Moral Disengagement as a Psychological Construct

Audun Dahl and Talia L. Waltzer

University of California, Santa Cruz

Author note

Audun Dahl, Department of Psychology, University of California, Santa Cruz;

Talia Waltzer, Department of Psychology, University of California, Santa Cruz.

Correspondence concerning this article should be addressed to Audun Dahl, Department of Psychology, University of California, Santa Cruz, CA 95064. E-mail: dahl@ucsc.edu

Moral Disengagement as a Psychological Construct

In 2013, Edward Snowden began leaking classified information to the press about government surveillance by the US National Security Agency (NSA) (Reitman, 2013). In considering whether to leak information about the NSA surveillance program, Snowden faced a dilemma. On the one hand, blowing the whistle on the NSA surveillance could protect privacy rights and prevent the government from abusing people's personal information. On the other hand, disclosing classified information would be illegal and violate written promises Snowden had made. It could also damage US counterterrorism efforts and international relations. Reflecting the complexity of the dilemma, people differ in their evaluations of Snowden's disclosures. Whereas some have praised him as a hero of an open society, others have condemned him for being disloyal (Brooks, 2013; Reitman, 2013).

While extreme, Snowden's situation highlights a key feature of morality: In many situations, individuals sometimes act and judge in violation of general moral principles, for instance by breaking promises yet generally holding that promises ought to be kept. In his recent book, *Moral disengagement: How people do harm and live with themselves,* Albert Bandura aims to provide an explanation for why people sometimes violate their general principles (Bandura, 2015). According to the theory of moral disengagement, people violate general principles by "disengaging," or turning off, the violated principle to avoid thinking that they have done something wrong: "It is the selective suspension of morality for harmful activities that enables people to retain their positive self-regard while doing harm" (Bandura, 2015, p. 3). Through selective moral disengagement, Bandura argues, people can engage in behaviors they would normally

have good reasons to condemn without experiencing conflict or losing self-respect.

When morally disengaging, people are said to first decide which actions they would like

to pursue, and then delude themselves into viewing those actions as permissible. This

idea is consistent with the notion, proposed by some contemporary theories, that people

typically do not reason about moral principles when making moral judgments (e.g.

(Greene, 2014; Haidt, 2001). Instead, reasoning is primarily thought to happen when

seeking to justify the actions to oneself or others, a phenomenon sometimes called

"post-hoc rationalization" (Haidt, 2001).

The present book review takes a critical perspective on moral disengagement as

a psychological construct. Our view is based on the notion that individuals are

sometimes forced to prioritize competing principles, violating one principle in order to

follow another (Turiel, 2003; Turiel & Dahl, in press). For instance, children and adults

hold general principles about keeping promises, not harming others, and not lying. Still,

most think it is sometimes permissible or even required to violate such general moral

principles when they conflict with other principles (Dahl, Uttich, Gingo, & Turiel, 2013;

Nucci & Turiel, 2009; Perkins & Turiel, 2007). Violating a general principle, for instance

by lying to protect someone's life, does not imply a lack of concern for honesty in that

situation, nor does it indicate the absence of internal conflict. Rather, it may reflect a

decision to prioritize other principles deemed more important than honesty, for instance

human life. In our view, then, the violation of one moral principle in order to adhere to

another principle is, by itself, not a sign of pathology or a lack of morality. On the

contrary, such balancing of competing principles is an inherent feature of moral

functioning (Dahl & Killen, in press; Turiel, 2003; Turiel & Dahl, in press).

After summarizing Bandura's theory, this review will question two key premises underlying theory and research on moral disengagement. First, we question the premise that people would always consider it wrong to violate moral principles if they were not biased by self-interest. Contrary to Bandura's claims, people often think it is required to violate one principle in order to give priority to another in situations of conflict: They do not "turn off" morality, but seek to coordinate competing evaluative concerns, such as honesty and prevention of harm. Second, we question the premise of moral disengagement theory that people who violate general principles are often not conflicted about their actions, or at least not as conflicted as they should be. Contrary to this assumption, research shows that people are often conflicted about violating moral principles, even when they judge that such violations are called for. The implication of our arguments is that many or most of the individuals said to be morally disengaged may in fact have been highly morally engaged, even if those individuals did not think or act in the way researchers think they should have.

Moral Disengagement

In *Moral disengagement*, Bandura (2015) aims to explain how people can violate moral principles and still "feel good about themselves" (p. 1). While his discussion of morality is focused on harmful behaviors, Bandura uses morality more broadly to refer to compassion, efforts to further the wellbeing of others, and the ability to refrain from "behaving inhumanely" (p. 1). To explain why people sometimes violate moral principles, Bandura has proposed the construct of moral disengagement (see also (see also Bandura, 2002). The present book sets forth a number of thought-provoking proposals that merit serious discussion. Moreover, the book contains erudite surveys of,

and impassioned contributions to, debates about gun rights, capital punishment, terrorism, and other issues of great societal importance.

Bandura (2015) begins by proposing that the desire for positive self-evaluation is the main motivation for adhering to moral principles. He argues that people avoid "behaving in ways that violate their moral standards because such conduct will bring self-condemnation" (p. 1) and that "[i]t is not the moral principles or standards per se but the investment of one's self-regard in how one lives up to those standards that governs the motivation and self-regulation of moral conduct" (p. 29). Hence, 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. Furthermore, it is assumed that people would lose self-respect if they viewed themselves as violating moral principles. Hence, Bandura argues, in order find harmful behavior acceptable in certain situations – for instance endorse capital punishment for some crimes – people must "morally disengage" in those situations.

Although the book does not provide an explicit definition of moral disengagement, it uses the term when the following two elements co-occur: First, a person must have deemed permissible an action that violates some evaluative principle. Second, the person must not view their action negatively, or at least not as negatively as the person was supposed to.

Bandura proposes several ways by which a person can avoid or reduce a negative self-evaluation following the violation of a principle. People are said to alter their perception of their own actions by justifying their actions on social or moral grounds, comparing their actions to worse acts, or using euphemistic language to

describe the acts. To illustrate the operation of moral disengagement through justifications, Bandura (2015) discusses the US Army Sergeant Alvin York, who received the Medal of Honor for his actions in World War I. York was initially a conscientious objector, but eventually came to think that fighting in a war was sometimes permitted by God. Bandura argues that such incidents illustrate how "[s]ocial and moral justifications sanctify harmful practices by investing them with honorable purposes" (p. 49).

Other forms of disengagement discussed by Bandura (2015) include displacement or diffusion of responsibility, attempts to minimize the negative consequences of harmful actions on others, and dehumanization or blaming of victims. According to this framework, moral disengagement can occur when any or all of these mechanisms are used by an agent who commits a harmful or immoral act. In the later parts of the book, Bandura applies the moral disengagement framework to explain how the entertainment industry, gun industry, and corporate world can continue practices considered harmful. The book also seeks to explain how individuals can participate in or endorse practices Bandura considers harmful or immoral, such as capital punishment, counter-terrorism, and gun rights activism.

In previous research, the construct of moral disengagement has been applied to several issues in research on morality and its development. For instance, some researchers have documented correlations between aggressive behaviors and responses to the Moral Disengagement Questionnaire in children and adolescents (e.g. Bandura, Barbaranelli, Caprara, & Pastorelli, 1996; Hyde, Shaw, & Moilanen, 2010). In this research, participants are deemed to be morally disengaging if they are accepting

of children fighting to protect their friends, lying to keep their friends out of trouble, or misbehaving after being under excessive pressure from their parents (Bandura et al., 1996, p. 374).

Questioning Two Key Premises of Moral Disengagement Theory

Bandura's (2015) book on moral disengagement has the laudable aim of using moral psychological research to explain phenomena of societal importance. However, the success of such efforts will hinge, in part, on whether the theory of moral disengagement provides an accurate portrayal of moral psychology and moral development. In fact, we will argue that the theory of moral disengagement is at odds with much psychological research on morality and its development.

In particular, we propose that most of the examples Bandura (2015) discusses may not have involved disengagement from moral considerations. Consider two examples taken to involve moral disengagement: Alvin York, the decorated sergeant from World War I, and a research participant endorsing the statement "It is alright to fight to protect your friends" on the questionnaire often used to assess moral disengagement (Bandura, 2015; Bandura et al., 1996). According to the theory of moral disengagement, these individuals must have turned off their moral principles in order to accept violence. The attribution of moral disengagement is based on the general assumptions that violating moral principles "causes self-chastisement" (Bandura, 2015, p. 29). This assumption, we will argue, is at odds with research showing that people sometimes judge it permissible to violate one principle in order to prioritize another principle, and thus may have little reason to chastise themselves. Secondly, we will argue that attributions of moral disengagement do not adequately consider how

individuals often experience a great deal of conflict about their own judgments and actions. Suggesting of such inner conflict, Sergeant York was later critical of US involvement in World War I (Lee, 1985). Experienced conflicts about moral decisions would be difficult to explain if people could simply turn off evaluative principles whenever they violated those principles.

People deem violations of one moral principle as acceptable when judging that another, conflicting principle should be prioritized. The theory of moral disengagement proposes that it is inherently wrong or irrational to apply a principle in one situation and violate it in another (for a related view, see Greene, 2014). According to Bandura (2015), people would lose self-respect if they viewed themselves as violating a moral principle. However, complex situations can make it necessary to violate one principle in order to prioritize another (Dahl et al., 2013; Turiel, Hildebrandt, & Wainryb, 1991).

Research on responses to hypothetical dilemmas provide clear examples of how children and adults accept violations of one principle in order to prioritize another principle. In one of the dilemmas used in Kohlberg's (1971) research on moral orientations, a man faces the choice of whether to steal a drug to save the life of his wife. The so-called trolley car dilemmas force a protagonist to choose between letting five persons be killed by a runaway train or saving the five persons by sacrificing the life of another person. Such situations require participants to weigh competing considerations and, ultimately, give priority to one principle over another (Dahl et al., 2013). Children and adolescents, too, sometimes weigh competing considerations when evaluating hypothetical actions, for instance those involving acts of self-defense or

balancing of fairness and group loyalty (Dahl & Killen, in press; Nucci & Turiel, 2009; Turiel et al., 1991).

The above balancing of competing principles happened in response to hypothetical situations, and cannot be explained by participants' need to retain self-respect following their own violations. (Note that such balancing also occurs in real-life situations: See Turiel, 2003). Hence, violating one principle in order to prioritize another does not mean that a person is acting out of self-interest or is unconcerned with morality. Yet, research and theory of moral disengagement typically assumes that persons who deem it permissible to lie or harm someone are disengaging their moral sense. In one study, saying that military force is justified when it prevents more suffering than it causes was taken as a sign of moral disengagement (McAlister, Bandura, & Owen, 2006). However, this research did not consider the alternative possibility that people genuinely thought the suffering prevented by an intervention warranted the use of military force, and were not simply trying to prevent themselves from losing self-respect from endorsing the use of military force. To take an extreme example, it is hard to imagine that a person would be at risk of losing self-respect by saying that the Allied invasion of France on D-Day in 1944 was morally warranted.

In most situations, children and adults act in accordance with their moral principles, and rarely steal from, lie to, or harm others. We have argued here that people do make exceptions to general principles when competing evaluative principles are deemed more important. They do so even in response to hypothetical situations, when their self-respect is not at stake. Thus, the finding that people sometimes judge harmful behaviors as permissible in no way implies that people are not morally engaged

in these situations. On the contrary, research on responses to moral dilemmas and multifaceted situations show that people are concerned with moral right and wrong as they seek to decide which principle to prioritize.

Violating moral principles does not mean disengaging them. A key assumption of moral disengagement is that people can disengage morally by avoiding self-evaluation. This assumption is based on the idea people care about moral and other evaluative considerations only insofar as it serves to maintain their self-image (Bandura, 2015, p. 1). People's main motive for adhering to principles, according to this view, is to avoid their own negative self-evaluation (or the negative evaluations or sanctions from others). In line with these notions, Bandura (2015) writes that "[t]he explanatory challenge is not the occurrence of crime, but rather why it is not more frequent, given such low rates of apprehension and punishment" (p. 204). Since it is often possible to avoid punishment for moral violations and moral disengagement theory argues that people are adept at turning off their moral sense, the theory appears to imply that people regularly engage in moral violations.

Contrary to moral disengagement theory, there is evidence that people are greatly concerned with moral violations, even when they are neither a victim nor a transgressor. The above-cited research on judgments about hypothetical situations, in which participants make evaluative judgments about situations in which they are not involved, provides one example of moral concerns in the near-absence of self-interest. Starting in early childhood, people also show emotional concerns when observing the suffering of others, sometimes acting to relieve the suffering of others or protesting

when observing violations against others (for reviews, see Dahl & Killen, in press; Hoffman, 2000).

Insofar as people care about moral issues, they would expectably be conflicted about situations in which they harmed or otherwise violated the rights of another person. Indeed, there are numerous examples of children and adults expressing doubt, regret, or distress about having caused harm to others (Turiel et al., 1991; Wainryb, Brehl, & Matwin, 2005). In our study of responses to moral dilemmas among adolescents and adults, participants frequently expressed conflict about whether it was permissible to sacrifice one life to save five others (Dahl et al., 2013).

Bandura (2015) cites several statements by people who seemingly express little regret or conflict about behaviors described as harmful or immoral. These examples include public statements by gun rights advocates and tobacco firm executives defending themselves against criticisms or lawsuits. However, a difficulty with taking such statements as evidence of moral disengagement is that people may be especially unlikely to express doubt or regret when they are under attack or otherwise seeking to defend themselves or what they view as their legitimate interests. In other words, these individuals may well have experienced conflict about their actions even when they did not say so in public. Similarly, defendants who choose to exercise their Fifth Amendment right not to self-incriminate themselves in a court could very well regret what they did without expressing such regret in court.

In short, there is evidence that people are can be conflicted even as they choose to violate general moral principles. Moreover, the evidence provided by Bandura, for the most part, does not demonstrate an absence of conflict or regret among those who

11

engage in or support the practices he terms harmful or immoral. The implication is people said to morally disengage may in fact not disengage their moral principles but rather seek to determine which principle to give priority in situations of conflict. This makes for a very different psychological account of the phenomena discussed by Bandura, and calls for a reconsideration of moral disengagement as a psychological construct.

Moral Disengagement, Moral Misengagement, and Coping with Conflict

We have argued that many attributions of moral disengagement to individuals are at odds with research on the moral psychology of children and adults. Still, the questions that moral disengagement theory was designed to address are undoubtedly important. For instance, many readers will probably share Bandura's stances on gun rights, environmental issues, and capital punishment. These readers may agree with Bandura (2015) that the administration of capital punishment involves "stripping morality from every phase," (p. 280) that there is no "humane way to execute people," (p. 280) and that anyone who endorses capital punishment must have a misguided moral view. It is then tempting to ask: "How do individuals come to hold misguided views, endorsing immoral or inhumane practices?"

However, this question, while interesting and important, is not an empirical psychological question. Terms such as "inhumane" or "immoral" are evaluative concepts not assessed through observable characteristics of human activities. For instance, Bandura (2015) argues that capital jurors morally disengage by "displacing agency" (pp. 292-295). Specifically, Bandura argues that these jurors are displacing responsibility when they say that they are responsible for following the instructions from the court, and

not for the death of the sentenced person. However, to argue that such conceptions of the role of jurors is a *dis*placement of responsibility – i.e. *not* what it ought to be – presumes a particular conception about what jurors should and should not do. Such conceptions of the rightful role of capital jurors are not based on research findings, but on evaluative stances that fall outside the boundaries of empirical psychology.

Elsewhere, Bandura (2015) argues that politicians in Western countries morally disengage when they encourage citizens to have more children. Bandura claims that such encouragements are misguided because they contribute to overpopulation. For instance, when the German minister of family affairs expressed delight about the increase in German birthrates, Bandura argues that she conferred "moral legitimacy on harmful practices" (2015, p. 392-393). However, it is hardly a psychological fact that increasing birthrates in industrialized countries is a harmful practice devoid of moral legitimacy. In these and other cases, the attribution of moral disengagement is not primarily based on a psychological examination of the persons involved, but also on the evaluation – made by researchers – that the persons were doing or saying something inhumane or otherwise morally wrong. In this sense, moral *dis*engagement is not an empirical psychological construct, and may equally well be called moral *mis*engagement – the wrongful application of moral considerations. In our view, the problem with these attributions of moral misengagement is *not* that there is no moral right and wrong; the problem is that empirical psychology is not equipped to determine what counts as moral rights and wrongs. Rather, responsibility for determining how people ought to treat each other rests with all human agents, researchers and non-researchers alike.

A way of salvaging moral disengagement as a psychological construct is to consider it a variant of *coping*: ways of attempts to deal with the moral conflicts and doubts inherent in many multifaceted events. For instance, Haney (1997) discusses ways in which the capital punishment system makes it psychologically possible for jurors to participate in the death-sentencing process. However, we argue that such individual and collective coping does not require the disengagement or turning off of moral considerations: As we argued above, people may think it is permissible to violate one moral principle when competing moral or other considerations are deemed more important. As illustrated by Haney's work, studying how individuals and social systems handle such moral conflicts, both psychologically and socially, can yield valuable insights with implications for policy as well as psychological theory.

An ongoing line of research in our lab explores how people deal with evaluative conflicts about academic integrity. In one study, we interviewed college students about their own experiences with academic misconduct. Nearly all participants said that cheating is generally wrong. However, all participants recalled at least one time they engaged in academic misconduct. In discussing these incidents, students frequently expressed uncertainty about what counted as academic misconduct, and about half said they did not think they were doing anything wrong at the time. Among those who did think they were doing something wrong at the time, many cited desires to help fellow students, lacking sufficient time to complete the assignment, and a need to perform well on the assignment as reasons for cheating. Although these retrospective self-report data need corroboration, the preliminary findings suggest that students did not cheat because they disengaged their concerns with academic integrity, but because they tried

to cope with informational uncertainty or balancing of academic principles and perceived rights to promote their own or others' academic survival.

It is a disservice to ethical discourse as well as to empirical psychology to blur the boundaries between moral evaluations of people's actions and a psychological account of those actions. In our view, a more fruitful empirical approach to studying crime, corruption, or academic integrity would investigate how those phenomena are viewed from the point of view of the people involved. We expect that such investigations would yield very few instances in which people turn off their general moral principles and numerous instances in which morally engaged individuals try to cope with factual uncertainties, competing evaluative principles, and conflicting goals.

References

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, *31*, 101–119. https://doi.org/10.1080/0305724022014322

Bandura, A. (2015). *Moral disengagement: How people do harm and live with themselves*. New York: Macmillan Higher Education.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology*, *71*, 364.

Brooks, D. (2013, June 10). The solitary leaker. *The New York Times*. Retrieved from http://www.nytimes.com/2013/06/11/opinion/brooks-the-solitary-leaker.html

Dahl, A., & Killen, M. (in press). Moral reasoning: Theory and research in developmental science. In J. T. Wixted (Ed.), *The Stevens' handbook of experimental psychology and cognitive neuroscience, 4th ed.* New York: Wiley.

Dahl, A., Uttich, K., Gingo, M., & Turiel, E. (2013). People can reason about moral dilemmas (when researchers allow them to). Presented at the Annual meeting of the Jean Piaget Society, Chicago, IL.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, *124*, 695–726. https://doi.org/10.1086/675875

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Haney, C. (1997). Violence and the capital jury: Mechanisms of moral disengagement and the impulse to condemn to death. *Stanford Law Review*, *49*, 1447–1486. https://doi.org/10.2307/1229350

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice*. New York, NY: Cambridge University Press.

Hyde, L. W., Shaw, D. S., & Moilanen, K. L. (2010). Developmental precursors of moral

    disengagement and the role of moral disengagement in the development of antisocial

    behavior. *Journal of Abnormal Child Psychology*, *38*, 197–209.

    https://doi.org/10.1007/s10802-009-9358-5

Kohlberg, L. (1971). From is to ought: How to commit the naturalistic fallacy and get away with it

    in the study of moral development. In T. Mischel (Ed.), *Psychology and genetic*

    epistemology (pp. 151–235). New York: Academic Press.

Lee, D. D. (1985). *Sergeant York: An American hero*. Lexington, KY: University Press of

    Kentucky.

McAlister, A. L., Bandura, A., & Owen, S. V. (2006). Mechanisms of Moral Disengagement in

    Support of Military Force: The Impact of Sept. 11. *Journal of Social and Clinical*

    Psychology, *25*, 141–165. https://doi.org/10.1521/jscp.2006.25.2.141

Nucci, L. P., & Turiel, E. (2009). Capturing the complexity of moral development and education.

    Mind, Brain, and Education, *3*, 151–159.

Perkins, S. A., & Turiel, E. (2007). To lie or not to lie: To whom and under what circumstances.

    Child Development, *78*, 609–621.

Reitman, J. (2013, December 4). Snowden and Greenwald: The men who leaked the secrets.

    Rolling Stone. Retrieved from http://www.rollingstone.com/politics/news/snowden-and-

    greenwald-the-men-who-leaked-the-secrets-20131204

Turiel, E. (2003). Resistance and subversion in everyday life. *Journal of Moral Education*, *32*,

    115–130. https://doi.org/10.1080/0305724032000072906

Turiel, E., & Dahl, A. (in press). The development of domains of moral and conventional norms,

    coordination in decision-making, and the implications of social opposition. In K. Bayertz

    & N. Boughley (Eds.), *The normative animal: On the biological significance of social,*

    moral, and linguistic norms. Oxford, UK: Oxford University Press.

Turiel, E., Hildebrandt, C., & Wainryb, C. (1991). Judging social issues: Difficulties,

inconsistencies, and consistencies. *Monographs of the Society for Research in Child Development*, *56*. https://doi.org/10.2307/1166056

Wainryb, C., Brehl, B. A., & Matwin, S. (2005). Being hurt and hurting others: Children's narrative accounts and moral judgments of their own interpersonal conflicts. *Monographs of the Society for Research in Child Development*, *70*. https://doi.org/10.1111/j.1540-5834.2005.00350.x