ABSTRACT

Moral disengagement refers to a set of eight cognitive mechanisms that decouple one's internal moral standards from one's actions, facilitating engaging in unethical behavior without feeling distress. A compelling predictor of a number of morally undesirable behaviors, including childhood aggression, workplace deviance, and misconduct in sport, this review focuses on more recent research that explores how moral disengagement operates, both as a process (mediator) and as a disposition (moderator) that affects individuals' responses to morally problematic opportunities. It also speaks to central questions in moral disengagement theory, such as its malleability over time, and interventions that can be used to reduce it.

MORAL DISENGAGEMENT

Originally described by Albert Bandura, first in his seminal book on social cognitive theory (1), and later elaborated in his work focused on moral behavior (2-4), moral disengagement refers to eight interrelated cognitive mechanisms that allow us to sidestep our internalized moral standards and behave immorally without feeling attendant distress. In social cognitive theory, internal controls only work effectively when they are activated. The mechanisms of moral disengagement decouple our internal standards from how we construe our behavior, rendering them ineffective. As an example, imagine Sam has an internal standard that prohibits theft, but has taken a newspaper without paying for it from Starbucks. Moral disengagement mechanisms help Sam construe taking the newspaper as no big deal (distortion of consequences), believe that everyone takes small things like a paper sometimes (diffusion of responsibility), that taking the paper is tiny compared to others' violations (advantageous comparison), or that he’s seen Starbucks employees take copies of the paper, so why shouldn’t he (displacement of responsibility)? He could think that in the grand scheme of things, being an informed citizen is more important than paying for the paper (moral justification). He could even plan on leaving the paper in the café when he was finished with it, so really he was just “borrowing” it (euphemistic labeling). He could think that Starbucks is a large heartless corporation that won’t notice the missing paper (dehumanization), or even deserves having the paper taken from it because it charges so much for their coffee (attribution of blame). These mechanisms facilitate understanding his behavior as unrelated to his internal standard against theft. Thus, he can leave the store, paper under arm, confident in the belief that he’s done nothing wrong.

Moral disengagement theory has been fertile ground for empirical research across a number of disciplines and contexts, including child and adolescent development (5-10), organizational behavior (11-18), criminology (19, 20), military psychology (21-23), and sports psychology (24-26). Individual predispositions to
morally disengage are associated with a host of negative behaviors, including criminal behavior (19), aggression and bullying (5, 7, 8, 10), workplace misconduct (11, 13, 27), and unethical behavior generally (13, 28), as well as a host of negative psychological states, including an increased ability to dehumanize others (29-31), greater likelihood of endorsing of violence towards them (23, 32).

Recent work on moral disengagement has explored the extent to which it is stable over the life course, and relatedly the extent to which interventions can affect it. In addition, more recent research has moved beyond testing moral disengagement as a main predictor of unethical behavior, and towards understanding when it operates as a mediator and moderator of other relationships. This review focuses on these two directions in the current literature, and offers a perspective on where future work is headed.

The stability of moral disengagement over time. Though Bandura’s theory largely discusses moral disengagement as a process, empirical explorations of moral disengagement typically measure it as an individual difference (13, 33-36). These efforts have fleshed out the nomological net of dispositional moral disengagement: positively associated with Machiavellianism, trait cynicism, external locus of control and moral relativism, and negatively associated with cognitive moral development, moral identity, moral idealism, empathetic concern, guilt, as well as honesty-humility, conscientiousness, and agreeableness (13, 15, 28, 37, 38). However, consistent with social cognitive perspectives on personality as “dynamic dispositions” (39), Bandura views the self-regulation of moral conduct and the tendency to morally disengage as part of a system of “triadic reciprocal causation” (1), in which behavior, cognition, and environmental influences all operate as continuously interacting determinants of each other. This perspective opens up the possibility that one’s context can influence one’s tendency to morally disengage. Studies documenting shifting levels of moral disengagement over longer time horizons have focused on moral development over the life course (particularly adolescence), finding that, for most, moral disengagement declines during the teenage years (40). However, Paciello et al. reached slightly more nuanced conclusions using data from Italian youth from the ages of 12 and 20, finding that moral disengagement declines specifically between the ages of 14 and 16, though some individuals showed intransigently high levels of moral disengagement across the study period (9).

Other studies have examined potential triggers of these longitudinal changes. Hyde et al. examined early influences in later moral disengagement in a sample of low-income boys followed prospectively from 1.5 to 18 years of age (41). They found that poor child-parent interactions at 1.5-2 years as well as living in an impoverished neighborhood were significant predictors of moral disengagement at age 15. Another study used longitudinal social network analysis to explore the role of social influence (peer groups) in moral disengagement among schoolchildren (6). Consistent with the literature showing that peers become more important influences in early adolescence, dispositional levels of moral disengagement among one’s friends influenced one’s own moral disengagement in the subset of the sample aged 11-14, but had no effect in children aged 9-10. Together these results indicate that moral disengagement is largely a function of home environments until early adolescence, when it becomes
influenced by one’s peers, peaking around the age of 14 before dropping again—for most, but not all.

**Intervening in the moral disengagement process.** There is growing interest in understanding the process of moral disengagement (42-44). The question of whether moral disengagement is pliable over shorter time horizons can be answered by studies focused on interventions or situational characteristics that either amplify or dampen moral disengagement tendencies.

*Triggering moral disengagement.* There is still little empirical evidence showing how moral disengagement is initiated. Without this evidence, there are nagging doubts about how moral disengagement functions as a process rather than as a disposition (45). In the best tests to date of moral disengagement processes, Shu et al. found that individuals who cheated showed higher levels of moral disengagement, as well an impaired ability to remember moral rules (43). Gino and Galinsky extended this finding to show that priming someone to feel psychologically close to someone who cheated increased moral disengagement about cheating (44). These studies provide the first evidence of moral disengagement as a motivated cognitive process: when it was in the participants’ interest to disengage from an internalized standard against cheating (either because they had cheated themselves, or felt close to someone who did), they were more likely to do so. In related work, Paharia et al. also found that morally disengaged reasoning resulted from a motivated process, finding that people were more likely to endorse moral justifications for poor labor practices when contemplating desirable goods or services that used them, compared to when they were contemplating the same good or service made under more favorable working conditions (46).

*Reducing moral disengagement.* Studies testing ways to reduce moral disengagement began in pedagogical contexts. For example, McAlister found that simply outlining the processes of moral disengagement reduced individuals’ tendencies to disengage (33). More recently, Bustamante and Chaux found that a critical thinking intervention reduced levels of moral disengagement in ninth grade students (47). These efforts suggest promising pedagogical avenues to reduce moral disengagement. Another set of studies has examined potential organizational interventions. Barsky (27) found that higher levels of participation in setting performance goals at work made individuals less likely to morally justify or displace responsibility (two moral disengagement mechanisms). Kish-Gephart et al. found that highlighting the harm that highly self-interested behavior would cause decreased the likelihood that individuals would morally disengage (42). Hodge and Lonsdale found that sports coaches who supervised athletes in a controlling way elicited higher levels of moral disengagement among their athletes, increasing later antisocial behaviors towards their teammates and opponents, but that supportive coaching elicited less moral disengagement among their players (25).

*Amplifying moral disengagement.* A third set of studies stress how psychological states affect moral disengagement. Chugh and colleagues (48) found that individuals were more susceptible to the negative behavioral consequences of moral disengagement when they had been primed feel anxious and insecure (vs. supported and secure). Similarly, Paciello et al. found that feelings of personal distress elicit moral disengagement (49). Worryingly, Waytz and Epley (31) found
that priming individuals to think about their social connections with others enabled
dehumanization. They theorize that priming social connectedness satisfies the human
motivation to connect, leaving room to think of others (particularly socially distal
others) as less human, thus endorsing their mistreatment. This result is consistent with
Gino and Galinsky’s finding that psychological closeness with someone who behaves
unethically also amplifies moral disengagement (44), and suggests we need to be
careful about how an individual’s immediate social context affects the extent to which
(and about whom) we morally disengage.

Some of the most interesting work on the mutability of moral disengagement
explores the role of video games in increasing moral disengagement or amplifying its
negative effects. This research shows that both the frequency and recency of one’s
exposure to violent video games (such as Grand Theft Auto) is associated with higher
levels of moral disengagement (50). In addition, high base rates of moral
disengagement amplify the extent to which, after playing such games, participants
later show worse self-control, as well as higher levels of cheating and aggression (51).
Manipulating moral disengagement cues within a violent video game (making the
targets zombies as a cue to dehumanization, or changing the cover story of the game
[fighting for the UN attacking a torture camp vs. fighting to protect a torture camp] as
a cue to moral justification), affected later guilt and enjoyment of the game. When
virtual violence was framed as morally justified, participants felt less guilt and
experienced fewer negative emotions after playing, suggesting that the availability of
cues to morally disengage influences individuals’ affective reactions to violent
behavior (52).

Mediator or moderator? The tension between understanding moral
disengagement as a process or disposition has implications for how it is tested
empirically. If moral disengagement is a process, it should be studied as a mediator. If
moral disengagement is a trait, it should be studied as a moderator. Both approaches
have met with success.

Moral disengagement as a mediator. A number of studies have explored moral
disengagement as mediating the effects of individual-level predictors on morally
problematic outcomes, both cognitive and behavioral. In Shu’s studies, moral
disengagement mediated the relationship between cheating and forgetting moral rules
(43). Leidner et al. found that moral disengagement mediated the relationship between
glorifying one’s ingroup and lesser demands for justice for those mistreated in the
Iraq war (30), echoing McAlister’s earlier findings that support for retaliatory strikes
after the September 11th terrorist attacks was mediated by moral disengagement (23).
Duffy et al. found that envy predicted social undermining behavior through moral
disengagement, in two multi-wave studies of hospital employees and student teams
(11), and Paciello et al. found that moral disengagement elicited by personal distress
allows individuals to absolve themselves of responsibility towards others in need (49).

Several studies have explored how role models influence negative behavioral
outcomes through moral disengagement. In Hyde’s longitudinal study of low-income
youth, the moral disengagement of study subjects at 15 years old mediated the
relationship between the experience of poor parenting at 1.5-2 years and adolescent
antisocial behavior at 16 and 17 years old (41). Hodge and Lonsdale found that moral
disengagement mediated the relationship between controlling coaching styles and
higher levels of anti-social behavior towards teammates and opponents (24). Moore et al. have similarly found that moral disengagement mediates the relationship between how ethical an employee’s leader is and the likelihood they will engage in unethical workplace behavior (53). Moral disengagement has also been studied as a mediator in the relationship between positive ethical antecedents and outcomes. For example, Ogunfowora and Bourdage found that individuals with high levels of the personality dimension honesty-humility were more likely to emerge as leaders in student groups, through lower levels of moral disengagement (38).

Some studies have found that mediating relationships involving moral disengagement are more nuanced, and depend on a third factor. In their study of youth offenders, DeLisi et al. found that the relationship between psychopathy and criminal behavior was direct for delinquents who showed high levels of psychopathy, but was mediated through moral disengagement for youths who had lower levels of psychopathy (20). Relatedly, Moore et al. found that the indirect relationship between ethical leaders and employee misconduct was moderated by employees’ moral identities: whether ethical leaders inspired more ethical behavior or unethical leaders encouraged more deviant behavior through moral disengagement depended on how important being moral was to the employee in the first place (53). These studies hint at the complex interactive processes that combine to produce our moral behavior: a function of who we are when we enter a given context, as well as how that context affects us.

Moral disengagement as a moderator. Dispositional moral disengagement seems to function as an accelerant in ethically dangerous circumstances: in facilitative contexts, high dispositional moral disengagement will amplify unethical behavior. For example, Samnani et al. found that the relationship between negative affect and counterproductive workplace behavior was stronger for those higher in dispositional moral disengagement (18). White-Ajmani and Bursik found that after being insulted, the likelihood of harming that individual (by making them drink hot sauce) was more likely among those high in moral disengagement (54). And Panasiti and colleagues (55) found that moral disengagement did not trigger increases in rates of lying in an experimental game across the board, but individuals who were more morally disengaged were less affected by reputational risks when making a decision to lie.

Dispositional moral disengagement may also make individuals less susceptible to positive moral influences. For example, Bonner et al. found that individuals who were high in moral disengagement were more immune to the effects of ethical leaders—only employees who were low in moral disengagement themselves were positively influenced by how ethical their leader was (56). Relatedly, moderation may take the opposite form: high levels of moral disengagement may predict unethical behavior directly, leaving only those low in moral disengagement to be negatively affected by contextual factors. For example, Kouchaki and Smith found that individuals low in moral disengagement were more susceptible to self-control failures when they were tired later in the day—individuals high in moral disengagement had lower levels of self-control all day long (57).

Together, these results suggest that moral disengagement both feeds into the process of unethical behavior in the moment, as well as influence how other factors may lead to unethical (and ethical) behavior.
Concluding Thoughts

The empirical evidence on moral disengagement amassed over the past five years suggests that it can be understood both as a relatively stable cognitive orientation (though one which is pliable to the influence of one’s context) as well as a state triggered by more immediate contextual factors. Yet understanding exactly how moral disengagement operates in the moment remains elusive, much to some researchers’ chagrin (45). Future work should explore how moral disengagement colors how we construe morally meaningful choices, and whether it follows motivated cognitive processing. Brain imaging (fMRI) studies may be necessary to test these possibilities. More thought should also be given to when it might be more fruitful to study individual mechanisms of moral disengagement, particularly to answer more specific theoretical questions—as in the work connecting social connection (31) and the valorization of in-groups (30) to dehumanization.

In terms of the stability of moral disengagement as a disposition, we know some about the trajectories of moral disengagement tendencies during the course of adolescence, but no studies to date have examined how moral disengagement may shift during adulthood. Where correlations between moral disengagement and age have been reported in adult samples, they are significant and negative (11, 45), suggesting that moral disengagement likely declines beyond adolescence as we age. Future research would be useful in this regard.

Finally, it would also be valuable to enrich our understanding of how moral disengagement may be related to outcomes beyond immoral or antisocial behaviors. What adaptive purposes might moral disengagement hold? How might it lead to benefits for individuals, such as leadership emergence (38), career advancement (58), or social status (59)? One recent paper found that moral disengagement buffered employees from the negative motivational consequences of unethical requests from their work colleagues (60). Yet, thus far, the literature has only hinted at the potential benefits of moral disengagement. Only when we understand the positive consequences that moral disengagement holds for people will we be able to truly neutralize its damaging effects.


*Finds MD plays a larger factor in criminal activity among individuals lower in psychopathic traits.


22. Beu DS, Buckley MR: This is war: How the politically astute achieve crimes of obedience through the use of moral disengagement. *The Leadership Quarterly* 2004, **15**: 551-568 (http://dx.doi.org/10.1016/j.leaqua.2004.05.007).


25. Hodge K, Lonsdale C: Prosocial and antisocial behavior in sport: The role of coaching style, autonomous vs. controlled motivation, and moral disengagement. *Journal of Sport and Exercise Psychology* 2011, **33**.


*Documents how feeling socially connected facilitates the MD mechanism of dehumanization.*


*Documents how making the harm one causes salient reduces the likelihood of morally disengaging.*


*Shows MD can be motivated by something as minimal as feeling psychologically close to a wrongdoer.


*Critiques whether MD can be considered a cognitive process rather than a trait or disposition.

46. Paharia N, Vohs KD, Deshpandé R: Sweatshop labor is wrong unless the shoes are cute: Cognition can both help and hurt moral motivated reasoning. *Organizational Behavior and Human Decision Processes* 2013, 121: 81-88 (http://dx.doi.org/10.1016/j.obhdp.2013.01.001).


*Shows individuals are more likely to resist MD when they feel socially secure rather than anxious.


*MD amplifies unethical behavior with situational triggers of aggression (violent video games).


