A Commentary on Moral Disengagement: The Rhetoric and the Reality

Albert Bandura
Stanford University

Albert Bandura
Department of Psychology
Stanford University
Stanford, CA 94305-2130
Email: bandura@stanford.edu
In responding to Dahl and Waltzer’s review of my book *Moral Disengagement*, I will briefly discuss the theory of moral disengagement and then address the substance of their evaluation.

**Social Cognitive Theory**

Social Cognitive Theory of moral functioning is grounded in an agentic theory of human behavior wherein moral behavior is motivated by and regulated through the exercise of self-sanctions (Bandura, 1991, 2008, 2016). During development individuals adopt standards of right and wrong that serve as guides and deterrents for behavior. They regulate their actions by the consequences they apply to themselves. They do things that provide satisfaction and a sense of worth while refraining from behaving in ways that violate their moral standards, because such actions evoke aversive self-sanctions. These sanctions take cognitive, affective, and behavioral forms. The cognitive form involves self-regard; there is no greater punishment than self-contempt. The affective form involves feelings of guilt, which when severe can be deeply disturbing. The behavioral form involves harsh self-treatment, including denial of pleasurable activities. Anticipatory self-sanctions keep behavior in line with moral standards.

Most of our traditional theories of morality focus heavily on cognitive aspects, mainly moral reasoning, to the neglect of moral behavior. Such truncated theorizing is based on the belief that commitment to moral standards compels moral conduct. However, the correlation between moral reasoning and moral action is essentially unknown because behavior was rarely measured. Cognitive theories of morality are only half the story because they fail to explain the gap between moral thought and moral action. Social Cognitive Theory addresses the second half of this story.
as well by specifying the motivational and self-regulatory mechanisms through which moral standards and moral thought are linked to moral action.

**Moral Disengagement**

Moral standards alone are insufficient to ensure moral behavior. They can be circumvented and even enlisted in the service of harmful activities. Indeed, we are currently witnessing a pervasive moral paradox in which people from all walks of life commit detrimental acts that violate their moral standards, but still retain positive self-regard. In my book, *Moral Disengagement: How People Do Harm And Live With Themselves*, the theory specifies eight psychosocial mechanisms that weaken or eliminate moral self-restraints over detrimental practices without self-censure (Bandura, 2016). Figure 1 presents schematically the eight mechanisms through which moral self-sanctions are selectively disengaged from harmful behavior at four points in moral self-regulation.

**Insert Figure 1**

At the justification locus, individuals sanctify harmful means by investing them with worthy purposes. Such justifications can take righteous, religious, social, economic, and civic forms. Harmful behavior can also be rendered benign or even altruistic through advantageous comparison. Clever contrasts portray the harm individuals commit as minor to the harm it prevents or that others commit. In the utilitarian form of advantageous comparison, the harmful activity is construed as acceptable because it benefits the common good. Euphemistic language in its sanitizing forms cloaks harmful behavior in beneficent language and removes humanity from it. In convoluted forms it conceals harmful behavior in unfathomable jargon, making it difficult to know what is being done. These three mechanisms differ from the others because they serve a
dual function. They engage morality for the mission or cause, but disengage morality for the implementation of the detrimental activities.

At the agency locus, people absolve themselves of personal accountability for harmful conduct by displacing responsibility onto others and by dispersing it widely so that no one bears responsibility. At the outcome locus, perpetrators disregard, minimize, distort, or even dispute the injurious effects of their actions. There is no moral issue if detrimental practices are judged to be harmless or even beneficial. At the victim locus, perpetrators dehumanize those they mistreat from their category of humanity by divesting them of human qualities or by attributing animalistic qualities to them. Additional moral disengagement at the victim locus blames victims for bringing maltreatment upon themselves or attributes it to compelling circumstances. When used together, these mechanisms operate as powerful moral disengagers (Bandura, 2016).

Bin Laden provides an excellent example of how extensive inhumanities can be perpetrated with equanimity using moral disengagement practices. Through moral justification, bin Laden sanctifies his global terrorism as serving a holy imperative: “We will continue this course, because it is part of our religion, and because Allah ordered us to carry out jihad so that the word of Allah may remain exalted to the heights.” He displaces the responsibility for the holy terror to Allah; they are carrying out their “religious duty”. Through attribution of blame, he construes terrorist strikes as morally justifiable defensive reactions to humiliation and atrocities perpetrated by “decadent infidels”: “We are only defending ourselves. This is defensive Jihad.” By advantageous comparison with the nuclear bombing of Japan, and the toll of economic sanctions on Iraqi children, the Jihad takes on an altruistic appearance: “When people at the ends of the earth, Japan, were killed by the hundreds of thousands, young and old, it was not considered a war crime, it is something that has justification. Millions of children in Iraq is
something that has justification.” He bestializes the American enemy as “lowly people”, perpetrating acts that “the most ravenous of animals would not descend to.” Terrorism is linguistically sanitized as “the winds of faith have come” to eradicate the “debauched” oppressors. His followers see themselves as holy warriors who achieve a blessed eternal life through their martyrdom.

**Deconstructing the Dialectic**

In their book review, Dahl and Waltzer (2017) state “that moral disengagement is not explicitly defined.” In point of fact, moral disengagement is well-defined as a phenomenon in which moral self-sanctions are disengaged from detrimental behavior. A moral paradox is defined as a phenomenon in which individuals behave harmfully, all while maintaining positive self-regard, without anguish over the harm they cause. The construct is complimented by its connectedness to theory that specifies the nature, form, and functional properties of the psychological mechanisms by which moral self-sanctions are disengaged from detrimental behavior.

**Prioritization As Moral Disengagement**

In Dahl and Waltzer’s prioritization conceptual scheme, they state that sometimes individuals find themselves in moral predicaments in which they are “forced to prioritize competing principles, violating one principle in order to follow another”. In their view transgressors still remain duty-bound to the violated principle. As a result, they are conflicted and distressed.

This conceptual scheme rests on the notion that continued allegiance to the violated principle leaves wrongdoers conflicted and anxious. Do the authors measure strength of allegiance to the violated principle independently? Or is its presence assumed from conflict and
distress? If the latter is the case, it creates a problem of explanatory circularity. If stress stands for allegiance, stress becomes a cause of itself. Dahl and Waltzer talk about prioritization in the abstract but they do not explain what individuals are doing when they are prioritizing, how it produces a competing moral principle, and how the principle gives rise to detrimental behavior. Without specifying the causal links in this intricate chain of events, prioritization is a descriptive construct, not an explanatory one.

Lying is often used as the moral predicament in hypothetical situations in studies of prioritization. Dahl and Waltzer used evidence from such a study as support for their view. A study of real-life deception by Wexler (2006) suggests that prioritization operates through moral disengagement. He studied real-life deceptions by résumé fraudsters in high managerial positions. He found that fraudsters enlisted the entire set of moral disengagement mechanisms in the service of self-exoneration. They trivialized lying about their qualification as a harmless minor matter: “It’s silly to hold me in a negative light when no one is complaining”. They linguistically sanitized their actions as “résumé enhancement”, not lying. In their view, by advantageous comparison, their deceit paled in comparison to the dishonesty, licentiousness, and hypocrisy of those who dismissed them. As further comparative self-exoneration, they invoked the corruptness of the corporate system itself: “I feel like a just man in an unjust system. Others are stealing, padding expenses and raising Cain, but are not losing their jobs”. They likened life in the fiercely competitive corporate world to a game in which everyone is driven to do shady things to win.

They even converted deceptive behavior into a positive trait: “I am precisely what the firm wants, needed, and advertised for – a bold, risk-taking innovator who pushes the boundaries”. Their résumé embellishment, they explained, was strong testimony to their
motivation and commitment to becoming valued members of the company. They maintained their positive self-regard as good, decent people who should be judged by the results they produce: “Good people get results and create more benefits than costs. I am using this definition of a good person”.

Viewed from the perspective of moral disengagement theory, during prioritization individuals are creating justifications designed to make transgressive behavior permissible. The fraudsters enlisted a variety of mechanisms for this purpose. The theory also provides an empirically testable proposition on whether or not individuals will be conflicted and distressed over choosing the transgressive path. The fraudsters not only were not distraught, but they elevated deceptive behavior to an admirable organizational trait. Distress is conditional on the self-persuasive power of the justifications. Weak justification leaves one perturbed, highly self-persuasive ones do not. Whether the source of distress is attributable to weak justification or strength of allegiance to the rejected principle cannot be empirically tested without a valid measure of strength of allegiance.

**On Detachment of Causal Factors**

In Dahl and Waltzer’s view of moral disengagement theory, they omitted the moral engagement component from the causal structure. As noted earlier, three of the mechanisms of moral disengagement, operating at the justification locus of moral control, differ from the others because they serve a dual function. They engage morality in the mission or cause but disengage morality in implementing the harmful practices by which the mission is achieved. These dual functions are well illustrated in bin Laden’s terroristic jihad cited earlier. He and his holy warriors were deeply morally engaged in doing Allah’s bidding in fighting a holy imperative to protect their religion from oppressive “debauched infidels”.
Linking one’s morality to a religious doctrine is the moral engagement function. Having to kill numerous innocent people, with families among them, requires additional moral disengagement methods whereby their victims are diminished as human beings. To enable his warriors to commit atrocities, bin Laden bestialized the enemy as the most “ravenous of animals”; he portrayed his global terrorism as a defensive jihad forced by “decadent infidels”; he made his terrorism look altruistic protective acts compared to the nuclear bombing of Japan and Iraqi children starved by economic sanctions; and sanitized his jihad as the “winds of faith”.

Nowhere in their critique do Dahl and Waltzer mention the moral engagement function. It is not that this subset of mechanisms is a minor, obscure one that can be easily overlooked. Quite the contrary. They are highlighted throughout the book, their functional role is explained, and the way in which they are manifested in different activity domains is illustrated with multiple examples. Omission of the moral engagement component is an important matter because it addresses how moral principles are adopted and elevated in importance for a given sphere of activity. With the omission, Dahl and Waltzer are critiquing a truncated version of moral disengagement theory.

**On Equating Harm with Immorality**

According to Dahl and Waltzer, Moral Disengagement Theory states that people “turn off” a moral principle and “delude themselves into viewing” harmful activity as permissible. They state that I consider many harmful activities immoral, “…such as capital punishment, counter-terrorism, and gun rights activism”. Moral disengagement is indeed concerned with explaining behavior that causes harm. The judgement of whether harm is immoral is a separate issue that falls within the domain of ethics. I am not equating harmful behavior with immoral
behavior, nor do I believe “…that anyone who endorses capital punishment must have a misguided moral view”.

To clarify this distinction, let’s take the example of soldiers at war. They face the task of killing fellow human beings. No one would deny that their behavior causes harm. However, this does not make it immoral. Moral rightness or wrongness is in the eyes of the beholder. Moral disengagement centers on the means by which a soldier can kill without being devastated. Society must convince soldiers that they are fighting a just cause. Usually the justification for war involves protecting the safety, freedom and security of its citizens. When there is unified public support for the cause, soldiers enter warfare with a powerful justification for killing. They are not “deluding themselves” and “turning off” a moral principle. They are engaging morality in the service of their country. The moral engagement aspect of Moral Disengagement Theory is critical for understanding how soldiers can kill without being devastated and even take pride in their role in battle. For example, both the U.S. and al Qaeda use the same mechanisms of moral disengagement to conduct war. The United States justified use of lethal attacks in self-defense against terrorist attacks by al Qaeda. Bin Laden justified use of lethal attacks in self-defense against invading “decadent infidels” threatening their Islamic faith. Mechanisms of moral disengagement enable both U.S. and al Qaeda combatants to kill each other without experiencing devastating intrapersonal consequences. In short, moral disengagement theory does not take sides; it seeks to explain how people are able to carry out warfare.

**Self-Regard as “Selfish Motivation”**

Dahl and Waltzer argue that moral disengagement theory is founded on the idea that people act to preserve their positive self-image. As they put it, “the theory presupposes that people do not inherently care about others’ welfare, rights and honesty, but conform to moral
principles only insofar as such conformity preserves their positive self-image.” They construe effort to pursue positive self-regard as incompatible with caring about the well-being of others.

This mistaken judgment is partly the result of them disregarding the moral engagement portion of the theory. It governs the causes to which people commit themselves. Self-sanctions do not operate unconditionally; they are linked to moral standards. In fact, the type of behavior self-sanctions support depends on the type of moral standard to which people are committed. For example, those who link their self-respect to social justice behave in socially just ways and fight for social justice for others. Those who link their self-respect to social inclusiveness behave in socially embracing ways, and fight discrimination. In short, the pursuit of self-respect is not inherently selfish.

**Hidden Conflict or Moral Disengagement?**

Dahl and Waltzer speculate that court testimony of corporate executives does not reveal the conflict they experience about promoting harmful products. In self-protective maneuvers, “…these individuals may well have experienced conflict about their actions even when they did not say so in public”. They conclude that conflict about violating moral principles may exist, even though it is not revealed. They use this to support their prioritization theory; we just can’t see the conflict, but it’s likely there.

I drew on a large volume of in-house memos when documenting moral disengagement in the tobacco industry. They are not statements made in court, but rather are documents of plans and actions of industry personnel and their consultants.

For example, when the surgeon general’s groundbreaking report was being prepared, an industry counsel advised withholding from the surgeon general the troublesome results of their own commissioned research regarding cardiovascular disease. The biggest challenge for the
tobacco industry was growing evidence of the toxicity of secondhand smoke. A public relations firm recommended a two-step action plan to negate an EPA report which classified secondhand smoke as a carcinogen. The first step was to create a “sense of doubt” by attacking the science as flawed. In the second step, they minimized the harm “by comparison with risks associated with everyday activities”.

In exoneration by advantageous comparison, tobacco is likened to other common consumer products. “The product is a ‘common consumer product’ intended for personal consumption, such as sugar, castor oil, alcohol, and butter”. To sanitize the addictiveness of smoking, other common substances that people consume are turned into addictions as well. “Many substances, including caffeine and alcohol, can produce physical or psychological dependence. ‘Chocoholics’ are ‘addicted’ to chocolate”. Staunch defenders of the industry argued comparatively that targeting young people with cigarette ads glamorizing smoking is nothing unusual: “Every product is produced with some ‘target audience’ in mind…Dental adhesive is aimed at people who wear false teeth”.

The harmful effects of smoking were masked by sanitizing language. Industry scientists converted the carcinogenicity of tobacco products to “specific biological activity”. Secondhand smoke became indoor air pollution or “ambient smoke”. The industry derogated research that demonstrates adverse health effects as “half-truths in the hands of fanatics”, “scientific malpractice”, and “Orwellian Official Science”.

These documented activities have nothing to do with legal proceedings. They reveal the practices of the tobacco industry and how they used mechanisms of moral disengagement to promote a product that contributes to the premature death of half a million people annually. It is
inappropriate for Dahl and Waltzer to discount the presence of moral disengagement within the tobacco industry because hidden moral conflicts “might exist” in court testimony.

**Moral Theorizing in Disparate Realities**

In this review, we are dealing with markedly different realities. Dahl and Waltzer focus heavily on minor transgressions and moral predicaments in hypothetical scenarios. Consider a few examples. A husband faces the choice of whether to steal a drug to save his wife’s life. A person is forced to choose between letting five persons be killed by a runaway trolley or saving the five persons by sacrificing the life of one person by throwing him on the tracks. Research based on this line of theorizing favors hypothetical reality.

Dahl and Waltzer subscribe to a purely cognitive theory of morality limited to prioritizing conflicting principles. The purely cognitive view is rooted in Emmanuel Kant’s moral theory. He argued that once people discover a cardinal principle they are duty bound to follow it regardless of situational circumstances. Because moral thought compels moral action, in their view, there is no need for motivational and self-regulatory processes mediating the link between moral thought and moral action. Kohlberg (1984) joined Kant with Piaget to create a stage theory of moral reasoning that dominated research on morality for years. Hypothetical reality is easy to study. The purely cognitive view of morality continues to be favored.

The reality of moral disengagement theory focuses on serious moral issues of real life. These include financiers spawning a global economic crisis in which millions of people lost their homes, pensions and livelihood; a tobacco industry whose products contribute to the premature deaths of half a million people annually; ordinary citizens as jurors condemning persons to death and other citizens having to kill them; terrorists arousing fear world-wide by unpredictable suicide bombings of civilian populations; individuals blocking restrictions on social and
Industrial practices that are changing the earth’s climate, destroying the ecological supports of life and passing on to future generations a less habitable planet.

Heavy reliance on the hypothetical reality is not the only concern about the scope of theorizing and research in the field of moral psychology. The problem of generalizability of findings from hypothetical scenarios to real life inhumanities does not seem to be of concern to Dahl and Waltzer in their claims that their findings challenge the soundness of moral disengagement theory.

Moral disengagement theory was designed to explain the moral paradox – how people can behave harmfully while retaining positive self-regard and a sense of well-being. Dahl and Waltzer state that conflicts between principles are an inherent feature of moral functioning. In this view, wrongdoers cannot disengage from a violated principle and live peacefully. With this line of reasoning, Dahl and Waltzer unknowingly argue for the irrelevancy of their conception to explaining the aforementioned moral paradox.
REFERENCES


Author Identification Footnote

Albert Bandura, Department of Psychology, Stanford University, Stanford, California.

Correspondence concerning this article should be addressed to Albert Bandura,

Department of Psychology, Stanford University, Stanford, California 94305-2130; e-mail: Bandura@stanford.edu.

I am grateful to Mary Bandura for her comments on an earlier version of this article.
Figure 1

Eight mechanisms through which moral self-sanctions are selectively disengaged from harmful behavior at four points in moral self-regulation (Bandura, 1986).