DOI: 10.1111/spc3.12472

ARTICLE

WILEY

Everyday dilemmas: New directions on the judgment and resolution of benevolence–integrity dilemmas

Alexander K. Moore | David M. Munguia Gomez | Emma E. Levine

The University of Chicago Booth School of Business, United States

Correspondence
Emma Levine, The University of Chicago Booth School of Business, 5807 S Woodlawn Ave, Chicago, IL 60637.
Email: emma.levine@chicagobooth.edu

Abstract

Many everyday dilemmas reflect a conflict between two moral motivations: the desire to adhere to universal principles (integrity) and the desire to improve the welfare of specific individuals in need (benevolence). In this article, we bridge research on moral judgment and trust to introduce a framework that establishes three central distinctions between benevolence and integrity: (1) the degree to which they rely on impartiality, (2) the degree to which they are tied to emotion versus reason, and (3) the degree to which they can be evaluated in isolation. We use this framework to explain existing findings and generate novel predictions about the resolution and judgment of benevolence–integrity dilemmas. Though ethical dilemmas have long been a focus of moral psychology research, recent research has relied on dramatic dilemmas that involve conflicts of utilitarianism and deontology and has failed to represent the ordinary, yet psychologically taxing dilemmas that we frequently face in everyday life. The present article fills this gap, thereby deepening our understanding of moral judgment and decision making and providing practical insights on how decision makers resolve moral conflict.

1 | INTRODUCTION

It's raining. The bus trundles up to the stop during the morning rush hour and opens its door. A man in a suit, drenched from head to toe, stumbles on, clutching his briefcase to his chest to keep it dry. He reaches into his pocket for his wallet. It's not there. He tries another. No wallet. He looks up at the driver. "I forgot my wallet. I really need to

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

get on this bus. If I don't, I'm gonna be late for work," he says, before dropping his gaze to his soaked loafers. "Would you let me on? Just today?" His voice is barely a whisper. Now, the bus driver faces a dilemma. On the one hand, it would display kindness toward this bedraggled man if she makes an exception and lets him onto the bus. She knows she won't get caught if she does so. On the other hand, making this exception may be unfair to all the other people who paid to ride the bus. And what if there's another person that forgets his wallet. Should she let him on the bus for free, too?

The bus driver in the story above is facing a specific example of a common class of moral dilemmas: benevolence–integrity dilemmas. She is faced with a choice between upholding an impartial principle (integrity) and violating that principle to be compassionate toward a specific individual in need (benevolence). We use the terms "benevolence" and "integrity" to characterize two broad families of moral motivation, which frequently (but do not always) conflict. We define benevolence as the motivation to improve the welfare of an individual in need based on the specific situation the individual faces (Mayer, Davis, & Schoorman, 1995), and we define integrity as the motivation to adhere to universal principles in an impartial manner (Mayer et al., 1995; McFall, 1987).

It is important to clarify that we conceptualize benevolence and integrity as *motivations*, rather than behaviors or moral principles (e.g., utilitarianism vs deontology, Conway & Gawronski, 2013; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001). Describing people's choices as having been motivated by benevolence or integrity does not necessarily provide information about the specific choice they made; rather, it describes *why* they made it. For example, when employees fail to meet expectations, managers must sometimes choose between delivering candid feedback and preserving an employee's self-esteem. In this situation, a manager who is motivated by benevolence makes her decision about whether and how to provide the feedback based on the rich set of cues relating specifically to the employee and the situation. A benevolent manager might choose to withhold harsh feedback based, for example, on her knowledge that the employee is struggling with a chronic illness and the feelings of compassion this knowledge elicits. On the other hand, she might provide the feedback based on the employee's facial expression when he enters the room and the determination that he would benefit from the feedback in the long run even if others may not. Despite leading to opposing behaviors, both cases represent benevolence because the decision relies on the specifics of the situation rather than a universal rule. In contrast, a manager who is motivated by integrity would base her decision on a rule. She may deliver harsh feedback based on a simple, deontological rule that she must always tell the truth. She may also withhold harsh feedback based on a more complex, rule that she should provide truthful feedback as long as it does not cause unnecessary harm (Levine, 2019). This example illustrates how conceptualizing benevolence and integrity as motivations means that they can lead to the same or different actions and outcome but rely on distinct psychological processes.

Benevolence and integrity have historically been positioned as two antecedents of organizational trust (Mayer et al., 1995), but these constructs are suitable for characterizing a broader range of phenomena. For example, benevolence is closely associated with values like loyalty (Haidt & Graham, 2007; Waytz, Dungan, & Young, 2013), mercy (Exline, Worthington, Hill, & McCullough, 2003), care (Gilligan, 1993; Haidt & Graham, 2007), and empathy (Zaki, 2014), whereas integrity is closely associated with values like honesty (Larzelere & Huston, 1980) and fairness (Greenberg, 1987), which are characterized by a motivation to act in accordance with a universal principle in all situations. Integrity and benevolence can also be thought of as synonyms of justice (Kohlberg, 1964) and care (Gilligan, 1993), respectively, as conceptualized by the moral sentiments literature. Like integrity, justice describes an adherence to universal moral principles (Kohlberg, 1964), while, like benevolence, the ethic of care focuses on caring for individuals in need (Gilligan, 1993).[1]

Benevolence–integrity dilemmas characterize a wide range of quotidian situations. They are likely to cause the greatest internal conflict when a decision prescribed by a salient rule conflicts with the welfare of an individual in need. This type of conflict exists in situations involving enforcing rules and meting out punishment (Aquino, Grover, Goldman, & Folger, 2003; Exline et al., 2003; Tyler, 2006), as well as speaking difficult truths, including providing critical feedback (DePaulo & Kashy, 1998; Lupoli, Jampol, & Oveis, 2017), delivering poor prognoses (Levine et al., 2018), or engaging in whistle-blowing (Dungan, Waytz, & Young, 2015; Waytz et al., 2013).

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

Despite the frequency with which benevolence–integrity dilemmas are encountered in everyday life, they are surprisingly understudied. Instead, existing research primarily examines conflicts between moral values and self-interest (for a review, see T. Zhang, Gino, & Margolis, 2018), or among deontological and utilitarian impulses (Conway & Gawronski, 2013; Duke & Bègue, 2015; Greene et al., 2001; Waldmann & Dieterich, 2007). For example, myriad papers in the past twenty years feature "sacrificial dilemmas" (Bartels & Pizarro, 2011) like the trolley problem (Foot, 1967; Thomson, 1985), which involve fantastical situations with high stakes. These types of dilemmas have been helpful in documenting a variety of phenomena relating to dual systems theories of morality (for a review, see Cushman, 2013). Nonetheless, they are often criticized for lacking external validity due to their lack of plausibility (Bauman, McGraw, Bartels, & Warren, 2014) and failing to represent everyday dilemmas (Kahane, 2015).

While our conceptualization of benevolence and integrity is related to the utilitarian–deontological distinction, it encompasses a broader range of conflicts that occur in organizations and routine interpersonal interactions. Within the utilitarian–deontological dilemmas generally studied in psychology, the deontological stance is typically operationalized as abiding by a "do no harm" principle, while the utilitarian stance focuses on maximizing overall welfare without necessarily abiding by any universal rules (see Conway & Gawronski, 2013; Greene et al., 2001). In these dilemmas, deontology bears a psychological resemblance to benevolence, which is focused on caring for or avoiding harm to those in need, whereas utilitarianism bears a psychological resemblance to integrity, which is focused on impartiality. Despite this resemblance, deontology is also philosophically motivated by impartial moral considerations. Moral rules like "do no harm" and "never lie" (e.g., Kant, 1785) advocated by deontology are to be applied equally under all circumstances. Neither deontology nor utilitarianism, as defined in most psychology literature, provide any room for considering partial moral concerns like loyalty or care, which depend on personal relationships (Everett, Faber, Savulescu, & Crockett, 2018). When any impartial moral system like deontology or utilitarianism motivates an action, it is motivated by integrity. In contrast to this motivation to adhere to moral systems, benevolence captures the partial moral considerations based on contextual information in a specific situation. For example, people motivated by benevolence may be influenced by the contextual cues contained within knowledge of specific individuals (Loewenstein & Small, 2007), social relationships (Shaw, DeScioli, Barakzai, & Kurzban, 2017), and emotions evoked by social situations in general (Van Kleef, 2009).

What might be the evolutionary origin of these motivations? Benevolence likely operates through the neurobiological systems that developed to ensure the survival of one's gene pool. A species' ability to respond to its offspring's emotional expressions of hunger, pain, and fear was more likely to ensure its genetic legacy (Decety, 2011). Hence, benevolence may reflect an ancient sensitivity to the distress of one's close kin, which we also observe in other mammalian species (e.g., chimpanzees, Parr, 2001; rats, Sato, Tan, Tate, & Okada, 2015). In contrast, integrity may have developed from the adaptive pressures of life in society. Ensuring cooperation with multiple individuals in large and complex groups may have required a distinct set of skills than those afforded by benevolence (Nowak & Sigmund, 2005). A desire to adhere to a set of shared rules may have facilitated cooperation by, for example, building and maintaining one's reputation. (For a richer treatment of evolutionary accounts of morality, see Haidt & Kesebir, 2010.)

Our broader conception of morality provides a useful lens through which to examine ethical dilemmas. In the present article, we discuss the defining features of benevolence and integrity and use them to make a novel set of predictions. Specifically, we identify three key attributes on which benevolence and integrity differ: the degree to which they reflect an individuating versus impartial orientation, the degree to which they appeal to emotion versus reason, and the degree to which they can be evaluated in isolation. By understanding these characteristics, we bridge research on trust, moral psychology, and judgment and decision making; we bring structure to the existing research on ethical dilemmas; and we make concrete predictions that can guide future research.

1.1 | Impartial versus individuating

Benevolence and integrity differ in the degree to which they consider the circumstances and attributes of individuals involved in a moral dilemma. Benevolence requires an actor to individuate a target—a moral actor must recognize a

target and determine that target's needs. Unlike benevolence, integrity does not require individuation. One need not know the specifics of an individual to apply an impartial principle. The dilemma faced by the bus driver in our opening example exemplifies the conflict that can arise between the individuating tendency of benevolence and the impartial tendency of integrity. The driver's desire to let the man ride without paying may stem from the individuating tendencies of benevolence. In relying on benevolence, the driver may dwell on the feeling evoked by the drenched man's woeful visage and her knowledge of his dire situation. Her motivation is driven by this man standing before her rather than a desire to abide by her principles. On the other hand, the driver's desire to deny the man's pleas may stem from the impartial tendencies of integrity. Beyond providing her with enough information to abide by her principles (like fairness), the particulars of the man and his situation do not matter. Her choice is not about this person begging to be let onto the bus but about all people who might ride the bus.

As illustrated above, a key distinction between benevolence and integrity is whether decisions are motivated by features of specific individuals and their respective situations or standards that would be applied to all similar cases. This difference between benevolence and integrity is further manifested in the importance that these motivations place on relationships with others. Interpersonal relationships are central to benevolence because of its individuating nature. By individuating people, benevolence pushes one to consider not only people and their particular circumstances but also one's relationship to them when engaging in a moral action. Is this person a stranger or a friend? A colleague or my boss? The norms that govern these relationships can compel partiality in some circumstances while exerting minimal pressure in others. For example, when an individual is identified as part of one's in-group, the “communal sharing” norms of the relationship urge attending to their needs (A. P. Fiske, 1992). Hence, people expect their close friends to be benevolent. Not being benevolent can be seen as immoral and damaging to the relationship (Rai & Fiske, 2011; Shaw et al., 2017). In the case of the bus driver, had the man standing in front her been her beloved brother-in-law, she might have felt even more compelled to let him ride without paying. In contrast to benevolence, integrity requires that one stay impartial regardless of relationship. Indeed, abiding by a universal principle often requires that one explicitly ignore relationships. In sum, benevolence requires individuation, which may entail sensitivity to the relationship between an actor and a target, whereas integrity requires impartiality.

1.2 | Emotion versus reason

Benevolence and integrity also differ in the degree to which they appeal to and rely on emotion and reason. A large body of research suggests that human judgment and decision making stems from two oft-competing processes: a more automatic and emotion-based process and a more deliberative, reason-based process (e.g., Epstein, 1994; Greene et al., 2001; Hsee, Rottenstreich, & Xiao, 2005; Hsee, Yang, Zheng, & Wang, 2015; Kahneman & Frederick, 2002, 2005; Levine, Barasch, Rand, Berman, & Small, 2018; Loewenstein & Small, 2007; Milkman, Rogers, & Bazerman, 2008; Sloman, 1996). We argue that, when motivating judgments and behaviors, benevolence uses cues based on one's emotional response to a specific situation more than integrity. In contrast, integrity relies more on reason-based principles.

Let us recall the bus driver from the opening example. She faces the choice to either allow the man to board the bus without paying or apply the same rule that all other riders abide by: no pass, no ride. If the driver is motivated by benevolence, she might base a decision to allow the meek, rain-soaked man onto the bus on the overwhelming empathy she feels for him. Indeed, helping decisions like this one often rely on emotional cues, sometimes to the exclusion of more reason-based processes (Greene, 2013; Haidt, 2001; Loewenstein & Small, 2007). Conversely, if she is motivated by integrity, she might base a decision to leave the man standing at the curb on a desire to uphold a reason-based principle of treating everyone equally.

Research on identifiable and statistical victims supports our proposition (Loewenstein & Small, 2007; Small, Loewenstein, & Slovic, 2007). This body of research finds that concrete beneficiaries typically evoke more compassion and empathy than abstract and diffuse beneficiaries (Loewenstein & Small, 2007). Actions motivated by

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

benevolence are directed at specific people; in other words, the beneficiaries of these behaviors are often singular and concrete (i.e., they are identifiable). Therefore, benevolence is inspired (at least in part) by one's affective response to a person in need. On the other hand, integrity reflects a desire to apply rules impartially across people and situations; in other words, the beneficiaries of these behaviors are diffuse and abstract (e.g., "society"). Therefore, integrity may rely more on rational thinking.

Importantly, we do not make claims about the speed at which benevolence and integrity-based motivations are activated, nor do we make claims about the cognitive effort they require. Though benevolence relies more on one's affective response to a situation than does integrity, affect may not be the only input that influences judgment. When relying on benevolence, emotion serves as only one piece of a potentially large amount of information that a moral actor considers about the current situation. Similarly, though we posit that integrity relies more on reason than benevolence, it may or may not require effortful deliberation. For example, the bus driver may quickly and effortlessly call on a specific principle (e.g., "no pass, no ride") and decide not to allow the man to ride, or she may engage in effortful deliberation to decide how her more general principles (e.g., "be fair") inform her behavior in the present situation. That is, translating a broad principle into a more specific principle that applies within a given situations and others like it may require deliberation. We posit that deliberation is likely to be more intense when situations are complex or novel, in which case it is less obvious how one should apply a given rule (Greene, 2013; Haidt, 2001; Sunstein, 2005). However, as a situation is encountered with greater frequency, acting in accordance with a rule can reduce cognitive effort by allowing a person not to attend to the specifics of the situation (Rand, 2016).

1.3 | Evaluability

Evaluability reflects the ease or difficulty with which an attribute can be evaluated independently (Hsee, Loewenstein, Blount, & Bazerman, 1999). An attribute is said to be evaluable when its value can be assessed without reference to other attribute values. For example, in the prototypical evaluability paradigm (Hsee, 1996), participants in different conditions were asked to evaluate the desirability of two dictionaries based on the number of entries they contained and whether they had any defects. On the one hand, the presence of a defect is highly evaluable. One only needs to look at a dictionary to determine whether it has a torn cover or not. On the other hand, the number of entries is not evaluable. It is hard to evaluate how many entries is "a lot" of entries without knowing the number of words contained in other dictionaries. Crucially, these differences in evaluability can lead to different choices that depend on whether people are thinking of a choice in isolation or in relation to other options (J. Zhang, 2015). Attributes that can be evaluated in isolation, like the torn cover, tend to drive judgment when a dictionary is presented on its own. On the other hand, attributes that can only be evaluated in the presence of other dictionaries, like the number of entries, tend to drive judgment when two or more dictionaries are presented together. Past research has found that evaluability can also influence moral judgments. For example, people tend to act more consistently when different variants of the trolley dilemma are evaluated simultaneously rather than separately. It has been hypothesized that since consistency with a given principle is difficult to evaluate in isolation, this effect may be caused by making the consistency or inconsistency of decisions more salient (Barak-Corren, Tsay, Cushman, & Bazerman, 2018).

We posit that actions based on benevolence and integrity differ in the ease with which they can be evaluated. As with the torn dictionary cover described above, benevolent actions are relatively easy to evaluate in isolation. This ease in evaluation stems from the highly contextual nature of benevolence. The information needed to determine whether a course of action was motivated by benevolence or not tends to be derived from the specifics of the situation under consideration. This context specificity implies that people motivated by benevolence do not necessarily need to look beyond the current context to inform their decisions. For example, an observer may attribute benevolence to a person who performs a single compassionate action based on the immediately observable gains to the welfare of the recipient. On the other hand, observers may attribute a lack of benevolent motivation to a person who

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

performs a single harmful action (Levine & Schweitzer, 2014, 2015). This response may be based on both the immediate harm caused to the recipient of the action and the emotional response this harm evokes within the observer.

Integrity, on the other hand, is harder to evaluate in isolation. It requires *consistent* adherence to general moral principles and therefore necessitates the observation of multiple decisions in order to be assessed (Frazier, Johnson, Gavin, Gooty, & Snow, 2010; Mayer et al., 1995). Furthermore, any single act that violates the simplest rule (e.g., passengers must pay the fare to ride the bus) may reflect adherence to a more complex rule (e.g., passengers must pay the fare to ride the bus, unless someone else on the bus is willing to cover their fare). As a result of this property, integrity becomes more salient in the context of a set of actions with which it is consistent.

2 | PREDICTIONS

The proposed three-factor framework generates novel predictions on ethical decision making. Our predictions explore how benevolence–integrity dilemmas interact with (1) one's relationships, (2) decision contexts, and (3) the number of decisions one experiences. These interactions influence the ways people resolve and judge these conflicts. We detail these predictions below and summarize them in Table 1.

2.1 | Relationships

Benevolence and integrity differ in their emphasis on partiality and impartiality. This difference informs predictions regarding the types of relationships in which benevolence and integrity will motivate judgment and action. We predict that within emotionally close relationships like friendships, people will rely on benevolence, while in more distant relationships, they will rely on integrity. This difference in motivation stems from the expectations that govern different types of relationships (A. P. Fiske, 1992; Rai & Fiske, 2011). In general, close relationships, whether between friends, family members, or colleagues, are centered around shared socio-emotional goals (Pillemer & Rothbard, 2018). They are characterized by mutual support in times of need (DeScioli & Kurzban, 2009) and an appreciation for a specific relationship with specific individuals (Ingram & Zou, 2008). Close relationships often demand partiality; people expect their friends to come to their aid whether they are blameworthy or not (Shaw et al., 2017). Integral to these features of friendships is the individuation of another person. In order to share socio-emotional goals and behave partially toward another, it is necessary to recognize the unique features and circumstances of another person.

In contrast, distant relationships, like those found within large groups or in relatively impersonal commercial transactions, are instrumental (Halevy, Berson, & Galinsky, 2011) and centered on norms of equality, fairness, and adherence to procedure (A. P. Fiske, 1992; Leventhal, 1980). These expectations stem from an inability to benefit from the partiality central to close relationships (Moore, Lewis, Levine, & Schweitzer, 2019). After all, when there are fixed resources or time constraints, if one person is benefiting from partial treatment, others will have resources diverted

TABLE 1 Defining features of benevolence and integrity and corresponding predictions

	Defining features		Predictions	
	Benevolence	**Integrity**	**Benevolence is favored ...**	**Integrity is favored ...**
Orientation toward people	Individuating	Impartial	1. in close relationships	1. in distant relationships
Psychological process engaged in decision making	Emotion driven	Reason driven	2. in actual decisions 3. when judging decisions, during or immediately after they have been enacted	2. in intended decisions 3. when judging decisions a while after they have been enacted
Evaluability	High evaluability	Low evaluability	4. in single-shot decisions	4. in repeated decisions

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

away from them (Pearce, 2015). As a result of these features, impersonal interactions tend to be characterized by features central to integrity: impartiality and lack of individuation. Hence, a person will favor different resolutions for similar dilemmas that arise in disparate relational contexts. For example, an actor may be motivated by benevolence when interacting with a friend but integrity when interacting with a stranger.

2.2 | Decision contexts that promote emotion versus reason

The degree to which benevolence and integrity differ in their reliance on emotion versus reason informs predictions regarding the decision contexts in which benevolence and integrity will be favored. Benevolence uses emotion as an input and therefore is more likely to be relied upon in contexts that are emotionally evocative (or affect-rich; Hsee & Rottenstreich, 2004; Rottenstreich & Hsee, 2001) and less likely to be relied upon in contexts that are emotionally barren. Conversely, integrity places importance on adherence to impartial rules and places less weight on emotional inputs, which arise from a situation. As a result of not considering emotional inputs, people are more likely to rely on, and appreciate, integrity in contexts where these inputs are muted or easier to ignore. Though many contexts may differ in the degree to which they promote emotional versus reason-based processes, we focus on two contextual differences that are particularly relevant to the resolution and judgment of ethical dilemmas: (1) intended versus actual decisions and (2) whether observers are judging the decision immediately or after a delay.

2.2.1 | Intended versus actual decisions

We predict that people will often intend, or plan, to act with integrity prior to a situation but end up acting benevolently when they experience it. Planning requires imagining a future event, which is relatively affect-poor (Bazerman, Tenbrunsel, & Wade-Benzoni, 1998; Loewenstein, 1996; O'Connor et al., 2002). However, experiences are often more affect-rich than people expect (O'Connor et al., 2002). When people actually engage with a decision, their behavior is often motivated by emotion (Bazerman et al., 1998; Milkman et al., 2008; Monin, Pizarro, & Beer, 2007). Conversely, planning promotes more rational behavior in future affect-rich contexts (e.g., Ariely & Wertenbroch, 2002; Taylor, Pham, Rivkin, & Armor, 1998). For example, past research shows that two forms of planning, commitment devices (Ariely & Wertenbroch, 2002; Gollwitzer & Sheeran, 2006) and implementation intentions (Gollwitzer, 1999; Gollwitzer & Sheeran, 2006; Milkman, Beshears, Choi, Laibson, & Madrian, 2011), can help people adhere to their goals and overcome emotional temptations they may feel in the moment.

Building on this research, we propose that people will rely on integrity when they plan future behavior. However, unless people are held accountable for executing these plans (as is the case with commitment devices), they may deviate from this behavior when they actually make decisions in more affect-rich environments. Recent research on allocation dilemmas (Cooney, Gilbert, & Wilson, 2016) lends support for this prediction. Allocators overweight the importance of fairness (an impartial rule, which aligns with our conceptualization of integrity) when predicting recipients' future affective responses to allocations. This overweighting occurs, in part, because allocators consider the allocation process before the outcomes are experienced (i.e., during the planning process).

2.2.2 | Immediate versus delayed judgments

Observers, who are responsible for judging and punishing moral actors, can form judgments at two different points in time: (1) when witnessing the event or (2) when reflecting on or learning about the event after it has happened. Recalling an event is different than experiencing an event (Loewenstein, 1996). People tend to be more emotional as they are experiencing or witnessing events than when they recall them (Bazerman et al., 1998; O'Connor et al., 2002). Furthermore, as time passes, more contextual information may become available and make integrity more salient. Consider the example of the bus driver. A passenger on the bus who witnesses the event may be compelled by the soaked man's state and reward a bus driver who responds to it benevolently. However, two weeks later, the

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

memory of the man may not be as vivid (Winningham, Hyman, & Dinnel, 2000), and the passenger may consider all of the other people who had to pay to ride the bus in subsequent rides. Therefore, the passenger may now think it would have been more appropriate for the driver to act in accordance with integrity by denying the man a free ride to uphold a standard of fairness. Broadly, we predict that people will judge behaviors that they infer to be motivated by benevolence more favorably when making a judgment immediately but will judge behavior that they infer to be motivated by integrity more favorably as time passes. Actors may engage in similar reversals; they may be motivated by benevolence in the moment, but when recalling the same dilemma from a relatively detached perspective, wish they had acted with integrity. Taken together, these predictions also suggest that actors may prioritize benevolence in their actual decisions, which may be resented by targets or observers in the long run.

2.3 | Single versus repeated decisions

The degree to which benevolence and integrity differ in evaluability informs predictions regarding their prioritization in single versus repeated decisions. Attributes that are high in evaluability tend to carry more weight in single decisions, whereas attributes that are low in evaluability tend to carry more weight in the context of multiple decisions (Hsee, 1996; Hsee et al., 1999). Building on this work, we posit that people will rely on benevolence more in single-shot decisions, whereas they will rely on integrity more in repeated decisions.

Beliefs about whether situations will be repeated may also lead to differential reliance on benevolence and integrity. Specifically, when people believe they will only encounter a dilemma once, we predict that they are more likely to rely on benevolence. In contrast, when people believe they will encounter a dilemma multiple times, they are more likely to rely on integrity. For example, if the bus driver believes that she will only be asked to allow someone onto a bus without paying once, she may not consider whether universal rules should govern her decision (i.e., integrity), and she may rely solely on benevolence. On the other hand, if a recent training session at work informs her that people frequently ask to be let onto buses without paying, when faced with the same dilemma, she may rely on integrity. These differences imply that providing people with a reference for the frequency with which they are likely to face certain dilemmas may help them to better understand when benevolence and integrity are warranted.

Evaluability is also likely to lead to different reliance on benevolence and integrity depending on whether multiple dilemmas are presented jointly or sequentially. We predict that when dilemmas are presented one at a time as opposed to all at once, the emotional benefits of benevolence are likely to feel compelling. When dilemmas are presented simultaneously, though, people are more likely to have a desire for moral consistency (Fishbach & Dhar, 2005; Miller & Effron, 2010), leading them to favor integrity.

Finally, differences in evaluability are likely to change observers' inferences about whether behaviors are consistent with benevolence or integrity depending on whether they are seen once or multiple times. People are likely to form inferences about desires, personality, and moral character based on actions (Inbar, Pizarro, & Cushman, 2012; Malle, Knobe, & Nelson, 2007). As such, we predict that people will judge actions that fail to signal benevolence (but could plausibly be motivated by integrity) more harshly within a single decision than they will when that decision is viewed as part of a series of decisions that consistently apply a rule. A single action consistent with integrity, committed at the expense of a specific individual, will be judged based on its lack of benevolence because it is easy to infer benevolence on its own, but it is hard to infer integrity without more information. Once this action is viewed in the broader context of similar and consistent behaviors, the integrity of the actor becomes clearer. It is now possible to see that the initial action was, indeed, motivated by integrity rather than unkindness. Levine and Schweitzer's (2015) work on prosocial lying provides some evidence in support of this prediction. In their studies, people are faced with a choice between lying for prosocial reasons, and therefore acting benevolently, or telling the truth, thereby acting with integrity by adhering to a principle of honesty. In these situations, observers saw a single act of prosocial lying as moral and motivated by benevolence, whereas they saw a single hurtful truth as malevolent and immoral. However, a follow-up study finds that people judge hurtful truths more positively when they are a part of a broader pattern of consistent truth-telling (Levine, unpublished data).

3 | CONCLUSION

In our day-to-day lives, we are often pulled between benevolence and integrity. It can be difficult to decide whether to lend a hand to a person in need or adhere to the principles we hold dear. Like life itself, these dilemmas are complex; they force us to make choices, sometimes heartbreaking ones, without giving us the satisfaction of moral clarity. Should we be loyal or honest? Kind or fair? Compassionate or principled? Over the course of this review, we developed a three-factor framework that differentiates these two moral motivations. Benevolence is defined by its individuating orientation, reliance on emotion, and the ease with which it can be evaluated. In contrast, integrity is defined by its impartial orientation, reliance on reason, and low evaluability. Using these properties, we posed four central predictions (see Table 1) about how preferences for benevolence and integrity vary based on one's relationships, the decision context, and the number of decisions a person experiences. We hope that testing these predictions can help us better understand the lived experience of people as they navigate the complexities of everyday moral life.

ACKNOWLEDGEMENTS

We are grateful for feedback and advice from Nicholas Herzog, Christopher Hsee, and Maurice Schweitzer.

ENDNOTE

[1]We chose to focus on the benevolence–integrity distinction to better bridge the gap between the moral psychology and organizational trust literatures, but care and justice would also be appropriate labels.

ORCID

Alexander K. Moore https://orcid.org/0000-0003-4046-3117
David M. Munguia Gomez https://orcid.org/0000-0002-0161-726X
Emma E. Levine https://orcid.org/0000-0001-7631-4231

REFERENCES

Aquino, K., Grover, S. L., Goldman, B., & Folger, R. (2003). When push doesn't come to shove. *Journal of Management Inquiry*, *12*(3), 209–216.

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, *13*(3), 219–224.

Barak-Corren, N., Tsay, C.-J., Cushman, F., & Bazerman, M. H. (2018). If you're going to do wrong, at least do it right: Considering two moral dilemmas at the same time promotes moral consistency. *Management Science*, *64*(4), 1528–1540.

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*(1), 154–161.

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536–554.

Bazerman, M., Tenbrunsel, A. E., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Making Decisions with competing internal preferences. *The Academy of Management Review*, *23*(2), 225–241.

Conway, P., & Gawronski, B. (2013). Deontological and utilitarian inclinations in moral decision making: A process dissociation approach. *Journal of Personality and Social Psychology*, *104*(2), 216–235.

Cooney, G., Gilbert, D. T., & Wilson, T. D. (2016). When fairness matters less than we expect. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, *113*(40), 11168–11171.

Cushman, F. (2013). Action, outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), 273–292.

Decety, J. (2011). The neuroevolution of empathy. *Annals of the New York Academy of Sciences*, *1231*(1), 35–45.

DePaulo, B. M., & Kashy, D. A. (1998). Everyday lies in close and casual relationships. *Journal of Personality and Social Psychology*, *74*(1), 63–79.

DeScioli, P., & Kurzban, R. (2009). The alliance hypothesis for human friendship. *PLoS ONE*, *4*(6), e5802.

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

Duke, A. A., & Bègue, L. (2015). The drunk utilitarian: Blood alcohol concentration predicts utilitarian responses in moral dilemmas. *Cognition*, *134*, 121–127.

Dungan, J., Waytz, A., & Young, L. (2015). The psychology of whistleblowing. *Current Opinion in Psychology*, *6*, 129–133.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709–724.

Everett, J. A. C., Faber, N. S., Savulescu, J., & Crockett, M. J. (2018). The costs of being consequentialist: Social inference from instrumental harm and impartial beneficence. *Journal of Experimental Social Psychology*, *79*(November), 200–216.

Exline, J. J., Worthington, E. L., Hill, P., & McCullough, M. E. (2003). Forgiveness and justice: A research agenda for social and personality psychology. *Personality and Social Psychology Review*, *7*(4), 337–348.

Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research*, *32*(December), 370–377.

Fiske, A. P. (1992). The four elementary forms of sociality: Framework for a unified theory of social relations. *Psychological Review*, *99*(4), 689–723.

Foot, P. (1967). The problem of abortion and the doctrine of double effect. *Oxford Review*, *5*, 5–15.

Frazier, M. L., Johnson, P. D., Gavin, M., Gooty, J., & Snow, D. B. (2010). Organizational justice, trustworthiness, and trust: A multifoci examination. *Group & Organizational Management*, *35*(1), 39–76.

Gilligan, C. (1993). *In a different voice: Psychological theory and women's development*. Cambridge, Massachusetts: Harvard University Press.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493–503.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in Experimental Social Psychology*, *38*(06), 69–119.

Greenberg, J. (1987). A taxonomy of organizational justice theories. *The Academy of Management Review*, *12*(1), 9–22.

Greene, J. D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York: Penguin Press.

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834.

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116.

Haidt, J., & Kesebir, S. (2010). Morality. In S. T. Fiske, D. T. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed.) (pp. 797–832). Hoboken, NJ: Wiley.

Halevy, N., Berson, Y., & Galinsky, A. D. (2011). The mainstream is not electable: When vision triumphs over representativeness in leader emergence and effectiveness. *Personality and Social Psychology Bulletin*, *37*(7), 893–904.

Hsee, C. K. (1996). The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, *67*(3), 247–257.

Hsee, C. K., Loewenstein, G. F., Blount, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, *125*(5), 576–590.

Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: On the affective psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 23–30.

Hsee, C. K., Rottenstreich, Y., & Xiao, Z. (2005). When is more better? On the relationship between magnitude and subjective value. *Current Directions in Psychological Science*, *14*(5), 234–237.

Hsee, C. K., Yang, Y., Zheng, X., & Wang, H. (2015). Lay rationalism: Individual differences in using reason versus feelings to guide decisions. *Journal of Marketing Research*, *52*(1), 134–146.

Inbar, Y., Pizarro, D. A., & Cushman, F. (2012). Benefiting from misfortune: When harmless actions are judged to be morally blameworthy. *Personality and Social Psychology Bulletin*, *38*(1), 52–62.

Ingram, P., & Zou, X. (2008). Business friendships. *Research in Organizational Behavior*, *28*, 167–184.

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience*, *10*(5), 551–560.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitutions in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge; New York: Cambridge University Press.

Kahneman, D., & Frederick, S. (2005). A model of heuristic judgment. In K. Holyoak, & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 267–293). New York: Cambridge University Press.

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles

Kant, I. (1785). In J. Timmermann (Ed.), *Groundwork of the metaphysics of morals* (Revised ed.). Cambridge: Cambridge University Press.

Kohlberg, L. (1964). Development of moral character and moral ideology. In L. W. Hoffman, & M. L. Hoffman (Eds.), *Review of child development research* (Vol. 1). New York: Russell Sage Foundation.

Larzelere, R. E., & Huston, T. L. (1980). The dyadic trust scale: Toward understanding interpersonal trust in close relationships. *Journal of Marriage and Family*, *42*(3), 595–604.

Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27–55). Boston, MA: Springer US.

Levine, E. E. (2019). Community standards of deception. Working Paper.

Levine, E. E., Barasch, A., Rand, D., Berman, J. Z., & Small, D. A. (2018). Signaling emotion and reason in cooperation. *Journal of Experimental Psychology: General*, *147*(5), 702–719.

Levine, E. E., Hart, J., Moore, K., Rubin, E., Yadav, K., & Halpern, S. (2018). The surprising costs of silence: Asymmetric preferences for prosocial lies of commission and omission. *Journal of Personality and Social Psychology*, *114*(1), 29–51.

Levine, E. E., & Schweitzer, M. E. (2014). Are liars ethical? On the tension between benevolence and honesty. *Journal of Experimental Social Psychology*, *53*, 107–117.

Levine, E. E., & Schweitzer, M. E. (2015). Prosocial lies: When deception breeds trust. *Organizational Behavior and Human Decision Processes*, *126*, 88–106.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292.

Loewenstein, G., & Small, D. A. (2007). The scarecrow and the tin man: The vicissitudes of human sympathy and caring. *Review of General Psychology*, *11*(2), 112–126.

Lupoli, M. J., Jampol, L., & Oveis, C. (2017). Lying because we care: Compassion increases prosocial lying. *Journal of Experimental Psychology: General*, *146*(7), 1026–1042.

Malle, B. F., Knobe, J. M., & Nelson, S. E. (2007). Actor–observer asymmetries in explanations of behavior: New answers to an old question. *Journal of Personality and Social Psychology*, *93*(4), 491–514.

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709–734.

McFall, L. (1987). Integrity. *Ethics*, *98*(October), 5–20.

Milkman, K. L., Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2011). Using implementation intentions prompts to enhance influenza vaccination rates. *Proceedings of the National Academy of Sciences*, *108*(26), 10415 LP–10420.

Milkman, K. L., Rogers, T., & Bazerman, M. H. (2008). Harnessing our inner angels and demons. *Perspectives on Psychological Science*, *3*(4), 324–338.

Miller, D. T., & Effron, D. A. (2010). Psychological license. When it is needed and how it functions. In M. P. Zanna, & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 43) (pp. 115–155). Cambridge, Massachusetts: Academic Press.

Monin, B., Pizarro, D. A., & Beer, J. S. (2007). Deciding versus reacting: Conceptions of moral judgment and the reason–affect debate. *Review of General Psychology*, *11*(2), 99–111.

Moore, A. K., Lewis, J., Levine, E. E., & Schweitzer, M. E. (2019). Trusting kind friends and fair leaders: How relationships affect the antecedents of trust. Working Paper.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature*, *437*(7063), 1291–1298.

O'Connor, K., De Dreu, C. K., Schroth, H., Barry, B., Lituchy, T. R., & Bazerman, M. H. (2002). What we want to do versus what we think we should do: An empirical. *Journal of Behavioral Decision Making*, *15*(August), 403–418.

Parr, L. A. (2001). Cognitive and physiological markers of emotional awareness in chimpanzees (Pan troglodytes). *Animal Cognition*, *4*(3), 223–229.

Pearce, J. L. (2015). Cronyism and nepotism are bad for everyone: The research evidence. *Industrial and Organizational Psychology*, *8*(1), 41–44.

Pillemer, J., & Rothbard, N. P. (2018). Friends without benefits: Understanding the dark sides of workplace friendship. *Academy of Management Review*, (April) *43*(4), 635–660.

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57–75.

Rand, D. G. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, *27*(9), 1192–1206.

Rottenstreich, Y., & Hsee, C. K. (2001). Money, kisses, and electric shocks: On the affective psychology of risk. *Psychological Science*, *12*(3), 185–190.

Sato, N., Tan, L., Tate, K., & Okada, M. (2015). Rats demonstrate helping behavior toward a soaked conspecific. *Animal Cognition*, *18*(5), 1039–1047.

Shaw, A., DeScioli, P., Barakzai, A., & Kurzban, R. (2017). Whoever is not with me is against me: The costs of neutrality among friends. *Journal of Experimental Social Psychology*, *71*, 96–104.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*(1), 3–22.

Small, D. A., Loewenstein, G., & Slovic, P. (2007). Sympathy and callousness: The impact of deliberative thought on donations to identifiable and statistical victims. *Organizational Behavior and Human Decision Processes*, *102*(2), 143–153.

Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences*, *28*, 531–573.

Taylor, S. E., Pham, L. B., Rivkin, I. D., & Armor, D. A. (1998). Harnessing the imagination. *American Psychologist*, *53*(4), 429–439.

Thomson, J. J. (1985). The trolley problem. *The Yale Law Journal*, *94*(6), 1395–1415.

Tyler, T. R. (2006). Restorative justice and procedural justice: Dealing with rule breaking. *Journal of Social Issues*, *62*(2), 307–326.

Van Kleef, G. A. (2009). How emotions regulate social life. *Current Directions in Psychological Science*, *18*(3), 184–188.

Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science*, *18*(3), 247–253.

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027–1033.

Winningham, R. G., Hyman, I. E., & Dinnel, D. L. (2000). Flashbulb memories? The effects of when the initial memory report was obtained. *Memory (Hove, England)*, *8*(4), 209–216.

Zaki, J. (2014). Empathy: A motivated account. *Psychological Bulletin*, *140*(6), 1608–1647.

Zhang, J. (2015). Joint versus separate modes of evaluation: Theory and practice. In G. Keren, & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (pp. 213–238). Chichester, UK: Wiley Blackwell.

Zhang, T., Gino, F., & Margolis, J. D. (2018). Does “could” lead to good? On the road to moral insight. *Academy of Management Journal*, *61*(3), 857–895.

AUTHOR BIOGRAPHIES

Alexander K. Moore is a first-year PhD student of behavioral science at The University of Chicago Booth School of Business. He holds a bachelor's degree in economics from the University of Chicago. He is broadly interested in the ways that stated rules and policies impact judgment, decision making, and motivation.

David M. Munguia Gomez is a second-year PhD student of behavioral science at The University of Chicago Booth School of Business. He holds a bachelor's degree in psychology from Princeton University. He is broadly interested in the psychology of policy-making, fairness, and causal attribution.

Emma Levine is an Assistant Professor of Behavioral Science and the Charles E. Merrill Faculty Scholar at The University of Chicago Booth School of Business. She holds a bachelor's degree in philosophy, politics, and economics from the University of Pennsylvania and a PhD in Decision Processes from The Wharton School, University of Pennsylvania. Her research seeks to understand how individuals make trade-offs between different values, and how this influences decision making and social perception.

How to cite this article: Moore AK, Munguia Gomez DM, Levine EE. Everyday dilemmas: New directions on the judgment and resolution of benevolence–integrity dilemmas. *Soc Personal Psychol Compass*. 2019;13:e12472. https://doi.org/10.1111/spc3.12472

17519004, 2019, 13, Downloaded from https://compass.onlinelibrary.wiley.com. By University Of Chicago Library- on [21/07/2022]. Re-use and distribution is strictly not permitted, except for Open Access articles