5 Moral Dilemmas

Joanna Demaree-Cotton and Guy Kahane

The demands of morality can seem straightforward. Be kind to others. Do not lie. Do not murder. But moral life is not so simple. We are often confronted with difficult situations in which someone is going to get hurt no matter what we do, in which we cannot meet all of our obligations, in which loyalties come into conflict, in which we cannot help everyone who needs it, or in which we must compromise on important values.

It is natural to describe such situations as *moral dilemmas.* This chapter is about the psychology of how we represent, process, and make decisions about what to do when moral life is difficult in this way. Our first aim is to provide some conceptual clarity on what exactly turns a choice situation into a moral dilemma. Our second aim is to critically survey existing psychological work, providing an overview of some important findings, while raising questions for future research.

5.1 Characterizing Moral Dilemmas

In moral psychology, "moral dilemmas" are typically defined as any situation in which you are required to make a moral trade-off: where a gain in some moral value comes at the cost of some other.[1] This definition has the advantage that it can be easily operationalized for experimental materials. However, this is far broader than the everyday sense of the term. Imagine you've promised to meet your friend for coffee. But on the way, you see a child collapse in the street. Do you simply continue on your way – potentially allowing the child to die but thereby ensuring you keep your promise? Or do you stop to help?

This situation demands a moral trade-off. But, outside of the lab, few would call this a moral dilemma. If you later said: "A child lay dying in the road – but I *did* promise my friend a cappuccino . . . it was a real moral dilemma!" you'd sound ridiculous, even sinister, and out of touch with the demands of morality.

[1] A great deal of current psychological research uses "moral dilemmas" to refer to variants of philosophical "runaway trolley scenarios," also known as sacrificial dilemmas. As we discuss later, sacrificial dilemmas are obviously merely examples of dilemmas and cannot be assumed to be typical or representative.

Part of the reason we wouldn't describe this trade-off as a moral dilemma is because it is relatively trivial to navigate, in at least two senses. Firstly, it's *epistemically* easy to figure out what the right thing is to do. A child's life is much, much weightier than a coffee date, so the obligation to help clearly outweighs the promise to your friend. Although the promise is what philosophers would call a *pro tanto* moral reason to keep walking, clearly the morally right thing *all things considered* is to help. Secondly, the trade-off is *psychologically* easy to manage, in the sense that choosing to help is not inherently emotionally aversive, and you shouldn't be kept up at night ruminating guiltily over this choice. Indeed, it seems there would be something normatively amiss if you did find the choice psychologically difficult: Breaking a promise in this kind of situation just isn't *that* big of a deal, morally speaking. Let's call situations that demand moral trade-offs that are epistemically and psychologically straightforward to resolve "trivial moral trade-offs."

By contrast, colloquially the term "moral dilemma" is reserved for more difficult situations. Let's call these "genuine" moral dilemmas or simply "moral dilemmas." A genuine moral dilemma can arise if moral considerations that are sufficiently weighty pull in the direction of incompatible courses of action. In some such situations, it can be difficult to know which outweighs the other. In other situations, you might know what the right thing is to do, all things considered; nevertheless, the competing moral considerations are sufficiently weighty so that whatever you do, you have to do something that at least *ordinarily* would be very wrong, even if it's not wrong here. This might be psychologically difficult to navigate even if it's not epistemically difficult: It may be uncomfortable, unpleasant, or even highly distressing, and give rise to regret or guilt. For instance, imagine you're a soldier who is forced to leave injured civilians to die, because attending to them would put your fellow soldiers' lives at risk and break direct orders. Even if you are sure you're doing the right thing, it's still distressing and heart wrenching (cf. Molendijk, 2018; Shortland & Alison, 2020); and, in normal circumstances outside of wartime, leaving people to die in the street would be a terrible thing to do.

And this is what makes the situation a genuine moral dilemma even though you know what you ought to do, all things considered. To be precise, what makes the situation a genuine moral dilemma is not that you *in fact* experience this psychological conflict; rather, it's the normative fact that, in some sense, you *ought to*. Insofar as you're a good person, it's appropriate to place heavy moral and emotional weight on the fact that your actions dictate the death of innocent civilians.[2]

In light of these considerations, we offer a definition of genuine moral dilemmas as situations in which:

[2] This is so even from a utilitarian point of view. The death of innocent civilians counts negatively in the utilitarian calculus. Moreover, this *normally* leads to worse consequences overall. Because of this, many utilitarians would say that at least an initial reluctance to make such choices will tend to make one the sort of person who maximizes good consequences, even if one ought to override this reluctance in specific situations.

(1) an agent's moral reasons conflict, such that any course of action open to them has at least some strong moral reason against it;
(2) this conflict is such that there is some sense in which it is morally appropriate to feel motivationally conflicted about what to do, and to feel psychologically conflicted about any given course of action they end up choosing.

Thus, the definition of moral dilemmas is inherently normative: it makes reference to the moral reasons that actually apply in a situation, and to the psychological response that would be morally *appropriate*. Note, that, in philosophical terms, the claim that sometimes there are strong *reasons* for and against a choice is not a claim about whether the agent engages in reason*ing*; it's a claim about the demands of morality, not the agent's psychology. So having to choose between innocent lives gives rise to strong conflicting reasons and is thus a moral dilemma even if the person deciding happens to be too heartless, jaded, or traumatized to actually *experience it* as such; a trivial trade-off is *not* a moral dilemma just because someone is irrationally wracked with unreasonable guilt.

But our normative account also tells us something about what, descriptively, it is to experience something *as* a moral dilemma – it is to experience a kind of moral psychological conflict arising from having to make a choice when it *seems* to you that moral reasons conflict. Thus, we expect there to be a number of descriptive markers that distinguish such situations from other moral decisions, including a tendency to evoke uncertainty, feelings of moral conflict, as well as what the philosopher Bernard Williams (1965) called "moral residue" – lingering negative affect later associated with the decision taken (see Figure 5.1).

Thus, even if the notion of a moral dilemma is best understood in normative terms, for the purposes of psychological research it may suffice to study what most people in a certain population experience *as* a moral dilemma, or what would count as a moral dilemma relative to certain commonly accepted moral norms.

There is a great deal of psychological research investigating how people respond to moral dilemmas. Most of this research in fact studies moral

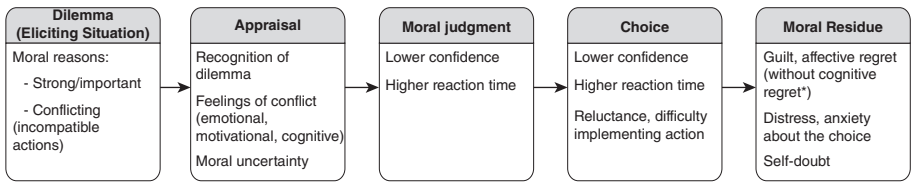

Figure 5.1 *Simplified illustration of the components and characteristic psychological markers of genuine moral dilemmas. Although the arrows indicate the paradigmatic order in which these steps unfold, there may be recursive relationships between the components (e.g., if one's choice leads one to revise one's moral judgment in order to reduce cognitive dissonance).* * *Experiencing markers of moral residue in the absence of cognitive regret is a special marker of an ultimate moral dilemma.*

judgments about hypothetical situations – which is rather different than actually facing a dilemma (Bostyn et al., 2018). Nevertheless, evidence suggests that at least some dilemmas employed by psychologists are experienced as genuine dilemmas: Moral dilemmas involving serious conflicts lead to subjective feelings of moral conflict, decision difficulty, longer reaction times, reduced confidence, and increased negative emotions (such as fear of making the wrong choice) compared to trivial trade-offs and straightforward moral decisions (Bialek & De Neys, 2016; Hanselmann & Tanner, 2008; Mandel & Vartanian, 2008), and they are associated with activity in brain regions involved in conflict detection and resolution (see Greene, 2008). The experience of moral conflict has cognitive and emotional components (Mata, 2019). Cognitive conflict involves feeling divided about what to do and feeling drawn to conflicting considerations and options (i.e., epistemic difficulty), and it predicts judgments that neither option is "totally acceptable" or "totally unacceptable" (Mata, 2019). Emotional conflict involves negative feelings and emotional discomfort (and is therefore a key component of what we termed psychological difficulty).

In these studies, the extent of conflict between moral values is manipulated by the experimenters. But a study by Krosch et al. (2012) illustrates how experiences of moral conflict depend on subjective commitment to competing moral values or reasons. Participants were presented with dilemmas and asked which values they would use to guide their decision and what choice they would make. The more a participant *also* endorsed values that supported the *opposite* choice to the one they favored, the more difficult they found the choice, and the more they expected to worry that they had made a mistake.

So far, we discussed the broad sense that psychologists give to "moral dilemma," which includes trivial moral trade-offs, and provided a characterization of "genuine" moral dilemmas, which only include more difficult trade-offs. We should also mention that when philosophers use the term "moral dilemma," they often refer to a much narrower category: situations in which *all* options are morally wrong. These are special kinds of genuine dilemmas. Let's call these "ultimate" moral dilemmas (see Figure 5.2). Philosophers disagree about

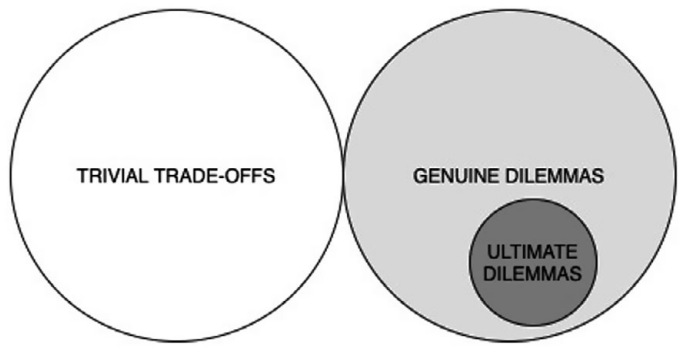

Figure 5.2 *We distinguish trivial trade-offs from genuine moral dilemmas. Ultimate dilemmas are a possible subset of genuine dilemmas.*

whether there are ultimate moral dilemmas. Some philosophers say there are – that some situations tragically force you to do something wrong no matter what you do (Barcan Marcus, 1980; Hursthouse, 2001; Nussbaum, 2000). Others hold that it's *always* possible to choose something that's not wrong, that is, morally permissible (Conee, 1982).

Consider the example of "Sophie's choice," where a mother is forced to choose one of her two children to be murdered, or else faces having them both murdered. On one view, it's an ultimate dilemma: Part of the tragedy is that Sophie is forced do something morally terrible, because it's absolutely morally wrong to select one of your own children for death, even if the alternative is even worse and they would have died anyway. By contrast, other philosophers would argue that although there is a genuine moral conflict, it's not an *ultimate* dilemma. According to this view, Sophie is doing the best she can under the circumstances, and so she isn't doing anything wrong when she sends one child to their death – even if it feels that way.

While psychological research on so-called moral dilemmas sometimes includes cases like Sophie's choice that some see as ultimate dilemmas, much of the psychological literature involves vignettes inspired by philosophical thought experiments – most famously, trolley problems – that would *not* normally be considered ultimate "moral dilemmas." For instance, most philosophers think that if you face a choice between allowing a runaway train to kill five workmen and pulling a switch to divert it onto a side-track where unfortunately there is one workman, you have the opportunity to do something that's morally right: pulling the switch. If you think that, then you don't think that it's an *ultimate* moral dilemma, where you're condemned to act wrongly no matter what.[3] We will consider whether there is evidence that nonphilosophers treat some cases as ultimate dilemmas later in the chapter.

5.2 The Psychological Source of Moral Dilemmas

We have offered an account of moral dilemmas and of what generates them at the normative level – a conflict between weighty opposing moral reasons – as well as of their key experiential markers. We now turn to possible accounts of the underlying psychological machinery: Why are certain kinds of moral trade-offs experienced as moral dilemmas that are epistemically or just psychologically difficult to resolve, whereas others are trivial?

[3] Why should anyone feel that Sophie's choice is an ultimate dilemma but not feel the same way about trolley-style dilemmas? After all, both involve choices between human lives. One possibility is if you think that *actively choosing* someone to die is intrinsically wrong, especially if it's your own child, but *allowing* someone to die (as is possible in some trolley-style dilemmas), especially a stranger, is not.

5.2.1 Emotion versus Reason and Conflicting Processes

To put one influential answer in slogan form: Moral dilemmas arise when reason tells us to do one thing, but the heart tells us to do another (Greene, 2008). According to classic dual-process theory, "System 1" processes (which are fast, automatic, and emotion-driven) run in parallel to "System 2" processes (which involve slow, deliberative reasoning), and each process type produces independent inputs into moral judgment. In spite of disagreement about how exactly to characterize the dual processes, the idea is that moral dilemmas arise when dissociable psychological processes produce conflicting outputs. As Cushman and Greene argue: "When two such processes yield different answers to the same question, that question becomes a 'dilemma.' No matter which answer you choose, part of you walks away dissatisfied" (Cushman & Greene, 2012, p. 267).

There is a wealth of psychological and neuroscientific evidence that dual processing plays a role in moral dilemmas. Famously, Greene and colleagues (see Greene, 2008, 2014) drew on philosophical thought experiments to create vignettes designed to pair emotionally evocative action types with positive consequences. The action generally involved sacrificing one person in order to obtain benefits for others; the types of situation depicted by these vignettes have come to be known as "sacrificial dilemmas." In "difficult dilemmas," the emotionally evocative action – the sacrifice – led to greater positive consequences, and thus had strong moral reasons in its favor. For instance, in Crying Baby, villagers are hiding when one of their babies begins to cry, alerting enemy soldiers to their location. The mother has to decide whether to smother and kill her baby to prevent all of their deaths at the hands of the soldiers (the "utilitarian" or "sacrificial" option) or whether to refrain from smothering the baby (the "deontological" or "nonsacrificial" option). By contrast, in a so-called easy dilemma (an oxymoron on our definition!) a new mother is deciding whether or not to kill her baby, simply because it's unwanted.

Greene and colleagues reported that "difficult" dilemmas activated neural areas associated with conflict and cognitive control (though see Baron & Gürcay, 2017, and Sauer, 2012, for criticism). In another study, they found that placing subjects under cognitive load selectively delayed pro-, but not nonsacrificial, judgments (see Greene, 2008). Such evidence is interpreted by Greene and others as showing that the psychological difficulty of moral dilemmas arises because of a conflict between the experience of automatic, emotional aversion to certain actions, and reasoned, deliberative recognition that this action will have the best outcome (for supporting evidence, see Greene, 2014; Patil et al., 2021) – thereby offering a purely *psychological* explanation of what, in the ethical debate, was characterized as a normative conflict between opposing moral *principles*.

Greene's dual-process model led to a proliferation of research examining dilemmas that pit emotionally evocative, harmful actions against some seemingly greater good. Indeed, it has led many to analyze *all* moral dilemmas in

terms of competing dual processes. For instance, Bartels and colleagues write that there is "an underlying agreement... [that] moral dilemmas exist because we have diverse psychological processes available for making moral judgments, and when two or more processes give divergent answers to the same problem, the result is that we feel 'of two minds'" (Bartels et al., 2015, p. 491).

5.2.2 Sacred Values and Tragic Trade-offs

Certainly, if our heads and our hearts are in agreement – or intuition and deliberation, or any other pair of processes – that makes for an easy decision, whereas conflict between processes requires resolution. But a focus on conflicts between values based in opposing processes may be an artifact of the "sacrificial" dilemmas that have dominated the literature, which have been designed precisely to pit reason-based processes against affective ones. But some kinds of moral conflict are especially difficult to resolve precisely because *similar* moral values are competing.

Consider the research on "sacred values" and "tragic trade-offs" (Tetlock et al., 2000). A "sacred value" (or "protected value"; Baron & Spranca, 1997; see also Chapter 8 in this volume) is a moral value that is regarded, in some sense, as non-negotiable and absolute – such as human lives, justice, or protecting nature. Whether a value is sacred is measured by explicit high levels of agreement with statements such as "it's something that we should not sacrifice, no matter the benefits"; "you can't quantify it with money"; or, "it involves principles that we should defend under any circumstances" (Hanselmann & Tanner, 2008). "Taboo" trade-offs involve sacrificing a sacred value for a so-called secular value – for instance, denying someone life-saving treatment because it costs too much money – and are widely regarded as morally wrong. "Tragic" trade-offs, by contrast, pit "sacred" values against one another.

Tragic trade-offs are genuine moral dilemmas in our sense; some may even be "ultimate" dilemmas. Take the "tragic trade-off" of a hospital director who must choose which of two ailing little boys will receive a life-saving liver transplant, when there is only one available (Tetlock et al., 2000). Such a case is difficult to resolve because serious and symmetrical moral reasons pull the agent in incompatible directions: Either way, the director is forced to deny a child a life-saving transplant. Tragic trade-offs can also arise when different sacred values are pitted against one another, such as protecting the environment versus ensuring safe working conditions. Participants find trade-offs between sacred values subjectively difficult (Hanselmann & Tanner, 2008), perceive agents willing to make them as untrustworthy and immoral (Everett et al., 2016; Uhlmann et al., 2013), and express moral outrage if others find such dilemmas easy to resolve (Tetlock et al., 2000). In this last respect, ordinary attitudes concur with arguments by certain ethicists that a virtuous person faced with a tragic dilemma acts only with great hesitation, reluctance, and regret (Hursthouse, 2001).

It's highly implausible that the feeling of moral conflict and difficulty to which such dilemmas give rise is due to a conflict between values rooted in

emotion, on the one hand, and values rooted in reason, on the other. In the liver transplant case, the very same value applies to two incompatible courses of action. The moral loss of allowing either little boy to die is presumably processed in the same way by the same mechanisms; consequently, both options are going to feel bad for similar reasons.

The "sacred values" framework typically contrasts nonmoral values with sacred values, neglecting the possibility of nonsacred moral values. Yet, the possibility that not all moral values are seen as sacred is important if we are to use this framework to explain moral dilemmas, since not all moral conflicts give rise to genuine dilemmas. For instance, if promise-keeping is a moral but not a sacred value, while saving a life is a sacred value, that would explain why the coffee versus child emergency scenario is a trivial moral trade-off rather than a genuine moral dilemma. An alternative hypothesis would be that "sacredness" comes in degrees; there could be a hierarchy of sacred values (cf. Shortland & Alison, 2020) such that moral dilemmas arise when there are conflicts between values of sufficient, or sufficiently similar, sacredness. On this hypothesis, the coffee promise versus child emergency scenario would fail to be a genuine moral dilemma, not because promise-keeping isn't perceived as "sacred," but because everyday promises are lower on the hierarchy than the sacred value of saving lives.

5.2.3 Value Commensurability and Absolute Constraints

How exactly should we understand the reluctance people have to violate sacred values? The language used to describe and measure sacred values – "absolute," "unquantifiable," "nonnegotiable" – might suggest that sacred values are regarded as (1) the subject of absolute moral constraints; (2) strictly *incommensurable* to each other, and also infinitely greater than nonsacred values. If people regard sacred values as the subject of absolute restrictions and/or incommensurability, this could explain why certain conflicts are especially difficult to resolve.

Indeed, in philosophy, value incommensurability – the idea that some values can't be compared and measured on a common scale (Chang, 1997) – is often taken to be intimately linked to *ultimate* moral dilemmas. For instance, you might think that the values of loyalty and justice are incommensurable: that they cannot be weighed against one another, so you can't "make up" for disloyalty by making things more just. Similarly, maybe there's just no fact of the matter as to how much happiness the death of an innocent would have to bring about for it to be "worth it." Such incommensurability could explain the special tragedy of ultimate moral dilemmas: You are forced to make a sacrifice that cannot be made up for by any of your choice's benefits, because these benefits cannot even be compared to the disvalue of the sacrifice (Nussbaum, 2000). In a different way, absolutism about moral prohibitions – that is, the idea that doing certain types of things is *always* morally wrong – could also give rise to ultimate moral dilemmas, as you may end up with a choice between two options, both of which are absolutely wrong to do.

Participants appear to endorse explicit claims of absolutism and incommensurability with regards to sacred values. And moral conflicts can therefore seem irresolvable. For instance, Shortland and Alison (2020) interviewed military veterans who described their experience of real-life dilemmas in combat settings. When conflicting moral values were *both* regarded as nonnegotiable, sacred values, this led to great epistemic and psychological difficulty, revealed by "redundant deliberations," "looping" cognitions, and great difficulty reaching a resolution. These veterans appear to have struggled to resolve the conflict either because both options were perceived to be absolutely prohibited and/or because they struggle to make a meaningful comparison between the conflicting values.

However, other evidence suggests that people don't treat conflicting values as incommensurable in the sense that they are unable to compare them. For instance, in many dilemmas people judge that it's better for one person to die if this is necessary to save sufficiently many others (e.g., Nichols & Mallon, 2006). Furthermore, these kinds of comparative judgments can be seen without the need to resort to large numbers if the *act* of sacrificing is disambiguated from comparisons regarding the *amount* to be sacrificed: Even if people are absolutely committed to avoiding the act of, for example, causing a child to die, they would still make trade-off judgments when it's impossible to avoid this act (Berman & Kupor, 2020).

Thus, conflicts between sacred values may be better captured in terms of commitment to a kind of absolutism: the idea that the *acts* involved (e.g. allowing children to die; razing acres of rainforest) are always wrong. Notice that the judgment that such choices involve facing an ultimate dilemma is compatible with the *comparative* moral judgments that are normally measured in research on moral dilemmas – for instance, that it's *better* to choose option A rather than B, or that choosing A *rather than B* is the "right" choice. Just as it's possible for someone to judge that they prefer Annie to Beth, and yet also believe that both Annie and Beth are bad people, it may be that in some dilemmas people judge that it's right to choose option A over option B, and yet still judge that either way you do something wrong (Hursthouse, 2001). This fits the finding that, when given the option of doing so, people sometimes judge both options in sacrificial dilemmas as wrong (Kurzban et al., 2012, Study 1).

If this is right, then one way of understanding the special difficulty of some sacrificial dilemmas would be not in terms of a conflict between reasoning-based values and emotion-based values, but as a special kind of tragic trade-off involving a conflict between sacred values that give rise to incompatible absolute obligations (e.g., when an absolute obligation to save lives conflicts with absolute prohibitions against intentional murder).

5.2.4 Automatic Aversion

We have argued that the values involved in moral conflict can, but needn't, have their source in different process-types or "systems." But we are still left with a

puzzle: Why do we feel motivationally conflicted even when we recognize it's the best choice under the circumstances?

Evidence suggests that performing certain types of actions feels intrinsically wrong and unpleasant *automatically*, even if doing so is justified or even harmless. For instance, Cushman and colleagues showed that people have strong automatic aversions to performing pretend harmful actions, such as shooting someone with a fake gun, as indicated by physiological measures (Cushman et al., 2012).

Automatic aversion is not sufficient for the experience of moral dilemmas (or pretend harmful actions would be considered wrong). Yet, combined with the perceived violation of moral requirements (cf. Nichols & Mallon, 2006) – perhaps requirements that are sacred or absolute – the automaticity of aversion explains the feelings of conflict characteristic of genuine dilemmas: If you are forced to perform a harmful action, whatever you do will automatically feel aversive, wrong and distressing even if you know it's the right thing to do under the circumstances. (This could also contribute to persistent "moral residue"; see Section 5.4).

While some theorists suggest that aversion to physically harming others is innate (Greene, 2008), learning may play a role in feeling automatic aversion to a wider range of actions. For instance, Graham and colleagues' moral foundations theory suggests that we have an innate propensity to feel aversion to violations of different moral foundations (including, for example, care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, and sanctity/degradation) which are then developed depending on which values are emphasized in our culture (Graham et al., 2013).[4] More recently, psychological and neuroscientific research has drawn on computational models to theorize how we might acquire moral aversion through reinforcement learning (Crockett, 2013; Cushman, 2013), where associating an action with harmful consequences or moral condemnation leads to automatic negative representations of that action in the future. Although much research has focused on physically harmful actions, such learning mechanisms could potentially explain aversion to a wider range of action types, like lying, stealing, showing disrespect, emotional harm, or purity violations (Cushman, 2013; Nichols, 2021). Indeed, it's possible that learning mechanisms could explain automatic aversion to *the breaking of a moral rule* as such. This would contribute to explaining the experience of moral dilemmas more broadly in which one is required to violate some moral rule where breaking that rule is typically associated with severe consequences or condemnation.

As well as aversion to performing certain acts, we also experience automatic negative affective reactions to the harm that befalls others. So-called outcome aversion is linked to empathic concern and is associated both with reluctance to harm and to fail to prevent harm (Jack et al., 2014; Reynolds & Conway, 2018).

[4] Greene (2014) allows that "deontological" aversion may be a product of cultural learning.

This suggests that the unpleasantness of moral dilemmas arises not just from an aversion to performing certain (e.g., violent) acts, but also from empathic aversion to the harm that would befall others as a consequence of our choice (see Chapter 11 in this volume for a detailed treatment of empathy).

5.2.5 The Sources of Moral Dilemmas: Summary

Moral dilemmas arise whenever important values exert motivational pull in conflicting directions. Sometimes, a value based in System 1 conflicts with a value processed and prioritized by System 2 (Greene, 2014). However, further research suggests that moral dilemmas can arise even when there's no such inter-system competition. This could happen due to conflict between values that are both processed by System 2 (which might give rise to cognitive conflict, and thereby to emotional conflict); or due to conflict between two values that are both processed by System 1 (which might give rise to emotional conflict first, recognition of which gives rise to cognitive conflict). Nevertheless, the automaticity of the aversion felt toward performing bad actions, to bringing about bad outcomes, to violating sacred values or important moral rules, or simply to doing something perceived as wrong, may be a key part of the explanation for why motivational conflict persists despite reflective beliefs that no better option is available, and it's in this more qualified respect that dual processes may be central to the experience of moral dilemmas.

5.3 How Do We Resolve Moral Dilemmas?

The question of how we resolve moral dilemmas involves three interrelated considerations. The first is *content*: What values or considerations do people bring to bear to resolve moral dilemmas? The second is *process*: What types of psychological processes are involved, and how do they interact? The third is the resolution that is reached, what we might call the *verdict*: What does the agent conclude about what they ought to do?

5.3.1 Content: Which Values Are Brought to Bear?

According to Greene's dual-process model (2008), these questions are intertwined: It says that the values one brings to bear, and thus one's all-things-considered resolution, depends critically on the psychological process used. On this view, moral dilemma resolution is the result of a battle between a default affect-driven response that promotes "deontological" aversion to certain options, and a reasoned response that promotes "utilitarian" preference for whichever action has the best consequences. If we cannot engage in reasoned resolution – because we lack the opportunity or capacity, or because our emotional responses are too strong – then System 1 emotional intuition dominates, and we simply pick whichever action feels least aversive.

Contrary to the dual-process model's claim of a strong link between intuitive resolutions and "deontological" prohibitions, we already saw that automatic, affect-driven responses can include aversion to bad outcomes. Thus, even if an agent responds on a purely intuitive basis, this doesn't always mean they will fail to prioritize better consequences. Indeed, the utilitarian solution to moral dilemmas is sometimes more intuitive than the deontological one (Kahane et al., 2012; for criticism, see Paxton et al., 2014; for a reply, see Kahane, 2014, p. 16). Additionally, "utilitarian" judgments can be made effortlessly even when participants are under cognitive load when the number to be saved is very large (Trémolière & Bonnefon, 2014); and the tendency to choose actions promoting the greatest consequences for distant others (impartial beneficence) is preserved when responding intuitively (Capraro et al., 2019). Indeed, Bago and De Neys (2019) used a two-response paradigm and found that the majority of participants endorsing the sacrificial ("utilitarian") option as their final judgment after reflection had already favored this option when initially required to make a quick, intuitive judgment.

It's similarly unclear that reasoned resolutions imply specific normative responses. Thus, which kinds of responses are intuitive, and which are deliberative, will vary for different contexts and individuals. While substantial evidence links deliberative reasoning to choosing the best consequences (e.g., Greene, 2008; Patil et al., 2021), reasoned resolutions are not limited to "utilitarian" cost-benefit analysis. Sometimes reasoning is used to override cooperative impulses to make more self-interested choices (Rand et al., 2012); in other contexts reasoning is used to promote deontological concerns over self-interested or utilitarian impulses (Knoch et al., 2006; see Kahane, 2015, note 8, for discussion). Reasoning can also involve considering other people's moral arguments (Paxton et al., 2012), or more domain-general techniques, such as consistency reasoning (Paxton & Greene, 2010), where one reflects on similar, less difficult cases in order to decide what ought to be done now. These reasoning techniques are unlikely to prioritize one kind of normative stance. Consistent with this claim, innovative paradigms that track participants' preferences over time (Gürcay & Baron, 2017; Parker & Finkbeiner, 2020) suggest that reasoning about sacrificial dilemmas can lead participants to change their verdict in different directions: Some participants move away from an initial "utilitarian" inclination toward a reasoned "deontological" solution, while other participants show the opposite trajectory.

Thus, reasoning about moral dilemmas can include relatively sophisticated endorsement of "deontological" rights, duties, or rules (e.g., Cushman et al., 2006; Gamez-Djokic & Molden, 2016; Holyoak & Powell, 2016; Körner & Volk, 2014).[5] For instance, Gamez-Djokic and Molden (2016) found that inducing a focus on goals relating to duties and responsibilities increased deontological judgments in sacrificial dilemmas. This increase was not

[5] See Holyoak and Powell (2016) for further discussion of deontological principles in moral reasoning.

associated with empathic concern or affect; instead, it seemed to be explained by reasoning.

Reasoning about deontological values can also concern "role-based" obligations, such as special duties toward close family and friends. Such values can play a role in resolving dilemmas where personal loyalties conflict with rules or the greater good (e.g., deciding whether to report one's own brother to the police; Lee & Holyoak, 2020), or when multiple personal loyalties conflict (e.g., deciding whether to sacrifice one family member to save other family members; Kurzban et al., 2012).

Religious values may also be brought to bear. For instance, belief that a moral rule is grounded in God's moral authority is associated with the judgment that one mustn't break that rule to promote better outcomes (Piazza & Landy, 2013), and the influence of such beliefs on dilemma judgments is reduced if religious participants cannot reflect because of time pressure or cognitive load (McPhetres et al., 2018).

These examples of moral values that may feature in reason-based resolutions are surely not exhaustive (see, e.g., Graham et al., 2013). Thus reasoning processes are unlikely to be linked to a single specific type of normative resolution across dilemmas generally (Kahane, 2012). Instead, the difference between resolving moral dilemmas intuitively or via deliberation may be better characterized in terms of the number and complexity of moral considerations one brings to bear. This is the central claim of Landy and Royzman's (2018) "moral myopia model," according to which purely intuitive responses will consist in singularly attending to one salient aspect of a moral problem, while having motivation and opportunity to engage in reasoning will lead to more complex, integrative responses. In our view, intuitive responses are not literally "myopic" – even automatic intuitions are sensitive to the presence of moral conflict (Bialek & De Neys, 2016), and thus responsive to a range of moral factors. But it does seem that reasoning often allows us to bring to bear a greater number of factors and to evaluate their relevance in more complex ways, including factors that are not immediately salient or intuitive (Kahane, 2012). For instance, Moore et al. (2008) found that participants' moral judgments regarding sacrificial moral dilemmas were sensitive to whether the one to be sacrificed would have died anyway, but only for participants with high working memory capacity – suggesting that reasoning was needed to give this factor weight in dilemma resolution.

Beyond "utilitarian" and "deontological" values, perception of *moral character* may influence moral dilemma resolution. Those who make pro-sacrificial judgments in moral dilemmas are judged to be less moral, less trustworthy, less praiseworthy, less caring, more self-interested, and generally worse than those who choose the nonsacrificial resolution (e.g., Critcher et al., 2020; Everett et al., 2016). Consequently, people may bring concerns about moral character to bear when deciding how to resolve a moral dilemma. Recent studies support this prediction. Participants are more likely to reject the sacrificial option in sacrificial dilemmas (like the Crying Baby dilemma) when told they are being

assessed for emotional competence (Rom & Conway, 2018) or that they are being observed by others, and this tendency is associated with increased sensitivity to words concerning "warmth"-related personality traits (Lee et al., 2018).

There are a number of (mutually compatible) ways that concerns about moral character could affect moral dilemma resolution. The first is instrumental: People just want to be *perceived* as good, and they self-interestedly moderate their choices to preserve their social reputation. The second is that reasoning about others' perceptions is used as a heuristic: Refraining from acts that make you *look* like a bad person might be a reliable heuristic for *in fact* doing the right thing. A third possibility is that concerns about character play a more direct role in moral reasoning: People may choose actions on the basis of what kind of person they would *be* if they performed these acts. Consistent with the latter hypothesis, Reynolds et al. (2019) found that assessing sacrificial dilemmas in front of a mirror increased tendencies to reject sacrificial actions (actions that participants tend to associate with immoral character; Uhlmann et al., 2013), even though it did not increase self-reported concern about how others would evaluate them for their decision.

To the extent that people consider how to act well or virtuously when comparing options, their reasoning would echo the philosophical tradition of virtue ethics (as opposed to deontology or utilitarianism), which grounds morality in what it is to be a good person, and according to which the right thing to do is what a virtuous person would characteristically do (e.g., to act honestly, kindly, courageously, etc.; Hursthouse, 2001). The distinctive psychological difficulties associated with moral dilemmas could therefore lie in the perception that one must do something seemingly inconsistent with virtue and, indeed, one's own self-conception (Strohminger & Nichols, 2014), as moral dilemmas require you to do something that *typically* only a vicious (callous, dishonest, disloyal, etc.) person would do (Hursthouse, 2001, p. 74).

5.3.2 How Do We Weigh Competing Values?

On some views, the psychological conflict experienced in moral dilemmas often does not reflect conflict between moral values that are viewed as genuine; rather, they involve an immediate, "pre-potent" emotional response that effortful reflection reveals is a kind of moral illusion, pulling us away from the moral values that we reflectively endorse (Greene, 2008).

By contrast, much of the research we've reviewed so far suggests that most people do recognize genuine competing values in sacrificial dilemmas and in other contexts. Most people therefore experience many situations as genuine moral dilemmas in our sense – cases where conflicting values need to be weighed against each other.

Some philosophers have offered normative accounts of how conflicts between values can be resolved. On one especially simple model, different values or principles are given different fixed weights that can provide a ranking of

"valuableness" that we can use to resolve moral dilemmas when those values conflict (e.g., Chang, 1997).

In psychology, the so-called conflict model of moral dilemma resolution suggested by Gürcay and Baron (2017) can be read along such lines. Gürcay and Baron argue that agents weigh up the competing alternatives until an option reaches a threshold of support, leading to its endorsement. Reaching this threshold can be seen as an additive process that depends on the background weight the agent attaches to different values, and on the trade-off of those values required by the dilemma. This model sits easily with the view that no value is treated in a qualitatively different way to others (e.g., on the basis of having roots in emotional or nonemotional processing, or on the basis of being a "default" value). In this respect, such a model complements the view that the competing moral values at stake in moral dilemmas are commensurable – contrary to certain interpretations of "sacred values" discussed earlier – at least in the descriptive sense that we do add up and weigh sources of moral value and disvalue to form a single overall valuation.

A natural way of interpreting this model is that people approach moral dilemmas already equipped with priorities or "weights" attached to different values that they simply apply to the situation at hand. For example, a "deontological" responder is one who places a larger negative weight on intentionally killing than the positive weight placed on saving a life. However, many philosophers have argued that the weight of different moral values and principles isn't fixed in this simple way but always depends on the context. We can't, then, come ready with a specific weight attached to a general value like "don't harm others" or "help people if you can." This idea is central to the pluralist moral system proposed by the philosopher W. D. Ross (1930). On Ross's framework, there are multiple "prima facie," nonabsolute moral duties. They don't have a fixed "weight," so there's no general rule that determines how to resolve conflicts between them; instead, which duty outweighs the other is determined according to the particular moral situation at hand, by exercising moral judgment – where this involves a kind of holistic pattern perception, not appealing to a higher-order rule or principle. In psychological terms, such an all-things-considered judgment could be seen as a System 1 intuition about a conflict between principled duties registered at the System 2 level.

Similarly, according to Aristotelian virtue ethics, the virtuous person must exercise "practical wisdom" – a kind of intuitive moral expertise developed over time through habit, reflection, and experience – in order to determine what ought to be done (Hursthouse, 2001). For instance, although there's no principle that prioritizes, for example, honesty over kindness, the virtuous person can "tell" that they ought to tell a white lie when their embarrassed friend asks whether the spill on their dress is very noticeable. Although it's unclear what exactly this "practical wisdom" consists of, it might be interpreted as involving System 1 intuitions (particularly those developed by learning). On the other hand, according to virtue ethics, those who have not yet acquired the intuitive skill of the truly virtuous might have to resolve the conflict using

reasoning (System 2), by reflecting on, for example, what moral exemplars have done in similar situations.

On these ethical views, determining the relative weights of competing values is primarily an intuitive process. What might this intuitive process look like in more detail? On one possible model, each moral factor is associated with an emotional response, and the factor that wins out is the one associated with the strongest emotion. For instance, we may simply endorse the option that feels the least emotionally aversive. Supporting this, many studies have found that the strength of emotional responses often predicts people's judgments in moral dilemmas (see Miller & Cushman, 2013, for an overview). One interpretation of these findings is that participants decide to forego a greater good (e.g., saving more lives) because sacrificing someone evokes stronger negative emotions.

However, other studies have found that the strength of various emotional responses has only limited correlation with overall judgments (Horne & Powell, 2016). This suggests that while emotions play a role in dilemma resolution, people tend to reason in a more complex way than simply choosing whichever option feels the least bad. Even when more complex emotions than aversion are taken into account, their strength appears to underdetermine judgment. Royzman et al. (2011) asked participants to evaluate dilemmas that forced a protagonist to choose between committing incest (a disgust-related violation) and allowing great harm to befall someone (a sympathy-related violation). Participants' moral judgments were consistently predicted by which action they believed would have the greatest costs for everyone. By contrast, their judgments were *not* significantly related to the levels of emotional distress or the disgust or sympathy evoked by contemplating the different options, or individual differences in participants' dispositional disgust sensitivity or dispositional sympathy.

5.3.3 The Verdict

Reasoning about complex moral considerations doesn't only affect which option is ultimately favored. It can also result in a different moral stance – a resolution that is, in a sense, more nuanced and forgiving. Royzman et al. (2015) found that performance on the cognitive reflection test (a marker of deliberative capacity) predicted the judgment that *both* options in a moral dilemma were morally permissible (thus, that neither was morally required over the other). In this vein, future research may consider moral dilemma resolution not only in terms of which action is preferred overall, but also in terms of different types of moral attitudes toward *both* actions.

Another question concerns what function reasoning plays beyond influencing all-things-considered moral judgments. After all, sometimes the intuitively preferred option is also reflectively endorsed (Bago & De Neys, 2019). One possibility is that people delay their response just to signal to others that they are moral and take the dilemma seriously (cf. Tetlock et al., 2000). But reasoning processes also allow people to explore justifications for their choice,

which can be important for justifying oneself to others later on (Paxton & Greene, 2010). Although this process remains unexplored, it could also have downstream implications for moral emotions like guilt and regret.

Finally, the dominant lab-based research paradigm that asks participants to evaluate two mutually exclusive options (e.g., sacrifice vs. don't sacrifice) artificially limits the resolution process. In real life, option sets are rarely limited to two mutually exclusive options restricted to a single point in time. Instead, agents can often undertake a series of actions that change the nature of the moral situation in more complex, creative ways. For instance, a "deontological" agent's recognition of the moral cost of refraining from sacrificing the one could manifest itself in other decisions – such as seeking other ways to save the five, trying to find ways to prevent such situations arising again in the future, and making amends for the harm caused to the five. Investigating more complex resolution options in future research could provide a more nuanced picture of dilemma resolution.

5.4 Moral Residue and the Aftermath of Dilemma Resolution

Most psychological research has focused on how moral dilemmas are resolved. But some of the central psychological characteristics of genuine moral dilemmas only manifest themselves after the resolution stage in "moral residue" or "moral remainder" – a sense of moral unease, remorse, or guilt that lingers no matter which option is chosen. Some philosophers further argue that, far from being irrational, these lingering moral feelings are sometimes morally appropriate, and even indicate you faced an ultimate moral dilemma – that you were forced to do something morally wrong (Barcan Marcus, 1980; Williams, 1965). Indeed, such feelings may be part of what it is to be a good, loving, loyal person who has been forced to do something terrible by tragic circumstances (cf. Hursthouse, 2001, pp. 73–77).

Goldstein-Greenwood and colleagues (2020) apply the distinction between affective regret and cognitive regret to hypothetical sacrificial dilemmas. Consistent with the notion of moral residue, many decisions, especially utilitarian decisions, produced affective regret (e.g., self-blame or feeling like "kicking yourself") but little cognitive regret (believing that a different decision would have been better, wishing you had decided differently). The combination of affective regret without cognitive regret is especially indicative of the kind of moral residue philosophers would expect of ultimate moral dilemmas, since it implies negative moral feelings about what one has done without believing one acted wrongly.

There are obvious limitations to investigating moral residue using merely hypothetical choices. However, the theoretical construct of "moral injury" from research on combat veterans provides a striking illustration of the psychological impact of real-life moral dilemmas. In this literature, moral injury is theorized as a distinct component of more general trauma, and is theorized to result from

witnessing or engaging in actions that transgress deeply held moral values – including killing, or being unable to help wounded civilians without risking the lives of fellow soldiers. These experiences are associated with deeply negative feelings such as hopelessness, shame, distress, and anger, as well as issues like depression, anxiety and social withdrawal (for a review, see Griffin et al., 2019).

Molendijk (2018) argues that the experience of seemingly irresolvable value conflicts (a marker of genuine moral dilemmas) is a common theme of moral injury. She also highlights the complexity of the resulting guilt. Consider this striking account from a Dutch veteran describing a situation where more desperate refugees approached his compound than it was possible to accommodate:

> People pressed against one another, against walls, all together. Terrified. Terror in their eyes. I'm going to die, these people thought. Help me, help me. Old men, women, passed out. So, I threw them into the wheelbarrow and drove [them to the compound]. You did what you could. . . . At that point, you're doing it all wrong. Everything. . . . You can't choose between one human life and another human life. So yes, you always do the wrong thing. Everybody in the compound, that didn't fit. (pp. 3–4)

An Afghanistan veteran similarly describes the dilemma of trying to give impromptu life-saving medical treatment, but then receiving military orders to leave. He says of the incident:

> So then you have to take off the oxygen mask and take out the IV. For a nurse, that doesn't make sense. I had taken an oath as a soldier, but as a nurse I also had an oath. But those two promises are not compatible over there, you have to choose . . . In the end I chose the [oath I took as a] soldier. (p. 4)

He goes on to write about feelings of guilt – feelings not quelled by his belief that he made the best decision under the circumstances.

These accounts seem to report the experience of ultimate moral dilemmas, as veterans describe the confusion, distress, and doubt that arise from knowing that they did the best thing they could under the circumstances, while nevertheless feeling that they transgressed a crucial moral requirement (Molendijk, 2018). Recent work has expanded research on moral injury to other real-life moral dilemmas, such as those facing health-care workers during the COVID-19 pandemic (Borges et al., 2020). Future research in moral cognition might consider drawing on this concept to further illuminate the phenomenology of moral dilemmas.

5.5 Conclusion

To understand morality in all of its real-world messiness and uncertainty, we need to understand what it is to face and resolve moral dilemmas. We have argued that the experience of genuine moral dilemmas arises from the recognition that any choice you make requires you to transgress serious moral

requirements or values – thus triggering negative affect, feelings of moral conflict, and specific forms of regret or guilt that persist as "moral residue" even if you did the best you could under the circumstances. This phenomenon, we suggest, is an all-too-common part of everyday moral life. Recognizing this, and recognizing the rich variety of values that feature in experiences and resolutions of such conflicts, calls for a broadening of dilemma research beyond cases of sacrificing strangers, beyond cases of violating requirements for the sake of a utilitarian greater good, and beyond conflicts between "emotion-based" and "reason-based" values. Correspondingly, future research would benefit from more investigation into how people compare and weigh values against one another. At the same time, a narrower target than "moral trade-offs" is needed if we are to fully understand moral dilemmas, with the characteristic experiences of strong conflict, psychological difficulty, and moral residue that they involve. Rather than resulting from any moral trade-off, the conflicting values most capable of generating these experiences may be those that are held equally sacred, absolute, or that strike at the core of our identities as virtuous, moral beings.

References

Bago, B., & De Neys, W. (2019). The intuitive greater good: Testing the corrective dual process model of moral cognition. *Journal of Experimental Psychology: General*, *148*(10), 1782–1801.

Barcan Marcus, R. (1980). Moral dilemmas and consistency. *Journal of Philosophy*, *77* (3), 121–136.

Baron, J., & Gürcay, B. (2017). A meta-analysis of response-time tests of the sequential two-systems model of moral judgment. *Memory & Cognition*, *45*, 566–575.

Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, *70*, 1–16.

Bartels, D. M., Bauman, C. W., Cushman, F. A., Pizarro, D. A., & McGraw, A. P. (2015). Moral judgment and decision making. In G. Keren & G. Wu (Eds.), *The Wiley Blackwell handbook of judgment and decision making* (Vol. 1; pp. 478–515). Wiley Blackwell.

Berman, J. Z., & Kupor, D. (2020). Moral choice when harming is unavoidable. *Psychological Science*, *31*(10), 1294–1301.

Bialek, M., & De Neys, W. (2016). Conflict detection during moral decision-making: Evidence for deontic reasoners' utilitarian sensitivity. *Journal of Cognitive Psychology*, *2*(5), 631–639.

Borges, L. M., Barnes, S. M., Farnsworth, J. K., Frescher, K. D., & Walser, R. D. (2020). A contextual behavioral approach for responding to moral dilemmas in the age of COVID-19. *Journal of Contextual Behavioral Science*, *17*, 95–101.

Bostyn, D. H., Sevenhant, H., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, *29*(7), 1084–1093.

Capraro, V., Everett, J. A. C., & Earp, B. D. (2019). Priming intuition disfavors instrumental harm but not impartial beneficence. *Journal of Experimental Social Psychology*, *83*, 142–149.

Chang, R. (Ed.). (1997). *Incommensurability, incomparability, and practical reason*. Harvard University Press.

Conee, E. (1982). Against moral dilemmas. *Philosophical Review*, *91*(1), 87–97.

Critcher, C. R., Helzer, E. G., & Tannenbaum, D. (2020). Character evaluation: Testing another's moral-cognitive machinery. *Journal of Experimental Social Psychology*, *87*, Article 103906.

Crockett, M. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*, 363–366.

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

Cushman, F. A., Gray, K., Gaffey, A., & Mendes, W. (2012). Simulating murder: The aversion to harmful action. *Emotion*, *12*, 2–7.

Cushman, F. A., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social Neuroscience*, *7*(3), 269–279.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science*, *17*(12), 1082–1089.

Everett, J. A., Pizarro, D.A., & Crockett, M. J. (2016). Inference of trustworthiness from intuitive moral judgments. *Journal of Experimental Psychology: General*, *145*, 772–787.

Gamez-Djokic, M., & Molden, D. (2016). Beyond affective influences on deontological moral judgment: The role of motivations for prevention in the moral condemnation of harm. *Personality and Social Psychology Bulletin*, *42*(11), 1522–1537.

Goldstein-Greenwood, J., Conway, P., Summerville, A., & Johnson, B. N. (2020). (How) do you regret killing one to save five? Affective and cognitive regret differ after utilitarian and deontological decisions. *Personality and Social Psychology Bulletin*, *46*(9), 1303–1317.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong & C. B. Miller (Eds.), *Moral psychology: The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–79). MIT Press.

Greene, J. D. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, *124*(4), 695–726.

Griffin, B. J., Purcell, N., Burkman, K., Litz, B. T., Bryan, C. J., Schmitz, M., Villierme, C., Walsh, J., & Maguen, S. (2019). Moral injury: An integrative review. *Journal of Traumatic Stress*, *32*, 350–362.

Gürcay, B., & Baron, J. (2017). Challenges for the sequential two-system model of moral judgement. *Thinking & Reasoning*, *23*(1), 49–80.

Hanselmann, M., & Tanner, C. (2008). Taboos and conflicts in decision making: Sacred values, decision difficulty, and emotions. *Judgment and Decision Making*, *3*(1), 51–63.

Holyoak, K. J., & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, *142*(11), 1179–1203.

Horne, Z., & Powell, D. (2016). How large is the role of emotion in judgments of moral dilemmas? *PLOS ONE*, *11*(7), Article e0154780.

Hursthouse, R. (2001). *On virtue ethics.* Oxford University Press.

Jack, A. I., Robbins, P., Friedman, J., & Meyers, C. (2014). More than a feeling: Counterintuitive effects of compassion on moral judgment. In J. Sytsma (Ed.), *Advances in experimental philosophy of mind* (pp. 125–179). Bloomsbury.

Kahane, G. (2012). On the wrong track: Process and content in moral psychology. *Mind & Language*, *27*(5), 519–545.

Kahane, G. (2014). Intuitive and counterintuitive morality. In J. D'Arms and D. Jacobson (Eds.), *Moral psychology and human agency: Philosophical essays on the science of ethics* (pp. 9–39). Oxford University Press.

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

Kahane, G., Wiech, K., Shackel, N., Farias, M., Savulescu, J., & Tracey, I. (2012). The neural basis of intuitive and counterintuitive moral judgment. *Social Cognitive and Affective Neuroscience*, *7*(4), 393–402.

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, *314*(5800), 829–832.

Körner, A., & Volk, S. (2014). Concrete and abstract ways to deontology: Cognitive capacity moderates construal level effects on moral judgments. *Journal of Experimental Social Psychology*, *55*, 139–145.

Krosch, A., Figner, B., & Weber, E. U. (2012). Choice processes and their post-decisional consequences in morally conflicting decisions. *Judgment and Decision Making*, *7*(3), 224–234.

Kurzban, R., DeScioli, P., & Fein, D. (2012). Hamilton vs. Kant: Pitting adaptations for altruism against adaptations for moral judgment. *Evolution and Human Behavior*, *33*, 323–333.

Landy, J. F., & Royzman, E. B. (2018). The moral myopia model: Why and how reasoning matters in moral judgment. In G. Pennycook (Ed.), *The new reflectionism in cognitive psychology* (pp. 70–92). Routledge.

Lee, J., & Holyoak, K. J. (2020). "But he's my brother": The impact of family obligation on moral judgments and decisions. *Memory & Cognition*, *48*, 158–170.

Lee, M., Sul, S., & Kim, H. (2018). Social observation increases deontological judgments in moral dilemmas. *Evolution and Human Behavior*, *39*(6), 611–621.

Mandel, D. R., & Vartanian, O. (2008). Taboo or tragic: Effect of tradeoff type on moral choice, conflict, and confidence. *Mind & Society*, *7*, 215–226.

Mata, A. (2019). Social metacognition in moral judgment: Decisional conflict promotes perspective taking. *Journal of Personality and Social Psychology: Attitudes and Social Cognition*, *117*(6), 1061–1082.

McPhetres, J., Conway, P., Hughes, J. S., & Zuckerman, M. (2018). Reflecting on God's will: Reflective processing contributes to religious peoples' deontological dilemma responses, *Journal of Experimental Social Psychology*, *79*, 301–314.

Miller, R., & Cushman, F. (2013). Aversive for me, wrong for you: First-person behavioral aversions underlie the moral condemnation of harm. *Social and Personality Psychology Compass*, *7*(10), 707–718.

Molendijk, T. (2018). Toward an interdisciplinary conceptualization of moral injury: From unequivocal guilt and anger to moral conflict and disorientation. *New Ideas in Psychology*, *51*, 1–8.

Moore, A. B., Clark, B. A., & Kane, M. J. (2008). Who shalt not kill? Individual differences in working memory capacity, executive control, and moral judgment. *Psychological Science*, *19*(6), 549–557.

Nichols, S. (2021). *Rational rules: Towards a theory of moral learning*. Oxford University Press.

Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*(3), 530–542.

Nussbaum, M. C. (2000). The costs of tragedy: Some moral limits of cost-benefit analysis. *Journal of Legal Studies*, *29*, 1005–1036.

Parker, S., & Finkbeiner, M. (2020). Examining the unfolding of moral decisions across time using the reach-to-touch paradigm. *Thinking & Reasoning*, *26*(2), 218–253.

Patil, I., Zucchelli, M. M., Kool, W., Campbell, S., Fornasier, F., Calò, M., Silani, G., Cikara, M., & Cushman, F. (2021). Reasoning supports utilitarian resolutions to moral dilemmas across diverse measures. *Journal of Personality and Social Psychology*, *120*(2), 443–460.

Paxton, J. M., Bruni, T., & Greene, J. D. (2014). Are "counter-intuitive" deontological judgments really counter-intuitive? An empirical reply to Kahane et al. (2012). *Social Cognitive and Affective Neuroscience*, *9*, 1368–1371.

Paxton, J. M., & Greene, J. D. (2010). Moral reasoning: Hints and allegations. *Topics in Cognitive Science*, *2*(3), 511–527.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science*, *36*(1), 163–177.

Piazza, J., & Landy, J. (2013). "Lean not on your own understanding": Belief that morality is founded on divine authority and non-utilitarian moral thinking. *Judgment and Decision Making*, *8*(6), 639–661.

Rand, D. G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430.

Reynolds, C. J., & Conway, P. (2018). Not just bad actions: Affective concern for bad outcomes contributes to moral condemnation of harm in moral dilemmas. *Emotion*, *18*(7), 1009–1023.

Reynolds, C. J., Knighten, K. R., & Conway, P. (2019). Mirror, mirror, on the wall, who is deontological? Completing moral dilemmas in front of mirrors increases deontological but not utilitarian response tendencies. *Cognition*, *192*, Article 103993.

Rom, S. C., & Conway, P. (2018). The strategic moral self: Self-presentation shapes moral dilemma judgments. *Journal of Experimental Social Psychology*, *74*, 24–37.

Ross, W. D. (1930). *The right and the good*. Oxford University Press.

Royzman, E. B., Goodwin, G. P., & Leeman, R. F. (2011). When sentimental rules collide: "Norms with feelings" in the dilemmatic context. *Cognition*, *121*, 101–114.

Royzman, E. B., Landy, J. F., & Leeman, R. F. (2015). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science*, *39*(2), 325–352.

Sauer, H. (2012). Morally irrelevant factors: What's left of the dual process-model of moral cognition? *Philosophical Psychology*, *25*(6), 783–811.

Shortland, N., & Alison, L. (2020). Colliding sacred values: A psychological theory of least-worst option selection. *Thinking & Reasoning*, *26*(1), 118–139.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159–171.

Tetlock, P. E., Kristel, O. V., Elson, B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology*, *78*(5), 853–870.

Trémolière, B., & Bonnefon, J.-F. (2014). Efficient kill–save ratios ease up the cognitive demands on counterintuitive moral utilitarianism. *Personality and Social Psychology Bulletin*, *40*(7), 923–930.

Uhlmann, E. L., Zhu, L., & Tannenbaum, D. (2013). When it takes a bad person to do the right thing. *Cognition*, *126*(2), 326–334.

Williams, B. (1965). Ethical consistency. *Proceedings of the Aristotelian Society* (Supplement), *39*, 103–124.