and verbal aggression	.38***	.29***
Prosocial behavior	−.17*	−.11***
Delinquency	.20*	.27***

Note. Dashes indicate data that were not collected for this sample of children.
*$p < .05$. **$p < .01$. ***$p < .001$.

However, there is no consistent relationship between moral disengagement and peer popularity. One can be rejected by prosocial peers and gravitate toward, and gain acceptance from, dissocial or deviant peers (Cairns, Cairns, Neckerman, Gest, & Gariépy, 1988; Dishion, 1990).

The correlations among the other set of variables were similar in direction and magnitude across the two educational levels. Except for one correlation, which will be noted in the text as significant beyond the .01 level, all of the relationships among the different variables are significant beyond the .001 level. Participants who are prosocially oriented anticipate guilt reactions for detrimental acts ($r = .32$), exhibit low aggression proneness as reflected in irascibility ($r = -.13$) and ruminate hostility ($r = -.19$), and are disinclined to behave aggressively ($r = -.27$) or transgressively ($r = -.32$).

Participants who experience guilt over detrimental conduct cannot stop thinking about their troublesome experiences and perceived grievances ($r = .09$, $p < .01$) and refrain from aggressive behavior ($r = -.15$) or engagement in delinquent activities ($r = -.25$). Those who ruminate hostilely behave more aggressively ($r = .32$) and transgressively ($r = .29$). Irascibility is similarly related to aggressive behavior ($r = .42$) and pursuit of delinquent activities ($r = .30$).

Although the various mechanisms of moral disengagement operate interrelatedly as a single factor, there is some indication that they differ in their relative contribution to detrimental behavior. The various disengagement mechanisms were grouped into the subsets depicted in Figure 1 depending on whether they operated on reconstruing the conduct, obscuring personal responsibility, misrepresenting injurious consequences, and vilifying the victims. Responses to items were summed within subsets and correlated separately with aggressive and delinquent behavior. For delinquent behavior, the correlates were $r = .42$ for moral reconstrual of harmful behavior, $r = .19$ for obscuring responsibility, $r = .21$ for misrepresenting harmful consequences, and $r = .39$ for vilifying the victims by blaming and dehumanizing them. For aggressive behavior, the correlates were $r = .34$ for moral reconstrual, $r = .16$ for obscuring responsibility, $r = .15$ for misrepresenting consequences, and $r = .33$ for vilifying the victims. Each of these correlations was significant beyond the .0001 level. Thus, across both classes of detrimental activities, the reconstrual of injurious behavior as serving worthy purposes and vilifying the victim exerted the greater disinhibitory impact. It is easy to hurt others when such conduct is viewed as doing worthy things with unworthy people.

Paths of Influence

The posited causal structure was tested with data from the junior high students because they provide the full set of theoretically relevant variables. The factors in the hypothesized model included moral disengagement, prosocialness, anticipatory guilt and restitutive reactions for harmful conduct, and a latent factor labeled Aggression Proneness. This factor was composed of the measured variables of hostile rumination and irascibility, which heighten propensity to detrimental action. The high loadings of .71 and .72, respectively, for the model concerning delinquency, and .67 and .76 for the model concerning aggression, show it to be a well-defined latent construct. The outcome variables in the structural model were physical and verbal aggression and engagement in delinquent activities. Delinquent acts, involving as they do more serious offenses than expression of verbal or physical aggression, require a greater exercise of moral disengagement. Hence, we conducted separate analyses on impact of moral disengagement on these two classes of behavior. Socioeconomic level was not included as a control variable because it was related neither to the predictive factors nor to the outcome variables. We tested the conceptual model on the covariance matrices using the EQS program (Bentler, 1989).

The results of the structural equation modeling for delinquent behavior are presented in Figure 3. The goodness of fit of the model to the data was corroborated by all of the fit indices considered. The tests yielded a nonsignificant $\chi^2(3, N = 659)$ of 3.29, $p = .35$, a Normed Fit Index (NFI) of .996, a Non-Normed Fit Index (NNFI) of .998, and a Comparative Fit Index (CFI) of 1.00. T tests revealed each of the constituent paths to be an essential part of the model.

In accord with the posited model, moral disengagement influenced delinquent behavior both directly and by reducing prosocialness and anticipatory guilt over transgressions and by fostering aggression proneness. Prosocialness and anticipatory guilt reactions influenced delinquency through their restraining effect on such conduct and by their impact on aggression proneness. Prosocialness increased feelings of guilt and counteracted propensity to aggression, whereas guilt roused ruminations about perceived grievances and irascibility. Heightened aggression proneness, in turn, increased the level of delinquent behavior. The full set of sociocognitive factors accounted for 31% of the variance in delinquent behavior. The direct (.20) and medi-

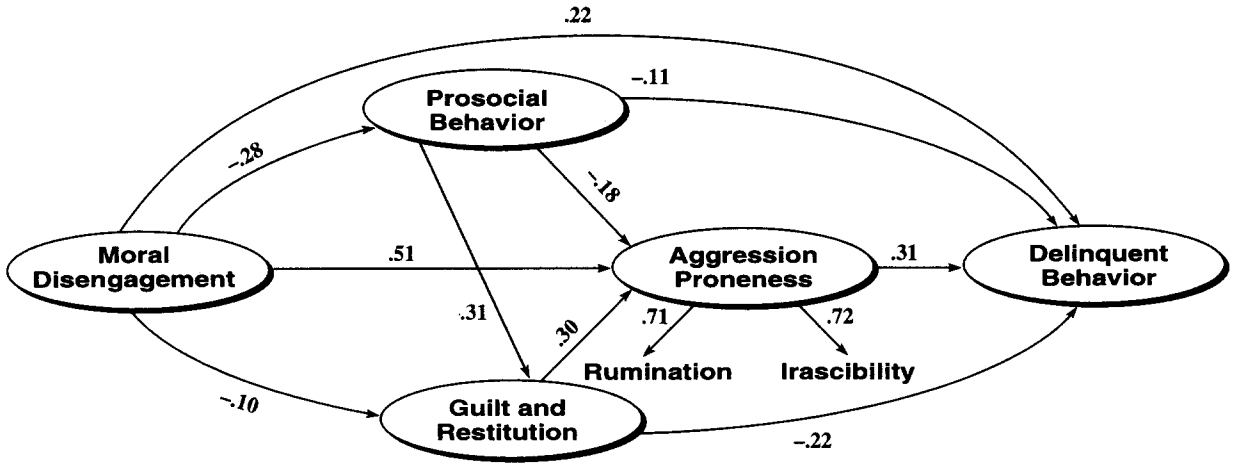

Figure 3. Contribution of moral disengagement to the multivariate determination of delinquent behavior. All paths of influence are significant at $p < .05$ or less.

ated (.23) effects of moral disengagement combined for a high total impact (.45) on delinquent behavior.

All but one of the posited paths of influence on aggressive behavior were significant. No direct link was found between moral disengagement and aggressive behavior. The various indexes of goodness of fit showed a good fit of the model to the data—an NFI of .982, an NNFI of .928, and a CFI of .986—but a significant $\chi^2(3, N = 666) = 13.37, p < .01$ was obtained. However, the dependence of this statistic on sample size makes it a less sensitive test with a large sample. Figure 4 presents the results of the structural equation modeling for aggressive behavior with the one nonsignificant path removed.

The influence of moral disengagement on aggressive behavior was mediated through prosocialness, guilt, and aggression proneness. High moral disengagement reduced prosocialness and guilt reactions and promoted cognitive and affective reactions that are conducive to aggression. The three mediating factors operated in the same way as they did in delinquent behavior, with prosocialness and anticipatory moral self-sanctions curbing aggression and aggression proneness heightening it. The set of factors in the model accounted for 34% of the variance in aggressive behavior.

Several alternative causal models also were tested. One such model assumes that weak prosocialness, low guilt, moral disengagement, and detrimental behavior are all simply coeffects of aggression proneness. This causal structure provides a very poor fit to the empirical data. The tests for goodness of fit for delinquent behavior yielded a highly significant $\chi^2(9, N = 659) = 215.54, p < .001$, and an NFI of .71, an NNFI of .52, and a CFI of .71. The results of the corresponding tests for aggressive behavior are $\chi^2(9, N = 666) = 130.35, p < .001$, NFI = .74, NNFI = .58, and CFI = .75.

Another alternative model posits that aggression proneness affects detrimental behavior both directly and through the mediation of prosocialness, guilt reactions, and moral disengagement. This causal structure also provides a poor fit to the empirical data. The fitness tests for delinquent behavior yielded a highly significant $\chi^2(6, N = 659) = 152.49, p < .001$, and an

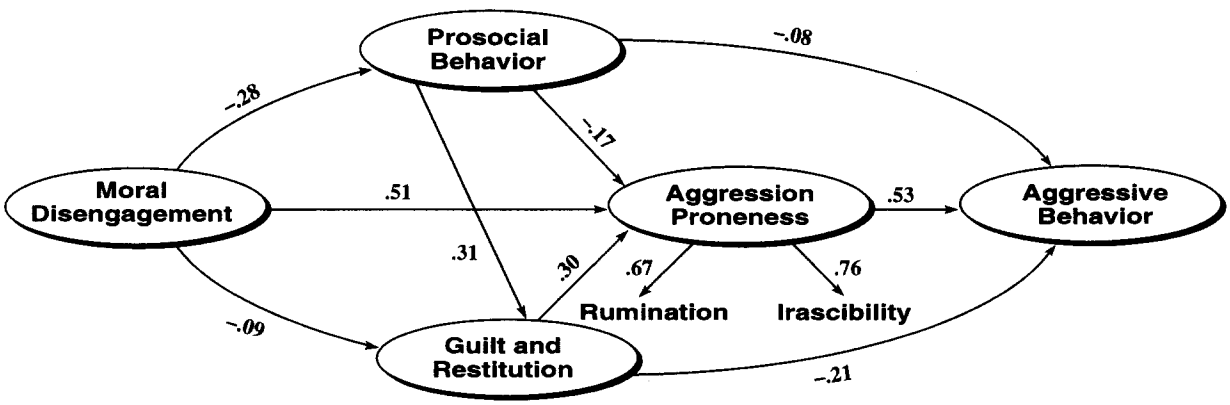

Figure 4. Contribution of moral disengagement to the multivariate determination of aggressive behavior. All paths of influence are significant at $p < .05$ or less.

NFI of .79, an NNFI of .49, and a CFI of .80. The results of the tests for aggressive behavior are $\chi^2(6, N = 666) = 163.99$, $p < .001$, NFI = .78, NNFI = .45, and CFI = .78.

Discussion

The findings of the present study lend considerable support to the influential role played by mechanisms of moral disengagement both in detrimental and prosocial conduct. In accord with prediction, high moral disengagers are more readily angered and behave more injuriously than those who apply moral self-sanctions to detrimental conduct. Ready moral disengagers are also more prone to engage in thought patterns that are conducive to aggression. They ruminate about perceived grievances and dwell on punitive retaliations. By contrast, when self-reproof remains engaged to detrimental conduct, adolescents are quicker to forgive and forget. Cognitive self-arousal through vengeful rumination perpetuates a high level of anger long after social slights or other provocations have ceased (Bandura, 1973). Anger arousal primes one for vindictive action (Berkowitz, 1990; Zillman, 1983). Both the findings of the present naturalistic study and laboratory tests of heightened punitiveness over a prolonged time span (Caprara et al., 1985; Zelli, 1984) corroborate the link between hostile rumination and interpersonal aggression.

Whereas moral disengagement weakens self-restraints over injurious conduct, adherence to moral self-sanctions fosters prosocial relations. It is difficult to hurt others who are humanized and not blamed entirely for their life predicaments. Adherence to self-sanctions against injurious conduct is strengthened not only by a sense of empathy but also by assuming personal responsibility for one's actions and not minimizing their injurious effects. The obtained relationship between low moral disengagement and prosocial behavior is consistent with evidence from controlled experiments in which personal responsibility and humanization are systematically varied (Bandura et al., 1975). People refuse to behave cruelly, even under high instigation to do so, if they act under personalized responsibility and recipients are humanized. Prosocialness increases anticipatory self-reproof for injurious conduct and attempts at restitution should self-restraint temporarily fail.

Psychological theorizing and research tend to emphasize how easy it is to bring out the worst in people through dehumanization and other self-exonerative means. Thus, for example, the aspect of Milgram's (1974) research on obedient aggression that is widely cited is the evidence that good people can be coerced into performing cruel deeds. However, to get people to carry out punitive acts, the overseer had to be physically present, repeatedly ordering them to act injuriously and absolving them of any responsibility for the effects of their actions as they voiced their mounting concerns and objections. Orders to escalate punitiveness to more intense levels are largely ignored or subverted when remotely issued by verbal command. As Helm and Morelli (1979) noted, this is hardly an example of blind obedience triggered by an authoritative mandate. Moreover, what is rarely noted is the equally striking evidence that most people steadfastly refuse to behave punitively, even in response to incessant authoritarian commands, if the situation is personalized by having them see their victim or requiring them to inflict pain directly rather than remotely (Milgram, 1974). The concern with suspension of self-restraints over detrimental behavior is understandable considering the prevalence of people's inhumanities to one another. However, the power of humanization to counteract human cruelty is of considerable theoretical and social significance, but it continues to receive comparatively little attention. Increased research efforts are needed to clarify how the affirmation of common humanity can bring out the best in others.

If moral standards are disengaged from transgressive conduct it can be carried out free from restraints of anticipatory self-censure. Through cognitive reconstruals and disownment of a sense of personal agency, negative self-sanctions are unlikely to be activated. There is little reason to engage in self-reproof for behavior that has been rendered acceptable or for which one professes no responsibility. Indeed, the findings confirm that the better the moral disengagement the weaker the felt guilt and the less the need to undo any harm caused by detrimental behavior.

So far the discussion has centered on aggressive and destructive conduct. Self-regulatory mechanisms preside over transgressive activities as well as interpersonal aggression. High moral disengagers were much more inclined to engage in delinquent pursuits than those who adhered to self-regulatory standards under conditions that lend themselves easily to self-exoneration. High moral disengagers were also less prosocial, less troubled by anticipatory feelings of guilt, and more prone to resort to vengeful ruminations and irascible reactions. These factors were, in turn, related to delinquent behavior. The use of these mechanisms is by no means confined to habitual freelancing delinquents. Gabor (1994) documented in considerable detail the widespread illegalities and inhumanities committed in all walks of life by ordinary citizens with self-exonerative justifications.

Analysis of the pattern of influences on delinquent behavior reveals that the proposed conceptual model provides an excellent fit to the empirical data. The model not only fits the data well, but it also yields a better fit than alternative causal structures. Moral disengagement affects delinquent behavior both directly and indirectly through its influence on prosocial behavior, level of guilt, and aggression proneness. Moral disengagement operates through essentially the same paths of influence on aggressiveness except that it involves no direct link. High moral disengagers are less prosocial and less guilty over detrimental conduct—both of which, in turn, lessen restraints over aggressive actions. Aggression proneness, of course, bears a close resemblance to aggressive action, which gives it a larger mediational role. This most likely accounts for a major share of the difference in its strength of influence across the two classes of detrimental behavior.

A number of features of the present research add to the reliability and generalizability of the obtained relationships between moral disengagement and detrimental conduct. It is replicated across diverse sources of data, different methods of measurement, distinct forms of detrimental conduct, and variant sociodemographic characteristics of the participants. Moreover, the predictiveness of proneness to moral disengagement is not confounded by socioeconomic factors. The causal analyses should be interpreted with some caution, however, because of the cross-sectional design of the research. However, the findings

of experimental investigations involving the key variables attest to the direction of causality posited in the guiding causal structure. On the moral disengagement side, controlled variations in displacement and diffusion of responsibility, dehumanization, and euphemistic labeling lead people to behave more aggressively (Bandura, 1991; Diener, 1977; Milgram, 1974; Zimbardo, 1969). With regard to the mediational paths, hostile rumination and irascibility have been shown to heighten aggressiveness (Caprara et al., 1985). Hence, there are experimentally validated bases for the posited paths of influence. The present study is part of a larger longitudinal project. The causal influence of moral disengagement will be further tested longitudinally in a multiple-panel design.

A recently completed study by Elliott and Rhinehart (1995) of serious assaults and transgressions with American youths attests to the generalizability of the theory of moral disengagement. In accord with the findings from the present study, males exhibit higher levels of moral disengagement than do females. The cross-cultural replication indicates that some of the gender differences in aggression may reside in differential proclivity to disengage moral self-sanctions from injurious conduct. Moral disengagement also differs by age, with older youths being more prone to adopt self-exonerative devices in regard to serious offenses. There were no significant differences in moral disengagement by race or by socioeconomic level. The latter finding also concurs with that obtained in the present study. Proclivity to moral disengagement predicted both felony and misdemeanor assaults and thefts. Ready moral disengagement retained high predictiveness regardless of age, sex, race, religious affiliation, and social class. This high predictive consistency attests to the pervading role of the self-regulatory system in detrimental behavior.

The differences between attribution theory and social cognitive theory regarding the mechanism of victim blaming warrants some comment. In attribution theory (Weiner, 1986), ascriptions of responsibility affect behavior through the mediation of emotional reactions. For example, blaming victims for their plight arouses anger toward them, whereas placing the blame on situational causes arouses pity (Zucker & Weiner, 1993). This interpretation begs the question of the mechanism governing when and how emotion gets translated into action. Sometimes people act on their anger, but oftentimes they go to great lengths to conceal how they feel and are not about to vent their anger in action. The differential reactions to anger indicate that emotion is linked to action through a self-regulatory mechanism. Moreover, many transgressive situations are not emotionally arousing, but they lend themselves readily to self-exoneration for transgressive acts. Consider, for example, the moral disengagement item "If people are careless where they leave their things it is their own fault if they get stolen." In this situation, transgressors do not have to arouse themselves to a state of anger to pilfer desired goods. They exempt themselves from restraining self-sanction for transgressive conduct by contending that, through their negligence, victims have only themselves to blame. This analysis indicates that emotional arousal may be facilitory, but it is neither necessary nor sufficient for the perpetration of transgressive conduct.

In the social cognitive theory of moral agency (Bandura, 1991), moral conduct is motivated and regulated mainly by the ongoing exercise of self-reactive influence. The major self-regulatory mechanism, which is mobilized in concert with situational factors, operates through three main subfunctions. These include self-monitoring of conduct, judgment of the culpability of conduct in relation to personal standards and environmental circumstances, and affective self-reaction. In this conceptual scheme, ascription of blame to victims functions as a disengager of moral self-sanctions. Indeed, the findings of the present research reveal that the disengagement mechanisms operate in concert on detrimental behavior both directly and by their effects on anticipatory guilt, prosocialness, and emotion-arousing ideation.

Theories of aggression typically characterize influences that reduce restraints over aggression as disinhibitory. *Disinhibition* describes a process but does not specify the mechanisms governing the behavioral effects. In social cognitive theory (Bandura, 1986), transgressive conduct is regulated by two major sources of sanctions: social sanctions and internalized self-sanctions. Both control mechanisms operate anticipatorily. In control arising from social sanctions, people refrain from behaving transgressively because they anticipate that such conduct will bring them social censure or other forms of punishing outcomes. In control rooted in self-sanctions, people behave prosocially because it brings self-satisfaction and self-respect, and they refrain from detrimental behavior because it will give rise to self-censure. What is called disinhibition largely reflects the disengagement of controlling self-sanctions from detrimental conduct. Understanding of the nature of disinhibition can, therefore, be advanced by increasing knowledge of self-regulative disengagement.

The discussion thus far has centered on the role of disengagement mechanisms in the regulation of one's own injurious conduct. These mechanisms also affect how the inhumanities perpetrated by others are viewed. For example, displacement of responsibility not only weakens restraints over one's own detrimental actions but also diminishes concern over the suffering of those mistreated by others (Tilker, 1970). Collective moral disengagement can have widespread societal and political ramifications by supporting, justifying, and legitimizing inhumane social practices and policies.

Psychological theorizing and research on aggression has focused heavily on impulsive aggression. The massive threats to human welfare stem mainly from deliberate acts of principle rather than from unrestrained acts of impulse. It is the morally justified and principled resort to destructiveness that is of greatest social concern but is largely ignored in psychological analyses of inhumanities. Over the years, much reprehensible and destructive conduct has been perpetrated by ordinary, otherwise considerate people in the name of religious principles, righteous ideologies, nationalistic imperatives, and ruthless social policies (Bandura, 1986; Rapoport & Alexander, 1982; Sanford & Comstock, 1971). There is much to be gained from understanding how the facility for moral disengagement develops and how institutional justificatory strategies are used to enlist people for exploitive and destructive purposes.

References

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85,* 1275–1301.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications for cross-informant correlations for situational specificity. *Psychological Bulletin, 101,* 213–232.

Andrus, B. C. (1969). *The infamous of Nuremberg.* London: Fravin.

Bandura, A. (1973). *Aggression: A social learning analysis.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1990). Mechanisms of moral disengagement. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, theologies, and states of mind* (pp. 161–191). Cambridge, England: Cambridge University Press.

Bandura, A. (1991). Social cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development: Theory, research and applications* (Vol. 1, pp. 71–129). Hillsdale, NJ: Erlbaum.

Bandura, A. (1992). Social cognitive theory of social referencing. In S. Feinman (Ed.), *Social referencing and the social construction of reality in infancy* (pp. 175–208). New York: Plenum.

Bandura, A. (1995). *Multifaceted scale of mechanisms of moral disengagement.* (Available from A. Bandura, Stanford University, Stanford, California 94305-2130.)

Bandura, A., Underwood, B., & Fromson, M. E. (1975). Disinhibition of aggression through diffusion of responsibility and dehumanization of victims. *Journal of Research in Personality, 9,* 253–269.

Bentler, P. M. (1989). *Theory and implementation of EQS: A structural equations program.* Los Angeles, CA: BMDP Statistical Software.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist, 45,* 494–503.

Bolinger, D. (1982). *Language: The loaded weapon.* London: Longman.

Brock, T. C., & Buss, A. H. (1962). Dissonance, aggression, and evaluation of pain. *Journal of Abnormal and Social Psychology, 65,* 197–202.

Brock, T. C., & Buss, A. H. (1964). Effects of justification for aggression and communication with the victim on postaggression dissonance. *Journal of Abnormal and Social Psychology, 68,* 403–412.

Cairns, R. B., Cairns, B. D., Neckerman, H. J., Gest, S. D., & Gariépy, J. (1988). Social networks and aggressive behavior: Peer support or peer rejection? *Developmental Psychology, 24,* 815–823.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., & Perugini, M. (1994). Individual differences in the study of human aggression. *Aggressive Behavior, 20,* 291–303.

Caprara, G. V., Coluzzi, M., Mazzotti, E., Renzi, P., & Zelli, A. (1985). Effect of insult and dissipation-rumination on delayed aggression and hostility. *Archivio di Psicologia, Neurologia e Psichiatria, 46,* 130–139.

Caprara, G. V., & Pastorelli, C. (1993). Early emotional instability, prosocial behaviour, and aggression: Some methodological aspects. *European Journal of Personality, 7,* 19–36.

Caprara, G. V., Renzi, P., Alcini, P., D'Imperio, G., & Travaglia, G. (1983). Instigation to aggress and escalation of aggression examined from a personological perspective: Role of irritability and of emotional susceptibility. *Aggressive Behavior, 9,* 345–353.

Caprara, G. V., Renzi, P., Amolini, P., D'Imperio, G., & Travaglia, G. (1984). The eliciting cue value of aggressive slides reconsidered in a personological perspective: The weapons effect and irritability. *European Journal of Social Psychology, 14,* 312–322.

Caprara, G. V., Renzi, P., D'Augello, D., D'Imperio, G., Rielli, G., & Travaglia, G. (1986). Interpolating physical exercise between instigation to aggress and aggression: The role of irritability and emotional susceptibility. *Aggressive Behavior, 12,* 83–91.

Cohen, D., & Nisbett, R. E. (1994). Self-protection and the culture of honor: Explaining southern violence. *Personality and Social Psychology Bulletin, 20,* 551–567.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115,* 74–101.

Darley, J. M., Klosson, E. C., & Zanna, M. P. (1978). Intentions and their contexts in the moral judgments of children and adults. *Child Development, 49,* 66–74.

Diener, E. (1977). Deindividuation: Causes and consequences. *Social Behavior and Personality, 5,* 143–156.

Diener, E., Dineen, J., Endresen, K., Beaman, A. L., & Fraser, S. C. (1975). Effects of altered responsibility, cognitive set, and modeling on physical aggression and deindividuation. *Journal of Personality and Social Psychology, 31,* 328–337.

Dishion, T. J. (1990). Peer context of child and adolescent troublesome behavior. In P. Leone (Ed.), *Understanding troubled and troublesome youth* (pp. 128–153). Beverly Hills, CA: Sage.

Elliott, D. S., & Rhinehart, M. (1995). *Moral disengagement, delinquent peers and delinquent behavior.* Unpublished manuscript, Institute of Behavioral Sciences, University of Colorado.

Ferguson, T. J., & Rule, B. G. (1983). An attributional perspective on anger and aggression. In R. G. Geen & E. I. Donnerstein (Eds.), *Aggression: Theoretical and empirical reviews* (Vol. 1, pp. 41–74). New York: Academic Press.

Gabor, T. (1994). *Everybody does it: Crime by the public.* Toronto: University of Toronto Press.

Haritos-Fatouros, M. (1988). The official torturer: A learning model for obedience to the authority of violence. *Journal of Applied Social Psychology, 18,* 1107–1120.

Helm, C., & Morelli, M. (1979). Stanley Milgram and the obedience experiment: Authority, legitimacy, and human action. *Political Theory, 7,* 321–346.

Keen, S. (1986). *Faces of the enemy.* San Francisco: Harper & Row.

Kelman, H. C. (1973). Violence without moral restraint: Reflections on the dehumanization of victims and victimizers. *Journal of Social Issues, 29,* 25–61.

Kelman, H. C., & Hamilton, V. L. (1989). *Crimes of obedience: Toward a social psychology of authority and responsibility.* New Haven, CT: Yale University Press.

Kohlberg, L. (1984). *The psychology of moral development* (Vol. 2). New York: Harper & Row.

Kramer, M. (1990). The moral logic of Hizballah. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, theologies, states of mind* (pp. 131–157). Cambridge, England: Cambridge University Press.

Lutz, W. D. (1987). Language, appearance, and reality: Doublespeak in 1984. In P. C. Boardman (Ed.), *The legacy of language—a tribute to Charlton Laird* (pp. 103–119). Reno: University of Nevada Press.

McHugo, G. J., Smith, C. A., & Lanzetta, J. T. (1982). The structure of self-reports of emotional responses to film segments. *Motivation and Emotion, 6,* 365–385.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: Harper & Row.

Perry, D. G., Williard, J. C., & Perry, L. C. (1990). Peers' perceptions of the consequences that victimized children provide aggressors. *Child Development, 61,* 1310–1325.

Rapoport, D. C., & Alexander, Y. (Eds.). (1982). *The morality of terrorism: Religious and secular justification.* Elmsford, NY: Pergamon Press.

Reich, W. (Ed.). (1990). *Origins of terrorism: Psycholgies, ideologies, theologies, states of mind.* Cambridge, England: Cambridge University Press.

Sanford, N., & Comstock, C. (1971). *Sanctions for evil.* San Francisco: Jossey-Bass.

Tilker, H. A. (1970). Socially responsible behavior as a function of observer responsibility and victim feedback. *Journal of Personality and Social Psychology, 14,* 95–100.

Weiner, B. (1986). *An attributional theory of motivation and emotion.* New York: Springer-Verlag.

Zelli, A. (1984). *Ruolo delle differenze individuali nella condotta aggressiva: Variabili di personalita, termperamentali, psicofisiologiche* [Role of individual differences in aggressive behavior: Personality, temperamental and psychophysiological variables]. Tesi di Laurea. Universita Degli Studi di Roma "La Sapienza," Roma.

Zillman, D. (1983). Arousal and aggression. In R. G. Geen & E. Donnerstein (Eds.), *Aggression: Theoretical and empirical reviews* (Vol. 1, pp. 75–101). New York: Academic Press.

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska symposium on motivation* (Vol. 17, pp. 237–309). Lincoln: University of Nebraska Press.

Zimbardo, P. G. (1995). The psychology of evil: A situationist perspective on recruiting good people to engage in anti-social acts. *Research in Social Psychology, 11,* 125–133.

Zucker, G. S., & Weiner, B. (1993). Conservatism and perceptions of poverty: An attributional analysis. *Journal of Applied Social Psychology, 23,* 925–943.

Appendix

Mechanisms of Moral Disengagement

1. It is alright to fight to protect your friends.
2. Slapping and shoving someone is just a way of joking.
3. Damaging some property is no big deal when you consider that others are beating people up.
4. A kid in a gang should not be blamed for the trouble the gang causes.
5. If kids are living under bad conditions they cannot be blamed for behaving aggresively.
6. It is okay to tell small lies because they don't really do any harm.
7. Some people deserve to be treated like animals.
8. If kids fight and misbehave in school it is their teacher's fault.
9. It is alright to beat someone who bad mouths your family.
10. To hit obnoxious classmates is just giving them "a lesson."
11. Stealing some money is not too serious compared to those who steal a lot of money.
12. A kid who only suggests breaking rules should not be blamed if other kids go ahead and do it.
13. If kids are not disciplined they should not be blamed for misbehaving.
14. Children do not mind being teased because it shows interest in them.
15. It is okay to treat badly somebody who behaved like a "worm."
16. If people are careless where they leave their things it is their own fault if they get stolen.
17. It is alright to fight when your group's honour is threatened.
18. Taking someone's bicycle without their permission is just "borrowing it."
19. It is okay to insult a classmate because beating him/her is worse.
20. If a group decides together to do something harmful it is unfair to blame any kid in the group for it.
21. Kids cannot be blamed for using bad words when all their friends do it.
22. Teasing someone does not really hurt them.
23. Someone who is obnoxious does not deserve to be treated like a human being.
24. Kids who get mistreated usually do things that deserve it.
25. It is alright to lie to keep your friends out of trouble.
26. It is not a bad thing to "get high" once in a while.
27. Compared to the illegal things people do, taking some things from a store without paying for them is not very serious.
28. It is unfair to blame a child who had only a small part in the harm caused by a group.
29. Kids cannot be blamed for misbehaving if their friends pressured them to do it.
30. Insults among children do not hurt anyone.
31. Some people have to be treated roughly because they lack feelings that can be hurt.
32. Children are not at fault for misbehaving if their parents force them too much.

Note. The following items correspond to the various mechanisms of moral disengagement. *Moral justification:* 1, 9, 17, 25. *Euphemistic language:* 2, 10, 18, 26. *Advantageous comparison:* 3, 11, 19, 27. *Displacement of responsibility:* 5, 13, 21, 29. *Diffusion of responsibility:* 4, 12, 20, 28. *Distorting consequences:* 6, 14, 22, 30. *Attribution of blame:* 8, 16, 24, 32. *Dehumanization:* 7, 15, 23, 31.

Received August 28, 1995
Revision received November 20, 1995
Accepted November 22, 1995 ■