



Reconceptualizing Moral Disengagement as a Process: Transcending Overly Liberal and Overly Conservative Practice in the Field

Ulf Schaefer¹ · Onno Bouwmeester²

Received: 30 November 2017 / Accepted: 14 April 2020 / Published online: 28 April 2020
© The Author(s) 2020

Abstract

Moral disengagement was initially conceptualized as a process through which people reconstrue unethical behaviors, with the effect of deactivating self-sanctions and thereby clearing the way for ethical transgressions. Our article challenges how researchers now conceptualize moral disengagement. The current literature is overly liberal, in that it mixes two related but distinct constructs—process moral disengagement and the propensity to morally disengage—creating ambiguity in the findings. It is overly conservative, as it adopts a challengeable classification scheme of “four points in moral self-regulation” and perpetuates defining moral disengagement via a set of eight psychological mechanisms, narrowing our understanding of the phenomenon. To address these problems, we propose to define process moral disengagement intensionally (specifying the necessary and sufficient conditions for correct application of the term) as intrapsychic cognitive reasoning processes through which people selectively reconstrue a moral judgment “behavior B by actor A is morally wrong” and shift it toward becoming “behavior B is not morally wrong” or “actor A is not responsible for behavior B.” This definition achieves disambiguation and increased concept clarity. We leverage the definition to motivate a classification scheme for psychological mechanisms of moral disengagement along two dimensions—reconstruing morality and reconstruing agency—and to initiate an open inventory of psychological mechanisms that specify how process moral disengagement operates.

Keywords Moral disengagement · Process moral disengagement · Ethical decision making · Inventory of process moral disengagement mechanisms

Introduction

Moral disengagement theory (e.g., Bandura 1990a) offers a compelling yet parsimonious explanation for how people transgress. The theory proposes that people are able to reconstrue moral judgments through intra-psychological processes, with the result that they “disassociate with their

moral standards” (Newman et al. 2019). Through moral disengagement, individuals absolve themselves of what they would evaluate as immoral conduct in the absence of morally disengaged cognition. Beyond the individual level, scholars have asserted moral disengagement at the group and organizational levels (Bandura 1999; Bandura et al. 2000; Gini et al. 2015; White et al. 2009) and have labeled it “the root cause of sanctioned corporate corruption” (Brief et al. 2001, p. 473). By now, moral disengagement has become a frequently adopted perspective to understand unethical behavior in the corporate world (e.g., Ashforth and Anand 2003; Bandura 2016; Moore 2008; for an overview see Newman et al. 2019, p. 2).

First formulated by Bandura in 1973 (pp. 210–221), the scholarly literature on moral disengagement has rapidly expanded only in the last decade, and it is in this decade that the term “moral disengagement” has found broader use in the business ethics literature. A recent review of this literature (Newman et al. 2019) provides an extensive overview of the antecedents, moderators, and outcomes of moral

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10551-020-04520-6>) contains supplementary material, which is available to authorized users.

✉ Ulf Schaefer
ulf.schaefer@sdabocconi.it

Onno Bouwmeester
o.bouwmeester@vu.nl

¹ SDA Bocconi School of Management, Via Sarfatti 10, 20136 Milan, Italy

² Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081RH Amsterdam, The Netherlands

disengagement, yet it concludes that the conceptualization of moral disengagement remains an area of continued debate.

Complementing the review from Newman et al., our article critically assesses how researchers currently conceptualize moral disengagement. We find that current scholarly practice is overly liberal, in that it not only accepts denoting two related but distinct conceptualizations with the same term, “moral disengagement”, but also intermixes the two constructs process moral disengagement and the propensity to morally disengage—as if they were identical. The two concepts are related, as both rely on a set of eight specific mechanisms of moral disengagement curated by moral disengagement pioneer Albert Bandura (e.g., 1973, 1999, 2016). We argue, however, that process moral disengagement (a state-like construct) and the propensity to morally disengage (a trait-like construct) are ontologically different, and thus denote distinct conceptualizations. The distinction has profound consequences on the narrative that moral disengagement supports. Process moral disengagement explains how normal and “morally healthy” people can behave unethically; the propensity to morally disengage, in contrast, supports “bad apple” narratives. As a result of a lack of disambiguation, the current literature is in large part ambiguous, and the field needs concept clarity.

Our assessment of the conceptualization of the current scholarly practice also finds that the field is overly conservative, as it retains a debatable classification scheme of moral disengagement mechanisms and perpetuates defining the key construct with reference to a list of eight mechanisms of moral disengagement curated by Bandura. In his writings, Bandura asserts a four-point structure (referred to as the behavioral process, the self-regulatory process, or moral self-regulation; Bandura 1977, 1990a, 1999) to classify mechanisms of moral disengagement. We argue that the structure is problematic, as it is vague and has invited misinterpretations. Furthermore, Bandura defined moral disengagement extensionally by reference to a set of eight specific mechanisms (e.g., Bandura 1990a, 1999, 2016). Extensional definitions, that is, definitions that provide “a list of those instances in which the expression being defined is applicable” (Cook 2009, p. 155), are rather vivid and intuitively accessible. However, extensional definitions are unspecific about the essential attributes that an entity must have in order to be considered part of the set captured by the definition, and they suffer from an inability to claim completeness. Bandura’s set of eight mechanisms has become the all-dominant reference frame to conceptualize moral disengagement. Observing that early scholars of moral disengagement did in fact suggest further moral disengagement mechanisms, we argue that the field needs a reconceptualization of process moral disengagement to allow for a more comprehensive, richer understanding of the core construct.

To address the ambiguity in the literature and to transcend the dominance of the restrictive extensional definition via eight specific mechanisms offered by Bandura (e.g., 1999), we present an *intensional definition* that provides the meaning of “process moral disengagement” by specifying the necessary and collectively sufficient conditions for the correct application of the term (Cook 2009). This reconceptualization defines process moral disengagement as intrapsychic cognitive reasoning processes through which people selectively reconstrue a moral judgment “behavior B by actor A is morally wrong” and shift it toward becoming “behavior B is not morally wrong” or “actor A is not responsible for behavior B.”

From this intensional definition, which establishes two strategies of moral disengagement—reconstruing morality and reconstruing agency—we derive a two-dimensional model that allows for classifying psychological mechanisms of moral disengagement. These mechanisms capture the variety of psychological processes through which moral disengagement operates. We discuss 12 specific psychological mechanisms of process moral disengagement proposed by Bandura and other early scholars of moral disengagement (Ashforth and Anand 2003; Haney 1997; Tsang 2002). Classifying them as reconstruing morality mechanisms, reconstruing agency mechanisms, or dual-function mechanisms (i.e., mechanisms that may serve either of the strategies), we initiate the development of an open inventory that may be augmented by other scholars.

Reconceptualizing process moral disengagement, we make two contributions to the literature. First, our paper contributes to increased concept clarity. We show the need for disambiguation between two related but distinct concepts, and we propose that scholars use two labels—“process moral disengagement” and “propensity to morally disengage”—in their research to unequivocally declare what they are studying to avoid ambiguity and misinterpretation of results, as well as internal inconsistencies between conceptualization, theory formulation, and operationalization. We provide an intensional definition of process moral disengagement that distinguishes process moral disengagement from other constructs and specifies the conditions for the correct application of the term. We further contribute to increased concept clarity by providing a classification scheme for psychological mechanisms of moral disengagement that is derived from the intensional definition and better accommodates these mechanisms, as compared to existent classifications.

Second, we offer a new taxonomy to the literature on moral disengagement: the open inventory of psychological mechanisms of moral disengagement. Current conceptualizations of moral disengagement suffer from an overly narrow focus on only eight specific mechanisms of moral disengagement. Leveraging the two-dimensional classification scheme derived from the intensional definition, we initiate

the establishment of an open inventory to help in developing a broader, and potentially richer, understanding in the literature of how moral disengagement processes operate. To start this development, we discuss 12 psychological mechanisms of moral disengagement; more importantly, we emphasize that the inventory is open for elaboration by future researchers. Psychological processes that meet the conditions specified in the intensional definition may be added and, over time, will provide an increasingly complete understanding of the plurality of psychological processes through which the human mind morally disengages unethical behavior.

In the following sections, we first briefly summarize the initial conceptualization of moral disengagement by its pioneer, Albert Bandura, to establish an anchor point. We then examine current conceptualizations of moral disengagement in business ethics research and show that current practice is both overly liberal and overly conservative. We then address these problems by proposing a new intensional definition of moral disengagement and leveraging the definition to develop a two-dimensional model. This model provides a classification scheme that categorizes mechanisms of process moral disengagement beyond the mechanisms suggested by Bandura. We discuss theoretical contributions and suggest directions for future research before we offer a conclusion.

Bandura's Conceptualization of Moral Disengagement

Bandura introduced moral disengagement theory (1990a, 1999) as part of his social learning theory (1971, 1977), which he later developed to become social cognitive theory (1986, 1991, 2018a). Bandura first used the term “neutralization of self-condemnation for aggression” (1973, p. 210), specifying eight “self-absolving practices” through which neutralization operates. In his presentation of social learning theory, he further developed the terminology and introduced a self-regulatory process of internal moral control to present the eight “[m]echanisms through which behavior is disengaged from self-evaluative consequences at different points in the behavioral process” (Bandura 1977, p. 156).

Bandura only later introduced and used the term “moral disengagement” (1990a, 1991, 1999, 2016) to denote his central idea: People generally act in line with (and refrain from conduct contrary to) internalized moral standards (e.g., Bandura 1990a, 2004). Internalized moral standards, however, can be morally disengaged through eight cognitive mechanisms that change people's assessment of the behaviors. Behaviors otherwise evaluated as unethical¹ can become reevaluated as justifiable or even obligatory, or as behavior for which the agent is not responsible. Through *moral justification*, *euphemistic labeling*, and *advantageous*

comparison, individuals can effectively reconstrue moral conduct, rendering otherwise immoral behavior as permissible, or even obligatory. Through the *displacement of responsibility* and the *diffusion of responsibility*, individuals effectively distance themselves from agency. By *minimizing, ignoring, or misconstruing consequences*, individuals dispute the harmful effects of their actions. Through the *dehumanization* of victims and by *blaming the victim*, the blame for own behavior is shifted to the target of the behavior.

Moral disengagement theory shares themes of cognitive dissonance reduction with dissonance theory (Festinger 1954), regarding an individual's need to resolve the conflict between one's own self-affirmation as a good person and one's own unethical behavior. Moral disengagement theory evolved from neutralization theory (Sykes and Matza 1957), as evidenced by its first label, “neutralization of self-condemnation for aggression,” and the adoption of mechanisms, for example, the denial of responsibility, or the denial of injury. Furthermore, moral disengagement theory was influenced by the work done on deference to authorities (Milgram 1963) and the dehumanization of victims (Kelman 1973).

In 45 years of theorizing and empirically investigating the key concept, Bandura (1973, 2018a) moderately changed the terminology to denote the main construct, the self-regulatory process, and the individual mechanisms. Bandura, however, consistently and firmly (e.g., 2018b) asserted the core of the concept: Moral disengagement is a *process*, it operates through a very specific set of *eight mechanisms*, which in turn operate at *four points* (loci) in the self-regulatory process.

Scholars are to be credited for the ideas they contribute; however, they do not own their ideas. Other scholars may adopt and adapt ideas. At times though, scholars are overly liberal in adapting an idea; at other times, they are overly conservative in adopting an idea. In the following section, we argue that moral disengagement has met both fates.

Problems with Current Conceptualizations

Overly Liberal: Mixing Related but Distinct Notions of Moral Disengagement

Reviewing measures of moral disengagement in the organizational literature, Newman et al. (2019) noted a “lack of consistency over how moral disengagement and its associated mechanisms have been studied” (p. 24). We argue that the lack of consistency does not enter at the level of measurement; it is rooted in different conceptualizations

¹ Adopting a convention in the field, we use the terms “ethical” and “unethical” synonymously with “moral” and “immoral,” respectively.

of moral disengagement. Scholars invoke two related yet psychologically distinct constructs, referring to both using the term “moral disengagement.” Bandura specified moral disengagement as a process, yet recently scholars have often studied moral disengagement as an individual difference. This conceptual ambiguity translates into inconsistencies in theorizing and operationalizing the construct, thereby threatening advancement of the field.

Bandura’s notion of moral disengagement refers to the process of reconstruing moral judgments (via the application of moral disengagement mechanisms), with the result that the moral evaluation of a behavior is changed. This process operates selectively, that is, it is specific to a behavior or type of behavior that is morally disengaged. When people selectively morally disengage a specific moral standard, “in other aspects of their lives, they adhere to their moral standards” (Bandura 2016, p. 3).

In contrast, the propensity to morally disengage—also referred to as dispositional moral disengagement—is a trait-like construct asserting that “[s]ome people will be more predisposed to moral disengagement than others” (Detert et al. 2008, p. 374f.). As a trait, it does not operate selectively, but it represents a general and rather stable entity. Detert et al. (2008) provided a widely used measure of the propensity to morally disengage. Building on this work, Moore et al. (2012) validated a measure of the general (context- and population-independent) propensity to morally disengage that has established itself as a second, widely used tool in empirical research on moral disengagement.

Although the two variables are related—most notably evidenced by the frequent reference to the same set of eight mechanisms of moral disengagement asserted by Bandura—they are, however, psychologically different constructs. Process moral disengagement can be best captured as a state variable: “[S]tates are conceptualized as temporary conditions of mentality or mood, transitory levels of arousal or drive, and currently evoked activity or process” (Messick 1987, p. 6). In contrast, the propensity to morally disengage (dispositional moral disengagement) presents a trait variable, a “relatively stable characteristic of a person—an attribute, enduring process, or disposition—which is consistently manifested to some degree when relevant, despite considerable variation in the range of settings and circumstances” (Messick 1987, p. 5).

Scholars often ignore the difference between process moral disengagement and the propensity to morally disengage. At times, they have even misattributed the latter to Bandura. To illustrate, Christian and Ellis state that “following the work of Bandura and others (e.g., Moore et al. 2012), we are primarily interested in the trait instantiation of the construct, which assumes that individuals exhibit dispositional differences in the expression of moral disengagement” (2014, p. 194), and Bonner et al. claim that “Bandura’s work

argues for a trait approach to moral disengagement” (2016, p. 733). Such misrepresentation of Bandura’s conceptualization—possibly caused by Bandura’s use of psychological jargon, which refers to individuals who morally disengage as high “moral disengagers” (Bandura et al. 1996, p. 368), and his use of the terms “proneness to moral disengagement” (Bandura et al. 1996, p. 367) and “facility for moral disengagement” (Bandura et al. 1996, p. 372)—may have recently prompted Bandura to reclaim ownership over the term’s interpretation: “Moral disengagement is not a character trait that is assessed by a one-size-fits-all measure.[...] Some studies are being published with faulty trait measures” (Bandura 2018a, p. 133).

Concept clarity demands that researchers do not use the same label for conceptually different constructs. Furthermore, the distinction has relevance for theorizing and consequences for study design, for example when moral disengagement constructs are invoked as a third variable to understand the relationship between an independent and a dependent variable; in these cases, “[i]f moral disengagement is process, it should be studied as a mediator. If moral disengagement is a trait, it should be studied as a moderator” (Moore 2015, p. 201).

We argue that a lack of disambiguation results in significant confusion in the literature. We analyzed the organizational literature on moral disengagement, reviewing the sample of the business ethics literature selected by Newman et al. (2019). Our analysis—reported in the Supplementary Material—found 16 of 56 studies to be ambiguous. We identified 6 conceptual and 21 empirical papers that consistently conceptualized and operationalized moral disengagement as process. Moreover, we identified 2 conceptual and 11 empirical papers that consistently conceptualized and operationalized moral disengagement as propensity to morally disengage. Sixteen papers (29% of the sample), however, could not clearly be classified into one of these literatures for different reasons. Eight papers—Baron, et al. (2015); Chen et al. (2016); Detert et al. (2008); He et al. (2019); Lee et al. (2016, 2019); Ogunfowora and Bourdage (2014); Zheng et al. (2019)—conceptualized moral disengagement as a process, studying it as a mediator, yet used trait measures in their correlational research. Although the use of trait measures (e.g., Detert et al. 2008; Moore et al. 2012) in experimental designs may be adequate to study process moral disengagement—as exemplified by Palmer (2013) as well as Valle et al. (2019)—they are insufficient for measuring process moral disengagement in correlational research. Four papers (Beaudoin et al. 2015; Knoll et al. 2016; Ogunfowora et al. 2013; Tasa and Bell 2017) conceptualize moral disengagement as a propensity (as an individual difference), yet they discuss models where moral disengagement is a mediator, which is confusing

because “[t]heoretically and conceptually, a mediator should be a responsive variable that changes within a person” (Wu and Zumbo 2008, p. 373). Conversely, Ntayi et al. (2010) conceptualize moral disengagement as a process. Their theoretical model, however, presents it as a moderator variable, which is confusing because a moderator is typically characterized as “an innate attribute,” “a relatively stable *trait*,” or “a relatively unchangeable background, environmental or contextual variable” (Wu and Zumbo 2008, p. 379). Three papers (Keem et al. 2018; Moore et al. 2019; Reynolds et al. 2014), present multiple studies in which they provide different conceptualizations. To illustrate, Reynolds et al. (2014) in four studies conceptualize moral disengagement as propensity while studying it as a mediator; only in one study do the authors conceptualize the focal construct as process, which is consistent with studying the variable as a mediator.

In sum, we find that more than a quarter of the business ethics review sample of Newman et al. (2019) consists of studies that provide ambiguous results due to a discrepancy between the conceptualization of the construct, the theoretical model specification, and the operationalization of the key variable.

We argue that the study of either construct—process moral disengagement or the propensity to morally disengage—may be a promising and fruitful endeavor; yet, we believe that there is a strong need for disambiguation. Because process moral disengagement and the propensity to morally disengage are different constructs, scholars need to indicate clearly which conceptualization they are invoking (and ideally, additionally indicate which conceptualization they consequently are not invoking) and choose their theoretical model and variable operationalization accordingly. When reviewing the literature, scholars may need to separate the findings; the field’s current practice of mixing the two constructs—even when the distinction has been noted (e.g., Reynolds et al. 2014; Moore et al. 2019; Newman et al. 2019)—is overly liberal and threatens advancement in the field.

The disambiguation of the two notions not only is mandated by methodological principles, but it also has profound consequences on the narrative that moral disengagement supports. Process moral disengagement helps us to understand how people do harm and live with themselves (Bandura 2016), how psychologically normal employees and executives become white-collar criminals (Coleman 1987), how good people become evil (Zimbardo 2007), and even how normal, “ordinary men” can become mass murderers (Browning 1992; Welzer 2007). The propensity to morally disengage, in contrast, supports accounts of “bad apple” narratives, that is, the claim that individual traits may explain why some people engage more easily in unethical behavior than others.

Too conservative: Holding on to eight mechanisms operating at four loci

Newman et al. (2019) noted that Bandura’s set of eight mechanisms overlaps significantly with mechanisms presented by criminal sociologists Sykes and Matza (1957), and Kish-Gephart et al. (2014) argued that “Bandura effectively cohered under one theoretical umbrella a number of rationalization and neutralization mechanisms discussed and studied by others” (p. 268). Though the terminology to denote the concept of moral disengagement and individual mechanisms showed some variation, Bandura (Bandura 1973, 1977, 1986, 1990a, b, 1991, 1999, 2002, 2004, 2016, 2018b; Bandura et al. 1996) consistently referred to a set of eight specific mechanisms in defining moral disengagement, as illustrated by the following two versions (the first specifying our focal construct for the first instance, and then refers to it as the “neutralization of self-condemnation for aggression”).

By engaging in self-deceptive machinations, [people] can behave inhumanely without self-condemnation. The different forms that self-absolving practices take are [...] Slighting aggression by advantageous comparison [...] Justification of aggression in terms of higher principles [...] Displacement of responsibility [...] Diffusion of responsibility [...] Dehumanization of victims [...] Attribution of blame to victims [...] Graduated desensitization [...] Minimization and selective forgetting of consequences. (Bandura 1973, pp. 210–217)

Moral control can be selectively activated or disengaged from inhumane conduct. Self-sanctions can be disengaged by reconstruing detrimental conduct through moral justification, euphemistic labeling, and advantageous contrast with other inhumanities; by obscuring personal agency in detrimental activities through diffusion and displacement of responsibility; by disregarding or misrepresenting the harmful consequences of inhumane conduct; and by blaming and dehumanizing the victims. (Bandura 1990a, p. 27)

Mechanisms that go beyond the eight that are consistently referred to by Bandura were suggested relatively early on by Haney (1997), who explored the capital jury system in the United States; by Tsang (2002), who investigated rationalization strategies in contexts such as the Holocaust; and by Ashforth and Anand (2003), who investigated the emergence and spread of corrupt practices in organizations. In recent studies of moral disengagement (both in its process or propensity variety), however, researchers have conservatively adopted Bandura’s set of eight specific mechanisms presented above, either as a complete set (frequently) or as a

subset (less often). All studies sampled for the business ethics literature review by Newman et al. (2019) either referred to the set of eight mechanisms listed by Bandura or a subset thereof (e.g., Barsky 2011; Kish-Gephart et al. 2014; Welsh et al. 2015; Petitta et al. 2017). The confinement to the set of eight mechanisms may suggest that Bandura's concept has "stood the test of time" (Newman et al. 2019, p. 6). However, we suspect that alternative explanations for the dominance of the eight mechanisms are more plausible: Scholars may be awestruck by the "stroke of genius" that Bandura's moral disengagement theory represents; reference to Bandura's mechanisms may have become traditionalized standard practice; or researchers may be avoiding the risk of working with different mechanisms, given the dominance of adhering to the set of eight in the scholarly community.

We argue that the field is overly conservative in perpetuating the set of eight mechanisms proposed by Bandura, and we find evidence for overly unreflective practice in the parallel adoption of the categorization of mechanisms along a four-point structure repeatedly affirmed by Bandura.

Bandura asserted that the eight mechanisms of moral disengagement fall into four classes that are based on where they operate in what he calls "the behavioral process" (Bandura 1977, p. 156), alternatively termed "the regulatory process" (Bandura 1990a, p. 28), "the self-regulatory process" (Bandura 1999, p. 194), and "moral self-regulation" (Bandura 2016, p. 3). This "process" consists of *three* steps, "REPREHENSIBLE CONDUCT → DETRIMENTAL EFFECTS → VICTIM" (e.g., Bandura 1986, 1999), alternatively presented as "DETRIMENTAL PRACTICES → INJURIOUS EFFECTS → VICTIM" (e.g., Bandura 2016, 2018b), yet Bandura states that moral self-sanctions are disengaged at *four* "points," alternatively called "loci." Three points (loci) result from direct connections to the three process steps: the "behavior locus" is linked to "reprehensible conduct," the "outcome locus" is linked to "detrimental effects," and the "victim locus" is linked to "victim." A fourth locus, the "agency locus," however, connects to two steps "reprehensible conduct" and "detrimental effects." Bandura's process raises numerous questions, and we argue that it may not provide an ideal basis to classify moral disengagement mechanisms.

The process model itself appears vague and ambiguous. The "process" appends a static target construct (victim) to the preceding causal sequence. It is thus unclear what the process sequence means, inviting misinterpretations such as "moral disengagement is regarded as a linear process in which one can only achieve the last stage of moral disengagement (victim dehumanization and victim blaming) by progressing through the other three loci" (Newman et al. 2019, p. 5).

Furthermore, the meaning of the links between loci and process steps remains unclear. If the dual link between,

for example, the displacement of responsibility and both reprehensible conduct and detrimental effects is to be interpreted as "people can displace responsibility for both their conduct *and* the outcomes of their conduct," then the question arises as to why euphemistic labeling points only to reprehensible conduct, but not to detrimental effects as well. Surely, the widely criticized euphemism "collateral damage" specifies detrimental effects. If the links are to be interpreted differently, then that interpretation remains opaque. Studiously scanning all of Bandura's works on moral agency, we could not locate an explication or a derivation of the process, nor a guide on how to interpret the links between loci and process steps.

As a result, there is considerable heterogeneity with regard to how scholars interpret and reproduce Bandura's categorization. Whereas most researchers who categorize moral disengagement mechanisms reproduce Bandura's four-locus model (e.g., Beu and Buckley 2004; Eriksson and Svensson 2016; Newman et al. 2019), many others provide deviating triad-categorizations. In such cases, triads result from merging two of Bandura's four loci: A first group merges outcome locus and victim locus (e.g., Moore 2008; Baron et al. 2015; Tasa and Bell 2017). A second group merges agency locus and outcome locus (e.g., Duffy et al. 2012; Christian and Ellis 2014; Kish-Gephart et al. 2014; Johnson and Buckley 2015; Chen et al. 2016; Bonner et al. 2016). A third triad variant merges behavior locus and outcome locus (Connelly et al. 2016). In deviating from Bandura's four-locus structure, scholars not only omit motivating their deviation, but they also often claim that they are adopting the structure provided by Bandura.

In conclusion, we find evidence that the current scholarly literature on moral disengagement is overly conservative in adopting and perpetuating a conceptualization of process moral disengagement rooted in extensional definitions of the construct affirmed by Bandura. Early attempts to include further psychological mechanisms under the umbrella of moral disengagement have not been taken up by other scholars. Instead, and despite the fact that Bandura never explicitly excluded the existence of further mechanisms, the conceptualization of moral disengagement as eight specific mechanisms and the classification of mechanisms along a four-point process has solidified as a result of continuous perpetuation. In the next section, we attempt a reconceptualization of process moral disengagement—consistent with the idea proposed by Bandura—via an intensional definition, which also suggests a classification structure for a new open inventory of moral disengagement mechanisms.

Reconceptualizing Process Moral Disengagement

An Intensional Definition of Process Moral Disengagement

Dahl and Waltzer alleged that Bandura “does not provide an explicit definition of moral disengagement” (2018, p. 241). What they might have meant is that Bandura never provided what philosophical logic refers to as an *intensional* definition (Cook 2009; Podsakoff et al. 2016), that is, a definition that gives the meaning of a term by specifying the necessary and sufficient conditions for correct application of the term. Such definitions focus on the intension of a concept, that is, on the internal content of a concept. Instead of an intensional definition, Bandura provided several versions of extensional definitions, as illustrated by the two examples presented in the last section. An extensional definition provides “a list of those instances in which the expression being defined is applicable” (Cook 2009, p. 155; cf. also the concept of family resemblance in Podsakoff et al. 2016). Extensional definitions focus on the extension of a concept, on the range of applicability for the concept, and on naming the particular instances that the concept denotes.

In specifying a construct—a “cognitive symbol that has meaning for the scientific community that uses it” (Podsakoff et al. 2016, p. 161)—extensional definitions are pragmatic and justifiable in principle. Especially in the early days of developing a psychological theory, more vivid extensional definitions may have advantages compared to rather abstract intensional definitions because they convey rather directly and intuitively the common fundamental attributes. Extensional definitions, however, may provide a challenge to concept clarity because “there is no way to know whether the exemplars provide a complete listing of the concept’s domain and/or whether new exemplars should be excluded from the concept’s domain without a clear statement of the concept’s underlying theme that ties the exemplars together” (Podsakoff et al. 2016, p. 185).

We propose to define process moral disengagement intensionally, decoupled from reference to specific mechanisms. To do so, we specify six key attributes—fundamental characteristics—of process moral disengagement that are individually necessary and collectively sufficient to define process moral disengagement.

First and most generally, process moral disengagement denotes *intrapyschic cognitive processes*, that is, processes that take place in people’s minds. Note that this attribute is sufficient to distinguish process moral disengagement from the conceptualization of the propensity to morally disengage discussed in the previous section, which specifies an individual difference, not a cognitive process.

Second and more specifically, process moral disengagement denotes *reasoning* processes, that is, cognitive processes “in which logical processes of an inductive or deductive character are used to draw conclusions from facts or premises” (APA 2019). Reasoning processes are characterized by accessing existing knowledge and general principles, making generalizations, drawing conclusions, or constructing explanations. When people morally disengage, they are reasoning. To illustrate, in moral justification, people invoke alternative moral codes or higher obligations to develop an (often biased) argument; in making an advantageous comparison, people provide an evaluative argument; in displacing responsibility, people engage in a sequence of arguments, identify alternative agents, make arguments for their responsibility, and offer conclusions about one’s own blamelessness. While process moral disengagement may be triggered by affect or intuition, moral disengagement manifests itself as reasoning, for example as post hoc rationalizations (Haidt 2001; Shu et al. 2011).

Third, in process moral disengagement, individuals *apply* their reasoning *to moral judgments*. A moral judgment is a “type of evaluative judgement that is based on assessments of the adequacy of one’s own and others’ behaviours according to socially shaped ideas of right and wrong” (Moll et al. 2005, p. 807). Moral judgments may themselves be the result of reasoning processes, or they may emerge through intuition (Haidt 2001).

Fourth, process moral disengagement *reconstrues* moral judgments. The verb “to reconstrue” combines an element of change with an element of continuity. Our intensional definition adopts the term “reconstrual” used by Bandura to emphasize that process moral disengagement does not operate by making drastic changes to prior moral beliefs; instead, moral disengagement provides a reinterpretation of existing beliefs, allowing people to convince themselves that they are still adhering to the same beliefs.

Fifth, process moral disengagement operates *selectively*, it provides a reconstrual with respect to a specific behavior or type of behavior: When people selectively morally disengage a specific moral standard, “in other aspects of their lives, they adhere to their moral standards” (Bandura 2016, p. 3). This attribute of moral disengagement is a second key characteristic, delineating process moral disengagement from a general propensity to morally disengage.

Sixth and last, process moral disengagement *shifts* a moral judgment “*behavior B by actor A is morally wrong*” toward a moral judgment “*behavior B is not morally wrong*” or “*actor A is not responsible for behavior B.*” This attribute specifies two strategies through which process moral disengagement operates—reconstruing morality and reconstruing agency. We argue that both strategies need to be made explicit in the intensional definition. People may effectively invalidate a moral accusation “behavior B by actor A is

morally wrong” either by arguing “A might have done B, but B is not morally wrong” or by arguing “behavior B is in fact morally wrong; yet it was not A doing it.” The strategies may be invoked simultaneously or consecutively (the “or” in the intensional definition is not an exclusive or). It is important, however, to acknowledge that the two strategies may operate independently from each other in process moral disengagement. We use the term “shift” here to emphasize the gradual character of process moral disengagement (Bandura 1990a, 1999; Shu et al. 2011; Welsh et al. 2015). An individual may morally reconstrue a certain behavior that is initially judged as “wrong” toward a judgment that renders the behavior as “not so morally wrong,” “not morally wrong,” or even as “morally right (obligatory).” To illustrate, consider a tax consultant who is helping a client to benefit from *cum ex* transactions that allow the client to reclaim tax payments from fiscal authorities beyond what has actually been paid in taxes (Correctiv 2017). The tax consultant may morally reconstrue own behavior by comparing it with larger transgressions (“it’s not murder”), rendering own behavior as “not so morally wrong.” Alternatively, consecutively, or complementarily, the tax consultant may morally reconstrue own behavior by regress to legality, pointing out that own behavior was within the boundaries of law, that nothing illegal was done, and that law can be the sole criterion to determine immorality in business, rendering own behavior as “not morally wrong.” Furthermore, invoking moral justification via the notion of fiduciary duties to a client, the tax consultant may even morally reconstrue own behavior as “morally right (obligatory).” Likewise, agency may be gradually displaced through process moral disengagement, rendering a moral judgment between “actor A is not solely responsible” and “actor A is not responsible at all” (Bandura 1999; Welsh et al. 2015). We further note that this key attribute of the intensional definition of process moral disengagement allows for substituting “actor A” with both the individual having the moral disengagement cognition, and actors (individuals or groups) other than the individual having the moral disengagement cognition. People not only morally disengage from their own behavior, but also the behavior of others, for example relatives, friends, colleagues, comrades, fellow countrymen, or nations (Beu and Buckley 2004; Osofsky et al. 2005; Leidner et al. 2010; Gino and Galinsky 2012).

Combing the six individually necessary and collectively sufficient key attributes, we thus propose the following intensional definition of process moral disengagement:

Process moral disengagement refers to intrapsychic cognitive reasoning processes through which people selectively reconstrue a moral judgment “behavior B by actor A is morally wrong” and shift it toward becoming “behavior B is not morally wrong” or “actor A is not responsible for behavior B.”

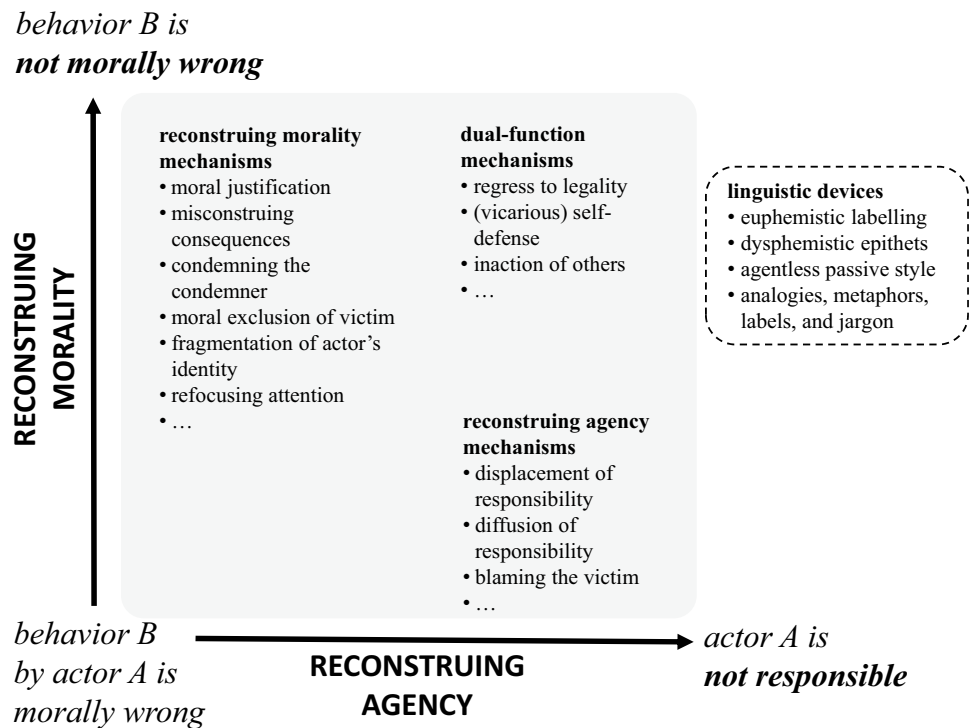
In defining moral disengagement, scholars often specify the relationship between moral disengagement and unethical behavior as part of the conceptualization. To illustrate, Newman et al. specify moral disengagement as “the process of cognitive restructuring that allows individuals to disassociate with their internal moral standards *and behave unethically* without feeling distress” (2019, p. 1; emphasis added). Note that our intensional definition specifies no relationship to unethical behavior, because the close positive link between process moral disengagement and unethical behavior is a central proposition of moral disengagement theory; yet for this reason, it is not a defining key attribute.

Although Bandura developed moral disengagement theory as part of social cognitive theory (e.g., 1986, 1991), it has been recognized by researchers—including Bandura (e.g., 1999, 2018b)—as a standalone theory (Newman et al. 2019). Our intensional definition refrains from using specific jargon of social cognitive theory, yet it is consistent with both approaches.

The intensional definition helps to distinguish the focal construct from related constructs. The first and fifth attributes discussed above help delineate process moral disengagement from the propensity to morally disengage. Furthermore, a conceptualization of moral disengagement as two emotions of disengagement—pride and indifference (Menesini et al. 2003, 2015)—places moral disengagement in the domain of affective cognition instead of in the domain of reasoned cognition. This is a different concept and, therefore, should not be termed moral disengagement.

The intensional definition abandons reference to specific mechanisms. Instead, it highlights the functional role of moral disengagement in changing the assessment of the morality or agency regarding a specific behavior. The intensional definition, however, provides a point of reference for the inclusion of moral disengagement research working with dissimilar sets of mechanisms, and for the exclusion of conceptualizations that deviate from the construct’s core meaning. The definition is broad enough to incorporate fully the existent process moral disengagement literature; it is consistent with conceptualizations that invoke mechanisms specified by Bandura, subsets of these mechanisms, but also extended subsets (which include mechanisms not specified by Bandura). These represent different specifications of process moral disengagement within the common frame provided by the intensional definition. The intensional definition safeguards researchers from overly conservative conceptualizations of process moral disengagement and provides scholars with necessary degrees of freedom to research isolated mechanisms, different groups of already identified mechanisms, or moral disengagement mechanisms not yet identified or labeled as such. The definition is open to the further discovery of mechanisms.

Fig. 1 The open inventory of psychological mechanisms of moral disengagement



The idea of having more than eight mechanisms might trigger aversion in researchers willing to whet Occam’s razor to demand parsimony in theorizing. We argue, however, that it is the intensional definition that provides a parsimonious theoretical framework. Researchers may explore the plurality of psychological mechanisms that the human mind is capable of devising in order to allow for reconstructing morality and agency. In the following section, we leverage the intensional definition to model a simple classification scheme to categorize mechanisms beyond those asserted by Bandura.

An Open Inventory of Psychological Mechanisms of Moral Disengagement

The inventory of psychological mechanisms of moral disengagement (Fig. 1) encompasses mechanisms proposed by Bandura and other early scholars of moral disengagement (Haney 1997; Tsang 2002; Ashforth and Anand 2003). The inventory is modeled along the two strategies of process moral disengagement specified in the intensional definition—reconstructing morality and reconstructing agency—and we graphically represent it as a two-dimensional space.

The model represents the moral judgment that is to be reconstructed by moral disengagement—“behavior B by agent A is morally wrong”—as the origin of the graph. When reconstructing morality, people shift a moral judgment “behavior B by actor A is morally wrong” toward a moral judgment “behavior B is not wrong,” which is represented as the vertical axis of the model. When reconstructing

agency, people shift a moral judgment “behavior B by actor A is morally wrong” toward a moral judgment “actor A is not responsible for behavior B,” which is represented as the horizontal axis of the model. In the resulting quadrant, process moral disengagement could be represented as any vector with positive magnitude and gradient angle of between 0° and 90° determining its direction. The vector analogy is helpful in reminding us that reconstructing morality and reconstructing agency are processes that may occur with different levels of intensity and may happen gradually, for example, by invoking different arguments in parallel or sequentially (in principle adhering to the axioms of vector addition).

Adding a layer to the space provided by the two dimensions of the model, we present three groups of psychological mechanisms through which process moral disengagement operates: reconstructing morality mechanisms, reconstructing agency mechanisms, and dual-function mechanisms. We added ellipses to the list in each group to indicate that the inventory does not claim to provide a full list of psychological mechanisms; we encourage scholars to specify further mechanisms as long as they fulfill the criteria specified in the intensional definition of process moral disengagement.

We adopt Bandura’s wording for five of the twelve psychological mechanisms, presented in more detail below. In two cases, we subsume Bandura’s mechanisms under the terms introduced by others: dehumanization (along with Haney’s difference exaggeration and Ashforth and Anand’s depersonalization) is subsumed under the more general term

“moral exclusion of victim” (Tsang 2002); and advantageous comparison is subsumed under the more general term “refocusing attention” (Ashforth and Anand 2003). Five further mechanisms are specified in the inventory of process moral disengagement: Condemning the condemner (Ashforth and Anand 2003; Sykes and Matza 1957), fragmentation of own identity (Tsang 2002; subsuming the “metaphor of the ledger” of Ashforth and Anand (2003) and Sykes and Matza (1957)), regress to legality (Ashforth and Anand 2003), (vicarious) self-defense (Haney 1997), and inaction of others (Tsang 2002).

Figure 1 also notes the existence of linguistic devices—for example euphemistic labeling (a separate mechanism in Bandura’s set of eight mechanisms) or agentless passive style—that play a role when people morally disengage. In the inventory, these are not represented as psychological mechanisms per se. We argue that these may be symptoms of process moral disengagement, they may be representative of moral disengagement processes, and they may support several moral disengagement mechanisms, but they are not to be regarded as psychological mechanisms in their own right.

The two strategies—reconstructing morality and reconstructing agency—do not operate in empirically distinct ways but may often intermingle in process moral disengagement. Depending on the context and the behavior to be morally disengaged, individuals may choose mechanisms from these two strategies, either to effectively reconstruct the behavior as being justified or to explain that the agent is not responsible for the behavior. Furthermore, a number of mechanisms—dual-function mechanisms—may reconstruct morality, reconstruct agency, or reconstruct both simultaneously or consecutively.

Psychological Processes Reconstructing Morality

The intensional definition of process moral disengagement specified reconstructing morality as a first strategy, whereby people shift a moral judgment “behavior B by Actor A is morally wrong” toward a moral judgment “behavior B is not morally wrong.” When people engage in *reconstructing morality*, they reevaluate the behavior in question, with the result that they convince themselves that the behavior is “less immoral,” “not immoral” (i.e., permissible), or even “morally obligatory.” In reconstructing morality, people weaken the moral frame on which the condemnation of behavior is based, or they invoke an alternative moral frame that reconstructs the behavior in question as having been morally obligatory. People may reconstruct morality via moral justification, misconstruing consequences, condemning the condemner, moral exclusion of the victim, fragmentation of the actor’s identity, and refocusing attention.

Moral justification (e.g., Bandura 1999), by and large, is identical with an appeal to higher loyalties (Ashforth and

Anand 2003; Sykes and Matza 1957). People morally justify a behavior by invoking an alternative moral frame, for example by referencing a religion, an ideology, or a higher obligation to a specific community that obligates an actor to commit the behavior in question. To illustrate, Reinhardt Siekaczek, a manager at Siemens involved in the company’s corruption scandal that surfaced in 2007, morally justified his behavior as an effort of wanting to help the company and co-workers: “We did it for the company, it was about keeping the business unit alive and not jeopardizing thousands of jobs overnight” (Schubert and Miller 2009).

People may also *minimize, ignore, or misconstrue consequences* of their doing (Bandura 1999), or even deny injury altogether (Ashforth and Anand 2003; Sykes and Matza 1957) to reconstruct morality. Misconstruing consequences may render immoral behavior as being venial or amoral. It may also operate to support a consequentialist moral justification, for example when the behavior in question is construed as the lesser of two evils, given an alternative behavior with (putatively) more detrimental consequences. In this case, misconstruing consequences becomes a special case of moral justification. To illustrate, engine developers at Volkswagen, involved in devising diesel engine defeat devices, countered allegations of unethical behavior by minimizing the harm caused by nitric oxides and also by juxtaposing the harm caused by diesel engines with the fact that a more efficient diesel was “a promising way to fight climate change” (Ewing 2017, p. 126).

People may reconstruct morality by *condemning the condemner* (Ashforth and Anand 2003), whereby people challenge the validity of others’ moral condemnation by disputing condemners’ legitimacy. This is done, for example, by contending that condemners themselves are dubious, crooked, or corrupt. The psychological mechanism operates by weakening a salient moral frame via an *ad hominem* argument, reasoning that those who invoke the accusation of morality invalidate it because of their own immorality. To illustrate, employees of an automaker involved in emissions cheating may psychologically invalidate condemnation from regulators and lawmakers by accusing them of having been complicit for decades.

Moral exclusion of victims (Tsang 2002) allows people to qualify victims as undeserving of the same moral status. Victims are dehumanized (Bandura 1999) or the difference between actor and victim is psychologically exaggerated (Haney 1997). Through moral exclusion, people psychologically divide individuals into those inside and outside the moral domain. Moral exclusion may be at work in business contexts when, for example, customers are referred to using technical terms (target, prospect) or referred to using terminology from the animal kingdom (e.g., as “lemmings” or “pigs to be hunted down”), making it psychologically easier to exploit them.

In reconstructing their own morality, people sometimes “fragment their own identity to preserve a moral self-concept” (Tsang 2002, p. 43). In cases of *fragmentation of an actor’s identity*, people psychologically compartmentalize their lives (e.g., private versus public; family versus work), allowing themselves to behave unethically in one domain of life while offsetting it with commendable behavior in another. To illustrate, members of the Sackler family, owners of Purdue Pharma, the producer and marketer of Oxycontin (the drug charged to have caused the US opioid crisis) may have morally disengaged in marketing the drug by psychologically fragmenting their identity and ascertaining their moral goodness by psychologically emphasizing their philanthropic activities for education and the arts.

Furthermore, people may reconstrue morality by *refocusing attention* away from negative aspects of their behavior. People “may willfully deemphasize, compartmentalize, or suppress knowledge of their acts in favor of more normatively redeeming features of their work” (Ashforth and Anand 2003, p. 22). An effective method to refocus attention is to engage in advantageous comparisons (Bandura 1999), which may have the effect of reconstructing problematic behavior as petty, trivial, or irrelevant. To illustrate, bank employees opening bank accounts without customer authorization may refocus attention on the much greater harm caused by reckless behaviors during the 2008 financial crisis, thereby psychologically silencing charges of immorality regarding own behavior.

The psychological mechanisms discussed so far all operate to reconstrue morality. We now turn our focus to a second group of psychological processes through which people may reconstrue agency to morally disengage.

Psychological Processes Reconstructing Agency

The intensional definition of process moral disengagement specified reconstructing agency as a second strategy, whereby people shift a moral judgment “behavior B by Actor A is morally wrong” toward a moral judgment “actor A is not responsible for behavior B.” When reconstructing agency, people attempt to psychologically distance themselves from the behavior, which is achieved through the displacement of responsibility to authorities, the diffusion of responsibility, or by blaming the victim.

In the *displacement of responsibility to authorities* (Arendt 1963; Milgram 1963; Kelman 1973; Bandura 1999), people refer to explicit or implicit orders from legitimate authorities, thereby avoiding the taking of responsibility for their own behavior. To illustrate, employees engaged in opening bank accounts without customer authorization blamed a cascade of management personnel for pushing the fulfillment of ever more aggressive sales targets in the

aftermath of the Wells Fargo *Fake Accounts Scandal* (Masunaga 2016).

Alternatively, people are able to reconstrue agency by *diffusing responsibility* across collectives (e.g., Bandura 1999), thereby minimizing their own agency. When people tell themselves that everyone is doing what they are doing, that they only contribute to a small, insignificant degree to an unethical behavior, or that they acted as a member of a decision-making collective, they are psychologically distancing themselves from having to accept responsibility. To illustrate, Reinhardt Siekaczek, the Siemens manager involved in corrupt practices, may have effectively psychologically distanced himself from having to take responsibility when he claimed that “paying provisions and bribes was customary in practically all business units of Siemens AG” (Schubert and Miller 2009). A third psychological mechanism that people invoke in reconstructing agency is *blaming the victim* (e.g., Bandura 1999), whereby “the status of the victim qua victim is refuted” (Ashforth and Anand 2003, p. 19) and, consequently, responsibility for the actor’s behavior is diminished. Victims of transgressions are presented as complicit (inviting the behavior), as cause (having provoked it first), or as being deserving of the behavior (e.g., in the case of revenge). To illustrate, psychologically blaming the victim happens in cases of sexual harassment at work, when transgressors exculpate themselves by telling themselves that harassment targets were complicit or provoked incidents.

The three reconstructing agency mechanisms discussed so far all focus on externalizing blame for unethical behavior with others. Further psychological mechanisms exist that may achieve the same result. These, however, also may serve to reconstrue morality. Thus, we now turn to discussing a third group of moral disengagement mechanisms, referred to as dual-function psychological mechanisms.

Dual-Function Psychological Mechanisms

Some moral disengagement mechanisms serve to reconstrue morality, reconstrue agency, or reconstrue both. People regress to legality, invoke arguments of (vicarious) self-defense, or point to the inaction of others to either reconstrue morality or agency.

People defend behavior via a *regress to legality* (Ashforth and Anand 2003) when they claim that explicit laws can be the sole test to evaluate whether a behavior or practice is legitimate. Allegations of immorality beyond the test of legality are discredited as having no relevant foundation. In regressing to legality, people reject the moral frame invoked to evaluate the behavior in question as immoral and argue that the law, as an alternative moral frame, can be the exclusive base for deciding what is right and wrong. To illustrate, Volkswagen CEO Müller pushed back against allegations of diesel engine manipulations constituting an “ethical

problem,” arguing that Volkswagen just did not have “the right interpretation of the American law” (Glinton 2016). Alternatively or complementarily, via a regress to legality, people may displace responsibility for a behavior with lawmakers. People reconstrue agency by arguing that it is lawmakers’ obligation to ensure legal frameworks, and that if inconsistencies or gaps exist, lawmakers are responsible for the behavior exploiting these.

In his analysis of the United States capital jury system, Haney (1997) proposed that jury members who have to inflict the death penalty upon a fellow citizen morally disengage when giving their verdict by construing it as vicarious self-defense. Jurors reconstrue their action as the duty to protect the larger community from future threats. In many contexts, self-defense—if proportionate to an attack of another party—is a moral right, or even constitutes a moral obligation. This is especially the case if its benefits are beyond the actor’s own person, for example when safeguarding one’s family, social group, community, or nation. People invoke (*vicarious*) *self-defense* reasoning as an alternative moral frame to justify transgressions. To illustrate, executives may argue that paying bribes is an act of self-defense to protect their organizations from losing out when competitors would do the same. Alternatively or complementarily, when people reconstrue behavior as self-defense, they frame their behavior as a reaction triggered by others. Agency is reconstrued when the responsibility is thus displaced onto others, for example other companies that are willing to bribe or buyers demanding them.

Furthermore, people may interpret the *inaction of others* (Tsang 2002)—for example, of bystanders who observe unethical behavior and who could intervene but fail to do so—as validation of the unethical behavior. People reconstrue morality when they reason that the inaction of others gives them reason to conclude that a behavior can be considered the right thing to do. Alternatively or complementarily, the inaction of others provides a basis to reconstrue agency. Bystanders’ inaction may lead to placing own responsibility onto the larger group of bystanders, for example by arguing that they share responsibility for not objecting, not intervening—for not *speaking up*.

Although the mechanisms provided in the inventory of process moral disengagement mechanisms are distinct, they are not mutually exclusive. More importantly, mechanisms may operate in concert, that is, individuals may reconstrue moral cognitions by employing several mechanisms simultaneously or consecutively, thus engineering highly self-convincing arguments.

Linguistic Devices

Bandura (e.g., 1990a) presented euphemistic labeling as a separate moral disengagement mechanism. Euphemisms are

not the only linguistic device that may play a role in process moral disengagement. Elsewhere, Bandura discussed the role of two further “linguistic tools” (1999, p. 195) in process moral disengagement: the agentless passive style and specialized jargon. More linguistic devices assist moral disengagement, for example dysphemistic epithets, analogies, metaphors, labels, and catch phrases. Linguistic devices may assist both reconstruing morality and reconstruing agency. To illustrate, euphemisms may support reconstruing morality when they are used to misrepresent detrimental consequences of behavior; dysphemistic epithets may support reconstruing agency when blaming victims.

We argue that linguistic devices are not psychological mechanisms of moral disengagement themselves. Rather they are symptoms of moral disengagement, as they may serve as tools in reasoning to reconstrue morality or agency. We thus follow Ashforth and Anand (2003), who observed that moral disengagement may be abetted by the malleability of language, and who thus did not account for it as a separate mechanism.

Discussion and Research Directions

Theoretical Contributions

Clear constructs—and clearly stating the scope conditions of constructs—contribute to good theory (Suddaby 2010). By increasing *concept clarity* and by providing a *new taxonomy*, we made two contributions to the literature.

We increased *concept clarity* in the field in three steps. First, we demonstrated that there is ambiguity in the literature resulting from researchers using the same label, “moral disengagement,” for related yet psychologically and conceptually different constructs, namely process moral disengagement and the propensity to morally disengage (dispositional moral disengagement). Current practice is unsatisfactory, in that scholars at times seem unaware of conceptual differences, mix the two related but distinct variables, and sometimes employ research designs that are at odds with stated conceptualizations and theoretical considerations. As a result, findings may be cloudy at best, misleading at worst. Researchers need to be careful in interpreting the results stated in existent individual studies; they also need to be careful in adopting conclusions from meta-analyses that do not distinguish between the two distinct constructs. We conclude that there is an urgent need for disambiguation. The field will benefit from scholars adopting the practice of declaring upfront whether they study or address process moral disengagement (Bandura 1990a, 1999, 2016) or the propensity to morally disengage (Detert et al. 2008; Moore et al. 2012).

Second, we provided an intensional definition of process moral disengagement as a remedy to the shortcomings in current conceptualizations. We presented and discussed six key attributes of process moral disengagement. Our intensional definition has a number of advantages over commonly used extensional definitions. In addition to being consistent with (and inclusive of) established conceptualizations based on extensional definitions, our intensional definition is compatible with conceptualizations based on variations of extensional definitions; furthermore, our intensional definition is open for the subsumption of further psychological mechanisms. Importantly, the intensional definition properly delineates the concept of process moral disengagement, excluding conceptualizations that use the term “moral disengagement” but are inconsistent with at least one key attribute of process moral disengagement, for example the propensity to morally disengage (e.g., Detert et al. 2008) and moral disengagement as emotions of disengagement (Menesini et al. 2015).

Third, we further contribute to concept clarity by introducing our open inventory of moral disengagement mechanisms to the literature. We showed that the literature remains dominated by an ambiguous and widely misunderstood classification scheme presented by moral disengagement pioneer Bandura. Our open inventory is directly derived from the intensional definition of process moral disengagement in adopting its two strategies—reconstruing morality and reconstruing agency—as dimensions, which provide the space in which we represent process moral disengagement mechanisms.

With the open inventory, we make a second contribution to the literature: *a new taxonomy* to categorize psychological mechanisms of process moral disengagement. We discussed the 12 mechanisms contributed by Bandura and other early theorists of moral disengagement, which explain how process moral disengagement operates. Our taxonomy not only synthesizes mechanisms from existent literature into a more coherent whole, it also pushes our thinking about moral disengagement forward, as it provides us with new avenues to investigate through which mechanisms moral disengagement operates and where we are likely to identify more mechanisms. The close link between the intensional definition of process moral disengagement and the inventory of process moral disengagement mechanisms demands the latter be open to housing further mechanisms to be specified by researchers in the future. The open inventory may initiate research to explore the plurality of psychological mechanisms that the human mind is capable of devising in order to allow for reconstruing morality and agency.

Taken together, our intensional definition, our two-dimensional classification scheme, and our open inventory provide a reconceptualization of moral disengagement as a process, which may alter how scholars of business ethics and other disciplines will conceive of and study process

moral disengagement. Adopting our reconceptualization, ultimately, promises decreased ambiguity of the research in the field and a broader understanding of the variety of processes through which moral disengagement operates.

Directions for Future Research

Extending the Open Inventory

We discussed 12 psychological mechanisms of process moral disengagement. By decoupling specific mechanism from the definition of process moral disengagement, the construct is open for the integration of further mechanisms. Future research could extend the inventory. We believe that searching for further mechanisms may be a fruitful endeavor. Scholars may take two approaches. First, our intensional definition specifies that process moral disengagement operates via reasoning processes; scholars could investigate other reasoned cognition processes regarding their “potential” for moral disengagement. Such investigation may be guided by the taxonomy underlying the inventory of process moral disengagement provided in this article. Second, scholars may investigate specific contexts to identify further mechanisms. We noticed that such studies appeared to be most fruitful in suggesting further moral disengagement mechanisms, for example: Sykes and Matza (1957), who observed delinquent behavior; Haney (1997), who studied the capital jury system; Tsang (2002), who analyzed the Holocaust; and Ashforth and Anand (2003), who analyzed the normalization of corruption in organizations.

Furthermore, while we presented mechanisms using the wording that has been chosen in the literature, scholars may want to establish explications of mechanisms grounded in a more uniform language frame. Scholars could thus leave behind the idiosyncratic meaning provided by the original scholars and aim at providing a unified terminology.

Researching Antecedents and Outcomes of Process Moral Disengagement

In their review of the business ethics moral disengagement literature, Newman et al. (2019) provided an overview of antecedents and outcomes of moral disengagement. In their analysis, however, Newman and colleagues decided not to distinguish between process moral disengagement and the propensity to morally disengage. As we have argued above, this may lead to a misunderstanding of process moral disengagement, especially when the empirical work on the propensity of moral disengagement or ambiguous empirical work is taken as evidence to claim that a finding applies to process moral disengagement. Of the 23 individual-level antecedents of moral disengagement reported (Newman et al. 2019), we could only identify nine—moral identity,

gender (Kennedy et al. 2017), leadership self-efficacy, affective motivation to lead, non-calculative motivation to lead (Hinrichs et al. 2012), psychopathy (Stevens et al. 2012), (negative) emotions (Fida et al. 2015), job security (Huang et al. 2017), and creativity (Keem et al. 2018)—from studies of process moral disengagement. For 14 reported individual-level antecedents of moral disengagement, the review sample would not provide unambiguous evidence for the process instantiation of moral disengagement. Researchers may need to calibrate the literature review's results regarding antecedents and outcomes by consulting Table 1 provided in the Supplementary Material; they should also research further variables to broaden the set of antecedents and outcomes of process moral disengagement.

Implications for Measuring Process Moral Disengagement

Measuring intrapsychic processes via latent variables is challenging. We found that scholars quite often conceptualize process moral disengagement, or claim to study process moral disengagement, but then employ trait measures. General, context-independent trait measures of the propensity to morally disengage (e.g., Moore et al. 2012) are relatively easy to administer. We might only suspect that researchers at times chose trait measures of moral disengagement because they are convenient, thereby committing the error of the drunk who is looking under a lamppost for a key that he lost elsewhere, because “that’s where the light is.” Current practice is often not satisfactory, and we call on researchers of process moral disengagement to choose measures wisely. We offer five pointers that may guide researchers to choose appropriate measures in future research.

First, new measures may incorporate or capture psychological mechanisms of moral disengagement beyond the eight mechanisms invoked by Bandura. The development of such measures needs to follow exploratory research (leveraging established methods such as case studies, interviews, focus groups, or content analysis), which must establish psychological mechanisms that are consistent with the criteria set forth by the intensional definition presented in this article.

Second, as a mental *process*—a sequence of steps that produce a result—process moral disengagement is not well-suited for cross-sectional research designs. If, nevertheless, researchers choose such designs, they must not use general trait measures, but instead need to employ context-specific scales, for example the work-moral disengagement scale, to study counterproductive workplace behaviors (Fida et al. 2015). Choosing context-specific scales comes at a cost, in that these scales often do not exist for the specific context that researchers aim to study, and thus require sumptuous validation. Process moral disengagement may be more adequately studied via longitudinal or experimental designs. In

such cases, measures may be either context-specific or general. In fact, trait measures are adequate to measure process moral disengagement in these cases, because differences (over time for longitudinal designs, and between conditions for experimental designs) may be interpreted as indicating process moral disengagement.

Third, throughout this article we advanced the argument of separating process moral disengagement and the propensity to morally disengage. That some people are more talented than others to engage in process moral disengagement is not central to the process. Nevertheless, in studying process moral disengagement, researchers may consider controlling for individuals’ propensity to morally disengage. This requires separate measures for the two constructs and has been tried by Kish-Gephart et al. (2014), who measured both “situationally-induced moral disengagement” (synonymous with process moral disengagement) and “dispositional moral disengagement.”

Fourth, instead of measuring process moral disengagement directly, researchers may also consider designs in which process moral disengagement is assumed to have been at work, as indicated by its expected effects (e.g., Mazar and Aggarwal 2011). Instead of capturing the precise psychological mechanism that is at work when people morally disengage, such designs measure how people evaluate the ethicality of a behavior or their perceived responsibility for the behavior.

Fifth, current measures predominantly aim at capturing moral disengagement as a higher-order construct that comprises a larger number (typically eight) of psychological mechanisms. We agree with Newman et al. (2019) that an understanding at the level of the individual mechanism may be meaningful and promising. We also call on researchers to conceptualize and measure moral disengagement on the level of the two moral disengagement strategies—reconstructing morality and reconstructing agency—as exemplified in a study by Barsky (2011), who separately measured moral justification and displacement of responsibility.

Conceptualizing Higher-Level Moral Disengagement

Moral disengagement may not only be an intrapsychic phenomenon. Scholars have asserted moral disengagement at collective (e.g., group and organizational) levels, yet related theory is not well-developed.

Bandura and collaborators repeatedly asserted the existence of moral disengagement at higher levels, referring to it as collective moral disengagement (Bandura et al. 1996; Bandura 2002), the moral disengagement of a network of agents (Bandura 1999), or moral disengagement at the social systems level (Osofsky et al. 2005), characterizing it as “an emergent group-level property arising from an interactive, coordinative, and synergistic group dynamics” (White et al.

2009, p. 43). Johnson and Buckley (2015) recently proposed to explore organizational moral disengagement, and in two empirical studies, Pozzoli et al. (2012) and Gini et al. (2015) researched (student-perceived) classroom moral disengagement. As of today, however, the field lacks a comprehensive theory of collective moral disengagement. Scholars have failed to provide clear definitions, specifications of bottom-up emergence processes, and theoretical arguments for top-down effects and collective level direct effects of collective moral disengagement. A multi-level theory of collective (group, organizational) moral disengagement needs to conceptualize (e.g., as a shared unit model; Klein and Kozlowski 2000) and define collective moral disengagement. Our intensional definition of process moral disengagement may serve as a guide to define higher-level moral disengagement. As an instance of reasoned cognition, morally disengaged thinking is readily available for exchange and mutual influence between people, and it may very well lead to coordinated reconstrual, for example via socially shared cognition among members of a collective (Haslam 2004; Tindale et al. 2001). Our intensional definition specified the function of process moral disengagement: selective reconstrual of a moral judgment “behavior B by actor A is morally wrong” and shift it toward becoming “behavior B is not morally wrong” or “actor A is not responsible for behavior B.” This function may serve as a basis to define terms such as “group moral disengagement” or “organizational moral disengagement,” constructs that may be similar in function but across levels may manifest themselves in a distinctly different manner (Morgeson and Hofmann 1999).

We concur with Islam (2019) and Newman et al. (2019) that collective (or more formally “higher-level”) moral disengagement is a promising phenomenon to study, and that via such study business ethics stands to gain a new avenue into understanding how unethical behavior may unfold in social and organizational contexts.

To study the emergence of collective moral disengagement as a process, researchers would ideally conduct field research and observe collectives (e.g., work groups, or a group of individuals who share a common identity) as they reconstrue morality or agency regarding a specific behavior that is qualified as unethical by group norms before the emergence of collective moral disengagement. Such research designs are obviously difficult to implement (How will researchers know when unethical behavior may be considered, thereby requiring moral disengagement?) or may require clearance from ethical research boards, for example when researchers attempt to provoke collectives of subjects to consider engaging in unethical behavior (e.g., Gino et al., 2009; Gino and Galinsky 2012; Shu et al. 2011; Welsh et al. 2015). Alternatively, researchers may investigate the past unethical behavior of a group and try to reconstruct group interaction and mutual influence via interviews with group

members or through content analysis (e.g., email correspondence; documents surfacing, for example as a result of fraud investigations). Such research would be important to model the process of collective moral disengagement emergence in order to understand how (intra-)psychological mechanisms are manifested on the collective level, and to verify whether similar mechanisms are at work in collective moral disengagement, as compared to individual-level moral disengagement.

To study emerged collective moral disengagement, which may best be understood as an emergent state (Marks et al. 2001), multi-level study designs (Klein and Kozlowski 2000) may be devised that capture individuals embedded in their collectives (e.g., work teams, departments, organizations) and that measure process moral disengagement on the individual level, and separately (perceived) moral disengagement of the collective (e.g., the team, the organization) on the collective level (e.g., Gini et al. 2015). To have sufficient statistical power, such an investigation requires sufficiently large numbers of collectives, and an even larger number of individuals who are members of the collectives. Alternatively, experimental studies may aim at simulating membership in a collective or prime identification with a collective. Collective moral disengagement could then be introduced as an independent variable to investigate the effects on a readiness to engage in unethical behavior.

There is no doubt that studying collective moral disengagement is vastly more difficult as compared to individual-level designs. Yet, the reality of unethical behavior in group and organizational contexts, which could only awkwardly be ascribed to individuals’ thoughts and actions, makes such research highly attractive.

Conclusion

“Moral disengagement theory focuses on serious moral issues of real life” (Bandura 2018b, p. 251). A rapidly growing moral disengagement literature, especially in business ethics (Newman et al. 2019), promises a deeper understanding of moral disengagement, unethical behavior, and the interplay between these constructs in organizational contexts. To be able to deliver the full explanatory power of moral disengagement theory, scholars need to agree on a clear and shared understanding of the focal construct. We showed that current researchers are overly liberal in adapting Bandura’s conceptualization, leading to ambiguity in the field. We also argued that the field is overly conservative in adopting a set of (and a classification structure for) eight specific mechanisms of process moral disengagement, preventing a more comprehensive understanding of the phenomenon.

With the help of a new intensional definition, we were able to provide a remedy for both problems. The definition provides a tool to separate two related but distinct constructs: process moral disengagement and the propensity to morally disengage. Although both fields of study are fruitful, we should not endanger scientific progress by mixing these literatures. Furthermore, the intensional definition provides concept clarity, specifying the key attributes that, together, clearly and distinctly define process moral disengagement. We leveraged the definition to offer a two-dimensional classification scheme that structures the open inventory of psychological mechanisms of process moral disengagement. The open inventory promises a broader understanding of the plurality of psychological processes that the human mind is capable of devising when morally disengaging unethical behavior. Our reconceptualization may provide researchers guidance and important new impulses for improving and continuing to explain unethical behavior in business contexts by adapting the perspective provided by moral disengagement theory.

Acknowledgements The authors wish to thank Johan Wempe, Urs Mueller, Michael Redies, Johannes Habel, and two anonymous reviewers for helpful comments on earlier drafts of this article.

Author Contributions U.S. conceived of the presented ideas, reviewed and analyzed literature, and wrote the original manuscripts; U.S. and O.B. worked on critical review, commentary, and revision; U.S. and O.B. read and approved the final manuscript.

Funding No funding was received for this study.

Compliance with Ethical Standards

Conflict of interest All authors declare that they have no conflicts of interest.

Ethical Approval This article does not contain any studies with human participants performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

References marked with two asterisks denote papers included in the review of Newman et al. (2019)

- **Alnuaimi, O. A., Robert, L. P., & Maruping, L. M. (2010). Team size, dispersion, and social loafing in technology-supported teams: A perspective on the theory of moral disengagement. *Journal of Management Information Systems*, 27(1), 203–230.
- APA (2019). *APA dictionary of psychology*. Retrieved November 30, 2019, from <https://dictionary.apa.org/>.
- Arendt, H. (1963). *Eichmann in Jerusalem*. New York: Penguin.
- Ashforth, B. E., & Anand, V. (2003). The normalization of corruption in organizations. *Research in Organizational Behavior*, 25, 1–52.
- **Astrove, S. L., Yang, J., Kraimer, M., & Wayne, S. J. (2015). *Psychological contract breach and counterproductive work behavior: a moderated mediation model*. In *Academy of management proceedings (Vol. 2015, p. 11094)*. Briarcliff Manor, NY: Academy of Management.
- Bandura, A. (1971). *Social learning theory*. Morristown.
- Bandura, A. (1973). *Aggression: A social learning analysis*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive perspective*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1990a). Selective activation and disengagement of moral control. *Journal of Social Issues*, 46(1), 27–46.
- Bandura, A. (1990b). Mechanisms of moral disengagement in terrorism. In W. Reich (Ed.), *Origins of terrorism: Psychologies, ideologies, states of mind* (pp. 161–191). Washington: Woodrow Wilson Center 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, Volume 1: Theory* (pp. 45–103). New York: Lawrence Erlbaum Associates.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193–209.
- Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31(2), 101–119.
- Bandura, A. (2004). Selective exercise of moral agency. In T. A. Thorkildsen & H. J. Walberg (Eds.), *Nurturing morality: Issues in children's and families' lives* (Vol. 5, pp. 37–57). Boston, MA: Springer.
- Bandura, A. (2016). *Moral disengagement: How people do harm and live with themselves*. New York: Macmillan Higher Education.
- Bandura, A. (2018a). Toward a psychology of human agency: Pathways and reflections. *Perspectives on Psychological Science*, 13(2), 130–136.
- Bandura, A. (2018b). A commentary on moral disengagement: The rhetoric and the reality. *American Journal of Psychology*, 131(2), 246–251.
- 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(2), 364–374.
- Bandura, A., Caprara, G. V., & Zsolnai, L. (2000). Corporate transgressions through moral disengagement. *Journal of Human Values*, 6(1), 57–64.

- **Baron, R. A., Zhao, H., & Miao, Q. (2015). Personal motives, moral disengagement, and unethical decisions by entrepreneurs: Cognitive mechanisms on the “slippery slope”. *Journal of Business Ethics*, 128(1), 107–118.
- **Barsky, A. (2011). Investigating the effects of moral disengagement and participation on unethical work behavior. *Journal of Business Ethics*, 104(1), 59–75.
- **Beaudoin, C. A., Cianci, A. M., & Tsakumis, G. T. (2015). The impact of CFOs’ incentives and earnings management ethics on their financial reporting decisions: The mediating role of moral disengagement. *Journal of Business Ethics*, 128(3), 505–518.
- **Beu, D. S., & Buckley, M. R. (2004). This is war: How the politically astute achieve crimes of obedience through the use of moral disengagement. *The Leadership Quarterly*, 15(4), 551–568.
- **Bonner, J. M., Greenbaum, R. L., & Mayer, D. M. (2016). My boss is morally disengaged: The role of ethical leadership in explaining the interactive effect of supervisor and employee moral disengagement on employee behaviors. *Journal of Business Ethics*, 137(4), 731–742.
- Brief, A. P., Buttram, R. T., & Dukerich, J. M. (2001). Collective corruption in the corporate world: Toward a process model. In M. E. Turner (Ed.), *Groups at work: Theory and research* (pp. 471–499). New York: Lawrence Erlbaum Associates.
- Browning, C. (1992). *Ordinary men: Reserve police battalion 101 and the final solution in Poland*. New York: Harper Perennial.
- **Chen, M., Chen, C. C., & Sheldon, O. J. (2016). Relaxing moral reasoning to win: How organizational identification relates to unethical pro-organizational behavior. *Journal of Applied Psychology*, 101(8), 1082–1096.
- **Christian, J. S., & Ellis, A. P. (2014). The crucial role of turnover intentions in transforming moral disengagement into deviant behavior at work. *Journal of Business Ethics*, 119(2), 193–208.
- **Chugh, D., Kern, M. C., Zhu, Z., & Lee, S. (2014). Withstanding moral disengagement: Attachment security as an ethical intervention. *Journal of Experimental Social Psychology*, 51, 88–93.
- **Claybourn, M. (2011). Relationships between moral disengagement, work characteristics and workplace harassment. *Journal of Business Ethics*, 100(2), 283–301.
- **Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2014). Moral character in the workplace. *Journal of Personality and Social Psychology*, 107(5), 943–963.
- Coleman, J. W. (1987). Toward an integrated theory of white-collar crime. *American Journal of Sociology*, 93(2), 406–439.
- Connelly, S., Dunbar, N. E., Jensen, M. L., Griffith, J., Taylor, W. D., Johnson, G., et al. (2016). Social categorization, moral disengagement, and credibility of ideological group websites. *Journal of Media Psychology*, 28(1), 16–31.
- Cook, R. T. (2009). *A dictionary of philosophical logic*. Edinburgh: Edinburgh University Press.
- Correctiv—Recherchen für die Gesellschaft. (2017). *The cumex files. A cross-boarder investigation. How Europe’s taxpayers have been swindled of €55 billion*. Retrieved November 30, 2019, from <https://cumex-files.com/en/>.
- **D’Arcy, J., Herath, T., & Shoss, M. K. (2014). Understanding employee responses to stressful information security requirements: A coping perspective. *Journal of Management Information Systems*, 31(2), 285–318.
- Dahl, A., & Waltzer, T. (2018). Moral disengagement as a psychological construct. *American Journal of Psychology*, 131(2), 240–246.
- **Dang, C. T., Umphress, E. E., & Mitchell, M. S. (2017). Leader social accounts of subordinates’ unethical behavior: Examining observer reactions to leader social accounts with moral disengagement language. *Journal of Applied Psychology*, 102(10), 1448.
- **Detert, J. R., Treviño, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: a study of antecedents and outcomes. *Journal of Applied Psychology*, 93(2), 374–391.
- **Duffy, M. K., Scott, K. L., Shaw, J. D., Tepper, B. J., & Aquino, K. (2012). A social context model of envy and social undermining. *Academy of Management Journal*, 55(3), 643–666.
- **Egels-Zandén, N. (2017). Responsibility boundaries in global value chains: supplier audit prioritizations and moral disengagement among Swedish firms. *Journal of Business Ethics*, 146(3), 515–528.
- **Eriksson, D. (2016). The role of moral disengagement in supply chain management research. *European Business Review*, 28(3), 274–284.
- **Eriksson, D., & Svensson, G. (2016). The process of responsibility, decoupling point, and disengagement of moral and social responsibility in supply chains: Empirical findings and prescriptive thoughts. *Journal of Business Ethics*, 134(2), 281–298.
- Ewing, J. (2017). *Faster, higher, farther: The Volkswagen scandal*. New York: Norton & Company.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140.
- **Fida, R., Paciello, M., Tramontano, C., Fontaine, R. G., Barbaranelli, C., & Farnese, M. L. (2015). An integrative approach to understanding counterproductive work behavior: The roles of stressors, negative emotions, and moral disengagement. *Journal of Business Ethics*, 130(1), 131–144.
- **Fida, R., Tramontano, C., Paciello, M., Ghezzi, V., & Barbaranelli, C. (2018). Understanding the interplay among regulatory self-efficacy, moral disengagement, and academic cheating behaviour during vocational education: a three-wave study. *Journal of Business Ethics*, 153(3), 725–740.
- **Fooks, G., Gilmore, A., Collin, J., Holden, C., & Lee, K. (2013). The limits of corporate social responsibility: techniques of neutralization, stakeholder management and political CSR. *Journal of Business Ethics*, 112(2), 283–299.
- Gini, G., Pozzoli, T., & Bussey, K. (2015). The role of individual and collective moral disengagement in peer aggression and bystanding: a multilevel analysis. *Journal of Abnormal Child Psychology*, 43(3), 441–452.
- Gino, F., & Galinsky, A. D. (2012). Vicarious dishonesty: When psychological closeness creates distance from one’s moral compass. *Organizational Behavior and Human Decision Processes*, 119(1), 15–26.
- Gino, F., Ayal, S., & Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science*, 20(3), 393–398.
- Glinton, S. (2016). ‘We Didn’t Lie,’ Volkswagen CEO Says Of Emissions Scandal. *National Public Radio*, January 11, 2016. Retrieved March 21, 2019, from <https://www.npr.org/sections/two-way/2016/01/11/462682378/we-didnt-lie-volkswagen-ceo-says-of-emissions-scandal?t=1553001760986>.
- Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Haney, C. (1997). Violence and the capital jury: Mechanisms of moral disengagement and the impulse to condemn to death. *Stanford Law Review*, 49(6), 1447–1486.
- Haslam, S. A. (2004). *Psychology in organizations*. New York: Sage.
- **He, P., Peng, Z., Zhao, H., & Estay, C. (2019). How and when compulsory citizenship behavior leads to employee silence: a moderated mediation model based on moral disengagement and supervisor-subordinate Guanxi views. *Journal of Business Ethics*, 155(1), 259–274.
- **Hiekkataipale, M. M., & Lämsä, A. M. (2017). What should a manager like me do in a situation like this? Strategies for

- handling ethical problems from the viewpoint of the logic of appropriateness. *Journal of Business Ethics*, 145(3), 457–479.
- **Hinrichs, K. T., Wang, L., Hinrichs, A. T., & Romero, E. J. (2012). Moral disengagement through displacement of responsibility: The role of leadership beliefs. *Journal of Applied Social Psychology*, 42(1), 62–80.
- **Huang, G. H., Wellman, N., Ashford, S. J., Lee, C., & Wang, L. (2017). Deviance and exit: The organizational costs of job insecurity and moral disengagement. *Journal of Applied Psychology*, 102(1), 26–42.
- **Huang, G. H., & Yan, M. N. (2014). Why groups engage in collective deviance? The role of unethical leadership. In: *Academy of Management Proceedings* (Vol. 2014, p. 13365). Briarcliff Manor, NY: Academy of Management.
- **Hystad, S. W., Mearns, K. J., & Eid, J. (2014). Moral disengagement as a mechanism between perceptions of organisational injustice and deviant work behaviours. *Safety Science*, 68, 138–145.
- Islam, G. (2019). Psychology and business ethics: A multi-level research agenda. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-019-04107-w>.
- **Johnson, J. F., & Buckley, M. R. (2015). Multi-level organizational moral disengagement: Directions for future investigation. *Journal of Business Ethics*, 130(2), 291–300.
- **Keem, S., Shalley, C. E., Kim, E., & Jeong, I. (2018). Are creative individuals bad apples? A dual pathway model of unethical behavior. *Journal of Applied Psychology*, 103(4), 416.
- Kelman, H. G. (1973). Violence without moral restraint: Reflections on the dehumanization of victims and victimizers. *Journal of Social Issues*, 29(4), 25–61.
- **Kempster, S., & Gregory, S. H. (2017). ‘Should I stay or should I go?’ Exploring leadership-as-practice in the middle management role. *Leadership*, 13(4), 496–515.
- **Kennedy, J. A., Kray, L. J., & Ku, G. (2017). A social-cognitive approach to understanding gender differences in negotiator ethics: The role of moral identity. *Organizational Behavior and Human Decision Processes*, 138, 28–44.
- **Kish-Gephart, J., Detert, J., Treviño, L. K., Baker, V., & Martin, S. (2014). Situational moral disengagement: Can the effects of self-interest be mitigated? *Journal of Business Ethics*, 125(2), 267–285.
- Klein, K. J., & Kozlowski, S. W. (2000). *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions*. San Francisco: Jossey-Bass.
- **Knoll, M., Lord, R. G., Petersen, L. E., & Weigelt, O. (2016). Examining the moral grey zone: The role of moral disengagement, authenticity, and situational strength in predicting unethical managerial behavior. *Journal of Applied Social Psychology*, 46(1), 65–78.
- **Lee, K., Kim, E., Bhave, D. P., & Duffy, M. K. (2016). Why victims of undermining at work become perpetrators of undermining: An integrative model. *Journal of Applied Psychology*, 101(6), 915–924.
- **Lee, A., Schwarz, G., Newman, A., & Legood, A. (2019). Investigating when and why psychological entitlement predicts unethical pro-organizational behavior. *Journal of Business Ethics*, 154(1), 109–126.
- Leidner, B., Castano, E., Zaiser, E., & Giner-Sorolla, R. (2010). Ingroup glorification, moral disengagement, and justice in the context of collective violence. *Personality and Social Psychology Bulletin*, 36(8), 1115–1129.
- **Liu, Y., & Berry, C. M. (2013). Identity, moral, and equity perspectives on the relationship between experienced injustice and time theft. *Journal of Business Ethics*, 118(1), 73–83.
- **Loi, R., Xu, A. J., & Liu, Y. (2015). Abuse in the name of injustice: mechanisms of moral disengagement. *Asian Journal of Business Ethics*, 4(1), 57–72.
- **Loyens, K. (2014). Rule bending by morally disengaged detectives: an ethnographic study. *Police Practice and Research*, 15(1), 62–74.
- Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26(3), 356–376.
- **Martin, S. R., Kish-Gephart, J. J., & Detert, J. R. (2014). Blind forces: Ethical infrastructures and moral disengagement in organizations. *Organizational Psychology Review*, 4(4), 295–325.
- Masunaga, S. (2016). November 3, 2016. Wells Fargo says the SEC is also investigating its accounts scandal. *The Los Angeles Times*. Retrieved October 27, 2017, from <https://www.latimes.com/business/la-fi-wells-fargo-sec-20161103-story.html>.
- Mazar, N., & Aggarwal, P. (2011). Greasing the palm: Can collectivism promote bribery? *Psychological Science*, 22(7), 843–848.
- **McFerran, B., Aquino, K., & Duffy, M. (2010). How personality and moral identity relate to individuals’ ethical ideology. *Business Ethics Quarterly*, 20(1), 35–56.
- Menesini, E., Palladino, B. E., & Nocentini, A. (2015). Emotions of moral disengagement, class norms, and bullying in adolescence: A multilevel approach. *Merrill-Palmer Quarterly*, 61(1), 124–143.
- Menesini, E., Sanchez, V., Fonzi, A., Ortega, R., Costabile, A., & Lo Feudo, G. (2003). Moral emotions and bullying: A cross-national comparison of differences between bullies, victims and outsiders. *Aggressive Behavior*, 29(6), 515–530.
- Messick, S. (1987). Validity. *ETS Research Report Series*, 1987(2), i-208
- Milgram, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4), 371–378.
- Moll, J., Zahn, R., de Oliveira-Souza, R., Krueger, F., & Grafman, J. (2005). The neural basis of human moral cognition. *Nature Reviews Neuroscience*, 6(10), 799.
- **Moore, C. (2008). Moral disengagement in processes of organizational corruption. *Journal of Business Ethics*, 80(1), 129–139.
- **Moore, C. (2015). Moral disengagement. *Current Opinion in Psychology*, 6, 199–204.
- **Moore, C., Detert, J. R., Klebe Treviño, L., Baker, V. L., & Mayer, D. M. (2012). Why employees do bad things: Moral disengagement and unethical organizational behavior. *Personnel Psychology*, 65(1), 1–48.
- **Moore, C., Mayer, D. M., Chiang, F. F., Crossley, C., Karlesky, M. J., & Birtch, T. A. (2019). Leaders matter morally: The role of ethical leadership in shaping employee moral cognition and misconduct. *Journal of Applied Psychology*, 104(1), 123.
- Morgeson, F. P., & Hofmann, D. A. (1999). The structure and function of collective constructs: Implications for multilevel research and theory development. *Academy of Management Review*, 24(2), 249–265.
- Newman, A., Le, H., North-Samardzic, A., & Cohen, M. (2019). Moral disengagement at work: A review and research agenda. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-019-04173-0>.
- **Nguyen, B. (2015). The effect of ethical leadership, behavioural integrity, and moral disengagement in predicting turnover intentions during newcomer socialization. *Unpublished doctoral dissertation*, University of Calgary.
- **Niven, K., & Healy, C. (2016). Susceptibility to the ‘dark side’ of goal-setting: Does moral justification influence the effect of goals on unethical behavior? *Journal of Business Ethics*, 137(1), 115–127.
- **Ntayi, J. M., Eyaa, S., & Ngoma, M. (2010). Moral disengagement and the social construction of procurement officers’ deviant

- behaviours. *Journal of Management Policy & Practice*, 11(4), 95–110.
- **Ogunfowora, B., & Bourdage, J. S. (2014). Does honesty–humility influence evaluations of leadership emergence? The mediating role of moral disengagement. *Personality and Individual Differences*, 56, 95–99.
- **Ogunfowora, B., Bourdage, J. S., & Nguyen, B. (2013). An exploration of the dishonest side of self-monitoring: Links to moral disengagement and unethical business decision making. *European Journal of Personality*, 27(6), 532–544.
- Osofsky, M. J., Bandura, A., & Zimbardo, P. G. (2005). The role of moral disengagement in the execution process. *Law and Human Behavior*, 29(4), 371–393.
- **Page, T. E., & Pina, A. (2015). Moral disengagement as a self-regulatory process in sexual harassment perpetration at work: A preliminary conceptualization. *Aggression and Violent Behavior*, 21, 73–84.
- **Palmer, N. (2013). The effects of leader behavior on follower ethical behavior: Examining the mediating roles of ethical efficacy and moral disengagement. *Unpublished doctoral dissertation*. University of Nebraska-Lincoln.
- **Petitta, L., Probst, T. M., & Barbaranelli, C. (2017). Safety culture, moral disengagement, and accident underreporting. *Journal of Business Ethics*, 141(3), 489–506.
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for creating better concept definitions in the organizational, behavioral, and social sciences. *Organizational Research Methods*, 19(2), 159–203.
- Pozzoli, T., Gini, G., & Vieno, A. (2012). Individual and class moral disengagement in bullying among elementary school children. *Aggressive Behavior*, 38(5), 378–388.
- **Pryor, C. G., Lopez, Y. P., Webb, J. W., & Porter, C. O. (2015). Blind justice? How group identification and moral disengagement influence punishment. In: *Academy of Management Proceedings* (Vol. 2015, p. 16702). Briarcliff Manor, NY: Academy of Management.
- **Reynolds, S. J., Dang, C. T., Yam, K. C., & Leavitt, K. (2014). The role of moral knowledge in everyday immorality: What does it matter if I know what is right? *Organizational Behavior and Human Decision Processes*, 123(2), 124–137.
- **Samnani, A. K., Salamon, S. D., & Singh, P. (2014). Negative affect and counterproductive workplace behavior: The moderating role of moral disengagement and gender. *Journal of Business Ethics*, 119(2), 235–244.
- Schubert, S., & Miller, T. C. (2009). At Siemens, bribery was just a line item. *Frontline* February 13, 2009. Retrieved March 19, 2018, from <https://www.pbs.org/frontlineworld/stories/bribe/2009/02/at-siemens-bribery-was-just-a-line-item.html>.
- Shu, L. L., Gino, F., & Bazerman, M. H. (2011). Dishonest deed, clear conscience: When cheating leads to moral disengagement and motivated forgetting. *Personality and Social Psychology Bulletin*, 37(3), 330–349.
- **Stevens, G. W., Deuling, J. K., & Armenakis, A. A. (2012). Successful psychopaths: Are they unethical decision-makers and why? *Journal of Business Ethics*, 105(2), 139–149.
- Suddaby, R. (2010). Editor's comments: Construct clarity in theories of management and organization. *Academy of Management Review*, 35(3), 346–357.
- Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. *American Sociological Review*, 22(6), 664–670.
- **Tasa, K., & Bell, C. M. (2017). Effects of implicit negotiation beliefs and moral disengagement on negotiator attitudes and deceptive behavior. *Journal of Business Ethics*, 142(1), 169–183.
- Tindale, R. S., Meisenhelder, H. M., Dykema-Engblade, A. A., & Hogg, M. A. (2001). Shared cognition in small groups. In M. A. Hogg & R. S. Tindale (Eds.), *Blackwell handbook of social psychology: group processes* (pp. 1–30). Oxford: Blackwell.
- Tsang, J. A. (2002). Moral rationalization and the integration of situational factors and psychological processes in immoral behavior. *Review of General Psychology*, 6(1), 25–50.
- **Valle, M., Kacmar, K. M., & Zivnuska, S. (2019). Understanding the effects of political environments on unethical behavior in organizations. *Journal of Business Ethics*, 156(1), 173–188.
- **Welsh, D. T., Ordóñez, L. D., Snyder, D. G., & Christian, M. S. (2015). The slippery slope: How small ethical transgressions pave the way for larger future transgressions. *Journal of Applied Psychology*, 100(1), 114–127.
- Welzer, H. (2007). *Täter: Wie aus ganz normalen Menschen Massenmörder werden*. Frankfurt: S. Fischer Verlag.
- **White, J., Bandura, A., & Bero, L. A. (2009). Moral disengagement in the corporate world. *Accountability in Research*, 16(1), 41–74.
- Wu, A. D., & Zumbo, B. D. (2008). Understanding and using mediators and moderators. *Social Indicators Research*, 87(3), 367.
- **Zheng, X., Qin, X., Liu, X., & Liao, H. (2019). Will creative employees always make trouble? Investigating the roles of moral identity and moral disengagement. *Journal of Business Ethics*. <https://doi.org/10.1007/s10551-017-3683-3>.
- Zimbardo, P. (2007). *The lucifer effect: How good people turn evil*. London: Rider.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.