

Citation:

Bartels, Daniel M., Christopher W. Bauman, Fiery A. Cushman, David A. Pizarro, and A. Peter McGraw (2014), “Moral Judgment and Decision Making,” In G. Keren & G. Wu (Eds.) *Blackwell Reader of Judgment and Decision Making*. Malden, MA: Blackwell.

## MORAL JUDGMENT AND DECISION MAKING

Daniel M. Bartels\*, Christopher W. Bauman<sup>†</sup>, Fiery A. Cushman<sup>‡</sup>, David A. Pizarro<sup>§</sup>, and A. Peter McGraw<sup>||</sup>

\**University of Chicago*, <sup>†</sup>*University of California-Irvine*, <sup>‡</sup>*Brown University*, <sup>§</sup>*Cornell University*, <sup>||</sup>*University of Colorado Boulder*

### 1. INTRODUCTION

Moral rules are rigid. The 10 Commandments of the Bible’s Old Testament, for example, include unambiguous prohibitions, such as, “Thou shalt not kill.” Similarly, Kant’s categorical imperative is absolute: “Act *only* in accordance with that maxim through which you can at the same time will that it become a universal law” (Kant, 1785/2002; emphasis added). In practice, however, people often struggle to determine what is right or wrong. Consider a doctor treating a terminally ill patient who is suffering from unrelenting pain. She may struggle to decide whether the right course of action is to honor the Hippocratic Oath (not to mention laws that explicitly forbid euthanasia in most states) or honor the patient’s request to provide drugs he can use to end his life, especially if the doctor believes that she would make the same request if she were in the patient’s position. She therefore faces a dilemma because multiple moral principles produce conflicting mandates. Decisions that involve tension between moral principles can generate cognitive conflict within a person and ignite disagreement between people. Ultimately, small variations in context across situations can tip the balance between competing moral forces and lead to principle-inconsistent decisions.

Our focus in this essay is *moral flexibility*, a term that we use to capture to the thesis that people are strongly motivated to adhere to and affirm their moral beliefs in their judgments and choices—they really want to get it right, they really want to do the right thing—but context strongly influences which moral beliefs are brought to bear in a given situation (cf. Bartels, 2008). In what follows, we review contemporary research on moral judgment and decision making and suggest

ways that the major themes in the literature relate to the notion of moral flexibility. First, we take a step back and explain what makes moral judgment and decision making unique. We then review three major research themes and their explananda: (i) morally prohibited value tradeoffs in decision making, (ii) rules, reason, and emotion in tradeoffs, and (iii) judgments of moral blame and punishment. We conclude by commenting on methodological desiderata and presenting understudied areas of inquiry.

We acknowledge that this chapter provides an incomplete view of the literature on moral psychology. We focus on moral judgment and decision making in situations that involve tension between moral principles. This focus reflects the intense theoretical curiosity these situations have garnered from behavioral scientists. We do not review one common type of moral choices people face—those that involve tension between moral principles and (material) self-interest—because they are (arguably) less perplexing for models of decision making than situations where moral principles are in conflict. Given our focus on moral judgment and choice, we also do not review research on (im)moral behavior (e.g., stealing, cheating, charitable giving, helping; for reviews see Tenbrunsel & Smith-Crowe, 2008; Treviño, Weaver, & Reynolds, 2006), research on the correspondence between moral behaviors (as in moral licensing; see Merritt, Effron, & Monin, 2010; Sachdeva, Iliev, & Medin, 2009), or research on the correspondence between moral principles and behavior (as in moral hypocrisy; see Monin & Merritt, 2012; Tenbrunsel, Diekmann, Wade-Benzoni, & Bazerman, 2010). Each of these omitted areas of research, as well as others we have not mentioned, illuminate features of morality that are worthy of further review (for broader treatments of morality see edited volumes by Bartels, Bauman, Skitka, and Medin, 2009; Sinnott-Armstrong, 2008, Vol 1, 2, 3).

### **1.1. Identifying the moral domain**

The urge to define is a commendable scientific impulse. Unfortunately, and despite countless attempts, no universally-accepted definition of the moral domain has been offered so far. Rather than toss more fuel onto this bonfire of dispute, we instead gesture at some prototypical features of the moral domain shared by many approaches. Psychological questions about morality are especially likely to address “judgments of justice, rights, and welfare pertaining to how people ought to treat each other” (Turiel, 1983, p. 3). Moral judgments often concern courses of action that

entail some harm, especially loss of life or other physical harm, loss of rightful property, loss of privacy, or other threats to autonomy. Moral judgments also tend to be triggered by actions that affect not only the actor but others as well. People can distinguish doing something that is unwise from doing something that is morally abhorrent, and the assessment that something is morally relevant has particular features that we discuss in the next section.

## **1.2. What makes moral judgment and decision making unique?**

Morality has long been treated as a distinct area of scholarship. Should it be? Put differently, why is it necessary to have a separate chapter on moral judgment and decision making in this handbook? Perhaps morality is distinctive only in its content, but not in any deeper structural way. By analogy, one could have a theory of how people buy houses and a theory of how people buy cars, but presumably these theories would differ little in few respects beyond those directly entailed by the content of the good being purchased. However, a theory of how people choose a spouse might be fundamentally different at a structural level from choices about laundry detergent (Goldstein & Weber, 1995; Rettinger & Hastie, 2001). We believe that moral judgments and choices have some distinctive properties.

Several past influential theories of morality simply asserted that the moral domain was unique based on normative criteria; that is, morality was set apart from other areas of inquiry based on philosophical definitions of what is or is not moral content (e.g., Kohlberg, 1976, 1981; Piaget, 1932/1997). But, in a substantial departure from these past theories, a rich body of evidence based on multiple methodologies and differing theoretical orientations now supports the claim that moral beliefs have distinct psychological properties from non-moral beliefs (e.g., Bauman & Skitka, 2009a; Skitka, Bauman, & Sargis, 2005; Smith, 1984; Turiel, 1983). Although the content of moral beliefs varies across time and place, the belief that there is such a thing as “right” and “wrong” seems to be present across widely different cultures (Shweder, Mahapatra, & Miller, 1987). The descriptive study of moral judgment is essentially the study of what (lay)people think is normative. (We elaborate more on normative approaches in the study of moral judgment and decision making in section 3.1.)

The domain theory of social judgment, for example, maintains that people act and reason differently as a function of whether a given situation is governed by moral rules, social conventions, or subject to personal discretion (e.g., Nucci, 2001; Turiel, 1983). The personal domain refers to aspects of behavior that are not subject to social rules and thus permits variability in belief and behavior. Whether you wear loafers or laces, for instance, is a matter left entirely to personal choice. Social conventions are norms and standards that are widely endorsed by and applied to a particular social or cultural group. Cultures differ considerably in the conventions they have for proper greetings (e.g., kissing, hand-shaking, bowing), and people within each culture are expected to conform to these norms. People from one culture, however, do not typically condemn those from other cultures who greet people differently; variability in conventions across social groups is largely accepted. Morals, in comparison, are the most compulsory rules. People are generally intolerant of morally-deviant behavior, irrespective of others' preferences. For instance, today many people consider slavery to be morally wrong, and so they condemn those who hold slaves even within a culture where slavery is the norm. When people view issues in terms of moral right and wrong, there is little room for compromise: Wrong is wrong (Skitka et al., 2005; Skitka, 2010; Turiel, 2002). Behaviorally, people exhibit moral intolerance by disparaging and distancing themselves from moral reprobates (Haidt, Rosenberg, & Hom, 2003; Skitka et al., 2005; Tetlock et al. 2000).

These distinctive features of moral conviction can be distilled into a few core dimensions (see Bauman & Skitka, 2009a; Skitka, 2010). One such characteristic is that people are objectivist about their moral beliefs. For instance, Goodwin and Darley (2008) asked participants to make judgments about statements from four categories—those that reference empirical beliefs (e.g., the earth is not at the center of the known universe), social conventions (e.g., talking loudly and constantly to the person next to you during a lecture is a permissible action), tastes (e.g., Madison, WI is America's best college town), and moral statements (e.g., anonymously donating a significant proportion of one's income to charity is a morally good action). Participants judged that many moral statements, like empirical beliefs, were likely to be either true or false and that if two people disagreed about a statement, then one of them must be mistaken. Moral statements were rated as second most objective after empirical beliefs, and significantly more objective than social conventions and tastes.

A second characteristic of moral conviction is that people feel that others should universally agree with their moral beliefs or would be persuaded to agree with them if only they knew “the facts” (Skitka et al., 2005). Consistent with this idea, one recent study found that self-reported moral conviction about a specific attitude object (e.g., abortion, HPV vaccines, same sex marriage) is associated with a willingness to assert that one’s own attitude position should be universally adopted; additionally, being asked to write about a moral attitude causes people to more strongly endorse moral universalism (Morgan & Skitka, 2013).

Third, judging that something is morally right or wrong carries an inherent motivational component (Skitka et al., 2005; see also Hume, 1888). Consistent with this idea, people are more inclined to act on their moral attitudes than they are to act on their strong but non-moral attitudes (Skitka et al., 2005; Skitka & Bauman, 2008).

Finally, moral motivation stems from internalized beliefs rather than concerns about how authorities or society feel or what they mandate (Bauman & Skitka, 2009b; Skitka & Bauman, 2009; see also Blasi, 1984). For example, children endorse obedience to moral requests (e.g., no hitting) made by any person, including other children, but they only endorse obedience to norms (e.g., seat assignments) from legitimate authorities (Laupa, 1994). Similarly, children say that hitting and stealing are wrong, even if a teacher says it is okay (Nucci & Turiel, 1978; Smetana, 1981, 1985). Even the mandates of expert authorities with a high degree of legitimacy (e.g., the US Supreme Court) do little to sway people’s moral attitudes (Skitka, Bauman & Lytle, 2009). Once established, research largely suggests that people’s judgments of right and wrong are authority independent (although see Kelly et al., 2007; Sousa, Holbrook, & Piazza, 2009; Piazza, Sousa, & Holbrook, 2013 for some indeterminacies about this claim).

### **1.3. What makes moral judgment and decision making flexible?**

Paradoxically, although people view morality as rigid, objective, universalizable, and so forth, moral judgment and decision making processes can be highly variable across contexts. One reason people may have flexible rather than rigid moral systems is that the moral world is extremely complicated. Most people are unlike philosophers in their ability and desire to achieve logical consistency across their beliefs (Converse, 1964; see also Chugh, Banaji, & Bazerman, 2005; Tenbrunsel & Messick,

1999). Therefore, it seems unreasonable to expect that a simple system would describe the way that most people experience morality.

Another reason to highlight the potential for moral flexibility is that people's moral concerns are complicated. Moral rules and principles represent abstract ideas that must be operationalized and applied to specific situations (cf. Kristiansen & Hotte, 1996; Rokeach, 1973), and the affordances of those situations can highlight one facet or another of a moral value. For example, parental love facilitates the apparent strict adherence to a moral rule of "do no harm." In some instances, however, some parents may believe it is necessary to show "tough love" in the short-term to serve the child's best interest in the long-term. In short, even a simple idea, like do no harm, can mean different things depending on context.

As we mentioned in the introduction, *moral flexibility* refers to the idea that people are often motivated to do the right thing, but a given moral principle can give rise to very different moral judgment and decisions across contexts. Therefore, from the perspective of moral flexibility, inconsistency across moral situations should not necessarily be interpreted as evidence of moral bias, error, hypocrisy, weakness, or failure. Instead, we believe observations such as these can reveal instances when decision making models are underspecified.

To fully understand moral judgment and decision making, we will likely need interactionist frameworks that take into consideration both aspects of the moral decision maker and the context in which her decisions take place. Most human behavior is the joint product of the person and the situation (e.g., Allport, 1937; Lewin, 1935, 1951; Mischel, 1968). A comprehensive, unified model does not exist for moral judgment and decision making, but research has identified important pieces of the puzzle. In what follows, we cover some of the substantial scientific progress in moral judgment and decision making from the past couple of decades, discuss contemporary theoretical cross-currents, and end by noting some areas where we think there is more to be learned.

## **2. ATTEMPTS TO UNDERSTAND MORAL JUDGMENT AND DECISION MAKING: MAJOR RESEARCH THEMES AND THEIR EXPLANANDA**

### **2.1. Moral value tradeoffs**

A major area of research in moral judgment and decision making is about moral tradeoffs, or how people choose between two courses of action that may both lead to morally undesirable effects. These cases are commonly explored using moral dilemmas that ask whether it is permissible to cause direct harm to one person to save the lives of several others. Many of the dominant psychological theories of moral judgment draw inspiration from the philosophical debate over how these moral dilemmas ought to be resolved, and so we begin with a brief survey of the philosophical landscape.

### *2.1.1. Normative ethics*

Several papers in the literature make reference to normative ethical theories, which provide recommendations about how people ought to judge whether acts are morally forbidden, permissible, or compulsory. Two of the most important normative ethical theories are consequentialism and deontology. Consequentialism is the view that the moral status of an action (or inaction) should be determined solely based on the outcomes it produces; actor's intentions and other features of actions or the circumstances in which they are undertaken are irrelevant. Utilitarianism is a well-known and influential version of consequentialism that results from "...combining consequentialism with *welfarism*. Since consequentialism holds that an act is right if and only if it leads to the best consequences, and welfarism holds that the goodness of an outcome is ultimately a matter of the amount of individual well-being, counting everyone equally, it follows that utilitarianism is the view that an act is right if and only if it leads to the greatest total amount of well-being" (Kagan, 1998, p. 52), where "well-being", or "welfare interests," are "abstracted from actual and possible preferences. Welfare interests consist in just that set of generalized resources that will be necessary for people to have before pursuing any of the more particular preferences that they might happen to have. Health, money, shelter, sustenance, and such like are all demonstrably welfare interests of this sort, useful resources whatever people's particular projects and plans." (Goodin, 1993, p. 242).

Deontology is the view that the moral status of an action should be evaluated based on qualities of the action, independent of its consequences. Actions are *intrinsically* wrong if they violate moral rules, such as those that specify rights, duties, and obligations. Deontology acknowledges that the consequences of an act are *relevant* for determining its moral status, but

considerations of moral rules outweigh considerations of the goodness of consequences (Kagan, 1998).

In many contexts, consequentialism and deontology yield the same judgments for harmful acts; doing harm leads to worse consequences, other things being equal. But strict consequentialism treats prohibitions of harmful acts as akin to rules of thumb that must be broken in cases where doing so would produce better consequences. And, conversely, an important characteristic of deontological judgments is that they are consequence-insensitive.

Importantly, consequentialism and deontology do not exhaust the full range of moral considerations identified by theories of normative ethics. Virtue ethics is a view that focuses on moral character and dispositions or traits that promote human flourishing. Virtues do not involve specific actions (e.g., telling the truth) as much as they represent a person's longstanding practices that are consistent with an ideal (e.g., being an honest person). Although this perspective has received much less attention from moral psychologists, a growing body of research suggests that people make moral judgments that are consistent with virtue ethics (e.g., Critcher, Inbar, & Pizarro, 2012; Goodwin, Piazza, & Rozin, in press; Tannenbaum, Uhlmann & Diermeier, 2011; see Pizarro & Tannenbaum, 2010 for a review). We will return to this point when discussing understudied aspects of moral judgment.

### *2.1.2. Protected values.*

Choice models in behavioral decision theory typically assume that people seek to obtain desirable outcomes. This supposition is central to both normative models (that describe how decision makers should choose; Savage, 1954; von Neumann & Morgenstern, 1947) and descriptive models (that describe how people actually choose; Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). However, research also shows that rules play a critical role in decision making (e.g. Amir & Ariely, 2007; March & Heath, 1994). Moral choices represent a useful context to investigate the conflict between rule-based and consequentialist decision strategies because moral choices are sometimes driven more by ideas about how sacred entities are to be treated (“companies should not allowed to buy the right to pollute the earth”) than by the direct consequences associated with the action (“even if pollution credits reduce pollution”). In other words, protected values are

associated with deontological rules (e.g., “do no harm”; Baron, 1996) and not the overall consequences of those actions (Baron & Ritov, 2009). A standard interpretation of such preferences is that protected values motivate rigid rule-based decision processes that ignore outcomes.

The protected values framework describes morally-motivated choice as constrained by an *absolutely restrictive* set of tradeoff rules (Baron & Spranca, 1997). Protected values are exempt from tradeoffs with other values; in theory, they cannot be traded off or violated for any reason, no matter the consequences. As such, they are typically measured by presenting respondents with statements concerning the acceptability of tradeoffs for some resource and asking them to evaluate its moral status. For example, Ritov and Baron (1999) asked participants to respond to a potential tradeoff, and classified people who respond “c” to the item below as those with a protected value for fish species.

*Causing the extinction of fish species:*

- a) *I do not object to this.*
- b) *This is acceptable if it leads to some sort of benefits (money or something else) that are great enough.*
- c) *This is not acceptable no matter how great the benefits.*

People who express a protected value for a given issue are more likely to exhibit “quantity insensitivity.” That is, decision makers with protected values relevant to a particular decision may disregard outcomes *entirely* and view a small violation of a protected value as being equally wrong as larger violations (e.g., Baron & Spranca, 1997). For example, Ritov and Baron (1999) presented participants with a scenario where the only way to save 20 species of fish upstream from a dam on a river would be to open the dam, but opening the dam would kill two species of fish downstream. Participants were then asked (i) whether they would open the dam in this situation, and (ii) to identify the maximum number of fish species they would allow to be killed downstream and still decide to open the dam. Participants who had a protected value about fish extinctions (based on the criterion above) were less likely to choose to open the dam (if doing so would kill two species), and more likely to indicate they were unwilling to open the dam if doing so would cause the loss of even a single species, *even though not opening the dam would lead to the loss of 20 species.*

The link between a non-consequentialism and protected values may not be clear at first blush. Presumably, people care deeply about the entities about which they have a protected value (e.g., family, endangered species). So, one would expect people to be sensitive to the consequences

that befall these protected entities. However, if protected values motivate non-consequentialism, then people who really care about an issue will fail to maximize the goodness of outcomes for these entities *and* they might even appear comparatively ignorant, insofar as they might not be taking stock of the consequences at all. These restrictive tradeoff rules, then, present a major problem that could undercut welfare-maximization, from the standpoint of a utilitarian policy-making perspective (Baron & Spranca, 1997).

Viewing the results of research on protected values through the lens of moral flexibility suggests that a non-consequentialist interpretation considers only part of a larger story. Moral decision makers sometimes affirm their protected values by judging a particular action to be wrong, even in the face of undesirable consequences. However, people with protected values are also capable of identifying situations where the benefits would justify tradeoffs (Baron & Leshner, 2000), and the relationship between protected values and evaluating moral actions is strongly determined by attentional processes (Bartels, 2008; Sachdeva & Medin, 2008; Iliev et al., 2009). For example, situations that direct people's attention to the consequences of their choices (and away from the actions that bring them about, like asking them whether they would make the tradeoff with varying sets of circumstances) make people with protected values *more* willing to endorse tradeoffs than those without protected values (i.e., the opposite of quantity insensitivity; Bartels & Medin, 2007). In short, it appears that features of the situation play a key role in determining whether and how much people base their choices on rules or consequences. People with protected values sometimes appear to be deontologists who strictly follow rules, and they sometimes appear to be utilitarians who ardently pursue the best consequences.

### *2.1.3. Sacred values and taboo tradeoffs*

As noted above, people sometimes reason and choose non-consequentially, such as when they are contemplating the extinction of an endangered species. Although the "protected values" framework (e.g. Baron & Spranca, 1997) addresses these situations, so does a seemingly parallel but distinct literature on "sacred values". Whereas the literature on protected values has largely focused on the problems that absolutely restrictive tradeoff rules pose for seeking utilitarian ends (e.g., crafting optimal policy) and in the cognitive and affective correlates of having protected values (Baron & Spranca, 1997), the literature on sacred values presents a framework, largely derived from

sociology, for understanding where such rules might come from (Fiske & Tetlock, 1997) and for understanding how people manage to flexibly navigate through a world that forces them to make value tradeoffs (Tetlock, 2002).

The sacred value framework primarily examines exchanges in which decisions are determined by the moral significance attached to the things being exchanged. Certain cherished goods, like human life, health, and nature, are treated by people in some communities as having intrinsic moral or transcendental value. In all but the most extreme circumstances, these *sacred values* are not to be exchanged for *secular values*, especially goods that can be bought or sold. For example, selling one's vote or paying someone else to take one's place in a military draft seems morally abhorrent to many people. The sacred values framework explains these instances of non-consequentialist judgments as the result of a person having internalized a set of culturally-defined norms that constrain the manner in which different types of goods can be permissibly exchanged for each other. Most research on sacred values focuses on the restrictive tradeoff rules that suggest that strongly held, situation-specific values engender deontological decision strategies. For example, people often have strong reservations about market exchanges for sex, organs, and adoption (Sondak & Tyler, 2001; Tetlock et al. 2000). People have similar reservations about religious organizations (i.e., a sacred entity) relying on commercial marketing strategies (i.e., a secular solution) to recruit and retain their members (McGraw, Schwartz, & Tetlock, 2012). These effects extend outside the laboratory. A field study involving Palestinians and Israeli settlers facing a (hypothetical) tradeoff involving a sacred value (e.g., returning land, recognizing a Palestinian state) reacted with greater outrage when the peace deal was "sweetened" with a monetary offer (e.g., receiving a \$1 billion a year for 100 years in money from the US; Ginges, Atran, Medin, & Shikaki, 2007; see also Deghani et al., 2009; 2010)

Sacred-secular exchanges are judged to be morally reprehensible (Tetlock et al., 2000). Tetlock's (2002) framework suggests that facing a "taboo tradeoff" — sacred-for-secular tradeoff, those where *only one of two or more resources are treated as morally significant* — decision makers view utilitarian considerations (i.e., costs and benefits of alternative courses of action) as off-limits. When posed with a taboo tradeoff, people instead adhere to deontological constraints, affirming their culture's proscriptive moral rules. People want to avoid taboo tradeoffs for interpersonal and intrapersonal reasons. Avoiding taboo tradeoffs means avoiding the negative judgments made by

others (Tetlock, Peterson, & Lerner, 1996); even knowing that someone *merely contemplated* such a tradeoff is aversive; eliciting contempt, disgust, and the judgment that such contemplation is unforgivable (Tetlock et al., 2000). Also, a decision maker facing a potentially taboo tradeoff experiences negative emotions (McGraw & Tetlock 2005; Tetlock et al. 2000), which they avoid by abandoning tradeoff reasoning (McGraw, Tetlock, & Kristel, 2003). For instance, contemplating secular-sacred tradeoffs, such as whether to save money on an apartment or a vehicle by accepting lower levels of safety, leads consumers to abandon tradeoff based reasoning to avoid negative emotions (Luce, Payne, & Bettman 1999, 2000).

According to the sacred values framework, not all morally significant exchanges are impermissible. In situations where *only* sacred values are at issue—“*tragic tradeoffs*”—people believe it is okay, perhaps even a good idea, to weigh utilitarian considerations. For example, people are not outraged when they learn about a hospital administrator agonizing over a decision about which of two dying children should be given the one life-saving organ, regardless of the ultimate choice (Tetlock et al., 2000).

The permissibility of tradeoff goes even further when one takes into account research that investigates how simple rhetorical messages can reframe a taboo tradeoff into acceptable, even routine tradeoff (McGraw & Tetlock, 2005; McGraw, Schwartz, & Tetlock, 2012). For instance, the public was outraged when they found that the Clinton administration provided major campaign contributors a night’s stay in the Lincoln bedroom. That outrage was mitigated when a reciprocity norm was invoked to explain the situations – “friends doing favors for friends” (McGraw & Tetlock 2005), and this mitigation was especially pronounced for the people—Bill Clinton supporters—who were most motivated to avoid experiencing outrage associated with the idea that rooms in the White House were up for sale as they would be in a common hotel.

What underlies the moral flexibility demonstrated in the literature on sacred values? Evidence suggests that people understand that moral concepts like harm, equality, or purity can have different meanings depending on the type of social relationship a given situation involves (Rai & Fiske, 2011). For example, even a slight change in framing a policy decision as military or diplomatic has a profound effect on how people view potential responses to a hostage situation (Ginges & Atran, 2011). According to this view, people’s flexibility in applying moral values across situations is a consequence of how people interpret social situations and implement a finite set of

schemata about the nature of the relationships therein. Specifically, social relationships can be grouped into four basic models: communal sharing, authority ranking, equality matching, and market pricing (Fiske, 1991, 1992; see also Haslam, 2004). The acceptability of a particular tradeoff depends on the relational system invoked. In communal sharing relationships, group members (e.g., a family)—but not outsiders—have equal status and expect equal access to shared resources. Authority ranking relationships include asymmetry among group members, such that lower-ranking members are expected to show deference to higher-ranking members. Equality matching relationships are characterized by efforts to balance outcomes across people based on comparisons along one dimension at a time (e.g., turn taking). In market pricing relationships, people strive to aggregate several dimensions of comparison (e.g., time spent, effort exerted, and output quality) using a common metric (usually money) that makes complex evaluations and exchanges possible. These four basic relational schema facilitate social interaction by allowing people to form expectations for their own and others' behavior, evaluate exchanges, and identify violations. From this perspective, taboo tradeoffs occur when relational schema conflict, such as when a market-pricing perspective comes into conflict with communal sharing (e.g., someone offers to pay for the part of the Thanksgiving dinner they ate at a friend's house; Fiske & Tetlock, 1997).

A relational regulation model provides a promising approach to understanding moral flexibility, as it represents a framework for understanding when and why moral rules and motives vary across situations and people (e.g., when people care about equality, when they care about equity, and whether they “keep score” at all). So, tradeoffs that are common and uncontroversial when viewed through the lens of one relational model can appear bizarre, unacceptable, or offensive when viewed through the lens of another. For example, people routinely set prices based on supply and demand without concern when they perceive the market pricing schema to apply to the situation. When it comes to fundamental needs—like being able clear one's driveway after a snowstorm—people often apply the communal sharing model, which causes them to judge people who set very high prices (e.g., for snow shovels) based on low supply and high demand as “price gougers” (Kahneman, Knetsch, & Thaler, 1986).

In sum, the sacred values framework thus links moral values (as internalized norms that afford certain goods moral significance) to deontology in some contexts; utilitarian considerations are off-limits when contemplating a sacred-for-secular exchange. The sentiment is that some goods

or services are not exchangeable for money, no matter what favorable consequences might be brought about by the exchange. The framework also identifies situations where consequentialist cognition is permissible (as in tragic tradeoffs, like the organ transplant example, and in secular-secular tradeoffs, like purchasing a computer). It further identifies that even in sacred-secular cases, it is possible to attenuate the outrage and contempt associated with taboo tradeoffs through rhetorical manipulations that change the social-relational context of the tradeoff.

## **2.2. Rules, reason, and emotion in moral tradeoffs**

A large amount of research into moral tradeoffs has investigated reactions to the trolley problem as a means to examine the contributions of emotion, reason, automaticity, and cognitive control in moral judgment (Foot, 1967; Thomson, 1985; see also Waldmann, Nagel, & Wiegmann, 2012). In the “bystander” version, a runaway trolley is on a path that will kill five workers on the track ahead, and study participants must decide whether to flip a switch that would divert the trolley onto a side track where it would kill only one person. In the “footbridge” case, five people are similarly threatened, but study participants must decide whether to throw a person in front of the trolley (killing him) to save the five on the tracks. A large majority of people say that the one person should be sacrificed for the sake of five in the bystander case, whereas a small minority say that the one person should be sacrificed in the footbridge case (Greene et al., 2001; Hauser, Cushman, Young, Jin, & Mikhail, 2007; Mikhail, 2009). If people were strictly following a broad deontological rule, like “It is absolutely forbidden to intentionally kill someone,” they would respond “no” in both cases. If people were strictly following utilitarianism (i.e., bring about the greatest good for the greatest number), they would respond “yes” in both cases. Therefore, the results show that most people are not rigidly deontological or utilitarian, and researchers have sought to explain what accounts for the flexibility people exhibit when dealing with these cases.

Several explanations have been offered for what underlies discrepant responses across versions of the trolley problem. For example, some have argued that the issue is whether an action directly (as in footbridge) or indirectly (as in bystander) causes harm (Royzman & Baron, 2002), whether the causal focus is directed on to the trolley or the people on the track (Waldmann & Dieterich, 2007; Iliev, Sachdeva, & Medin, 2012), whether the action is interpreted as violating a rule in the social contract (Fiddick, Spampinato & Grafman, 2005), and whether the outcomes are

viewed as gains or losses (Petrinovich, O’Neill, & Jorgensen, 1993). Below, we will provide expanded descriptions of three interpretations that have received considerable attention. One posits opposing roles for affect-laden intuition vs. reflective thought (Greene, 2007). Another decomposes these scenarios into their causal structure and invokes concepts like “assault”, “battery”, and “homicide” to account for judgments (Mikhail, 2007). A third invokes affectively-tagged moral rules and consideration of consequences (Nichols & Mallon, 2006).

### 2.2.1. *Dual process morality*

One influential model for understanding people’s responses to sacrificial dilemmas like the trolley problem is a dual system theory that contrasts reflective and emotional, intuitive processing (Greene, 2007; Greene et al., 2001). According to the model, controlled cognitive processes are responsible for welfare-maximizing (i.e., utilitarian) choices, such as flipping the switch and pushing the man in the bystander and footbridge versions of the trolley problem. The automatic emotional processes are responsible for choices that correspond to deontological rules, such as the aversion to causing direct harm to the man on the footbridge. So, this view maintains that although deontological *philosophy* depends on explicit, conscious rules, many deontological-seeming, non-utilitarian *judgments* depend on automatic, emotional intuitions. Central to the theory is a distinction between *personal* moral dilemmas that involve a strong affective component (such as the footbridge version of the trolley problem, where it is necessary to actually push a person to their death) versus *impersonal* dilemmas that do not involve such an affective component (such as the bystander version, where you are merely required to flip a switch that redirects the train). Personal dilemmas are proposed to evoke a conflict between utilitarian and deontological considerations, while impersonal dilemmas do not.

Several sources of evidence support this dual process hypothesis. Some studies underscore the role of controlled cognition: fMRI reveals correlates of controlled cognition for utilitarian choices (Cushman, Murray, Gordon-McKeon, Wharton, & Greene, 2011; Greene et al., 2004), time pressure and cognitive load decrease the frequency and speed of utilitarian choices (Suter & Hertwig, 2011; Trémolière, De Neys, & Bonnefon, 2012). Others underscore the role of affect: brain damage to regions that process emotions increases utilitarian responding, (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs et al., 2007; Moretto, Ladavas, Mattioli, & di

Pellegrino, 2010) and people who exhibit low levels of affective concern for others make more utilitarian judgments (Bartels & Pizarro, 2011; Gleichgerrcht & Young, 2013). Moral judgments are remarkably malleable: for instance, pharmacological interventions that enhance aversive learning and inhibition promote deontological responses (Crockett, Clark, Hauser & Robbins, 2010), as do manipulations that encourage participants to imagine the harmful consequences of action in vivid detail (Amit & Greene, 2012; Bartels, 2008). Additionally, people with higher working memory capacity and those who are more deliberative thinkers are more likely to judge a harmful utilitarian action as permissible (Bartels, 2008; Feltz & Cokely, 2008; Moore, Clark, & Kane, 2008; although, notably, people prone to reflective thought also tend to judge that it is permissible not to do so, a pattern of normative indifference referred to as “moral minimalism”; Royzman, Landy, & Leeman, in press).

At the same time, evidence for dual process model of morality has been challenged on empirical and methodological grounds (McGuire, Langdon, Coltheart & Mackenzie 2009; Baron, Gürçay, Moore & Starcke 2012; Kahane, Wiech, Shackel, Farias, Savulescu & Tracey 2011) and also on conceptual grounds (Kvaran & Sanfey, 2010; Moll, Oliveira-Souza & Zahn 2008; Nucci & Gingo 2010). A common theme among these critiques is that, ultimately, a sharp division between “cognitive” and “emotional” systems cannot be maintained. Utilitarian judgments require some kind of motivational grounding, while characteristically deontological judgments require some kind of information processing. In fact, this point is echoed even by proponents of a dual-process approach (e.g., Cushman, Young & Greene 2010).

Recently, an alternative dual process model has been proposed that draws on current neurobiological models of reinforcement learning (Crockett, 2013; Cushman, 2013). These models rely on a broad division between two algorithms for learning and choice. One algorithm assigns values to actions intrinsically based on past experience (e.g., “I just don’t feel right pushing a person”, due to having been scolded for pushing on the playground years ago) and provides an explanation for an intrinsic aversion to harmful actions performed in ways that are more “typical”, such as up-close and personal harms. The other algorithm derives the value of actions from an internally represented causal model of their expected outcomes (e.g., “if I flip this switch, it will send a train down the side track and will kill the person standing there”) and provides an explanation for utilitarian moral judgments. These models of reinforcement learning were

developed quite independently of the literature on moral judgment, but they revolve around a distinction between *good versus bad actions* and *good versus bad outcomes* that resonates deeply with dual process models of morality (Miller, Hannikainen & Cushman, 2013).

### 2.2.2. *Moral grammar*

The dual process model of moral judgment relies largely on a coarse distinction between more “personal” and “impersonal” harms, but a wealth of evidence indicates much greater subtlety and organization in moral intuitions. For instance, people condemn actively causing harm more than passively allowing harm (Baron & Ritov, 2009; Cushman et al., 2006; Spranca, Minsk, & Baron, 1991), and they are also more willing to make tradeoffs (e.g., sacrificing one life to save another) in cases that involve passive rather than active harm (Goodwin & Landy, in press). People more strongly condemn actions that involve the direct transfer of bodily force from the actor to the victim (e.g., pushing the man off the footbridge) than when no such transfer of “personal force” occurs (e.g., flipping the switch in the bystander version; Cushman et al., 2006; Greene et al., 2009). And, the bystander and footbridge versions also differ in whether the choice involves an action that causes harm to someone as a means to save others or as a side-effect of saving others (Cushman et al., 2006; Foot, 1967; Mikhail, 2000; Thomson, 1985). In the footbridge version, the potential victim would be used as a “trolley-stopper,” an instrument to accomplish the goal of saving five others. In the bystander case, however, the potential victim would be collateral damage; his death would be an unfortunate consequence of diverting the train, but saving the five people on the other track would not be a consequence of his death.

The theory of universal moral grammar provides a descriptive account of these detailed principles (Mikhail, 2009). It maintains that moral judgment is the product of a single, relatively discrete psychological system (i.e., dedicated to morality) that distills situations into their causal and intentional structure and makes use of rules and legal concepts like battery, assault, and homicide to convert important features of situations morally-valenced judgments (for another interesting exploration of judgment processes involving legal concepts, see Koehler & Meixner, this volume). This system is postulated to be innate and to operate below the level of conscious awareness (cf. Chomsky, 1957, 1965), and whereas some of its rules specify the relative moral value of different outcomes (e.g., human life is good, causing harm is bad), other specify moral constraints on actions

that bring about those outcomes (e.g., intentionally killing someone is bad, allowing people to die is bad if an actor could save them without incurring unreasonable costs). Compared with the dual process model of moral judgment, universal moral grammar has the virtue of explaining many detailed patterns in judgment that have been repeatedly identified in the literature, and it also provides an appraisal theory for what kinds of actions are likely to trigger specifically *moral* judgment. On the other hand, universal moral grammar is less suited to explain dissociations between the apparent contributions of automatic versus controlled processes to moral judgment.

### *2.2.3. Rules and emotions: A potential reconciliation*

An alternative model that may reconcile seeming disparate approaches to understanding affect and cognition in moral judgment contends that moral cognition depends on an “affect-backed normative theory” (Nichols, 2002; Nichols & Mallon, 2006). Under ordinary circumstances, an act is judged wrong only if it both violates a proscriptive moral rule (e.g., “don’t kill”) and also elicits an affective response. For example, people consider it worse to violate a rule of etiquette that evokes disgust (e.g., spitting in one’s soup and then eating it) than to violate a rule of etiquette that does not (e.g., playing with one’s food; Nichols, 2002). However, in a scenario where a little girl has been instructed by her mother not to break a teacup, people consider her decision to break a teacup to prevent five others from being broken a violation of a rule, but not a moral violation, because a broken teacup does not elicit a strong emotional response (Nichols & Mallon, 2006).

While this theory offers a promising framework, there is substantial evidence that the framework must be a flexible one. Nichols and Mallon (2006) found that even affect-backed moral rules could be overwhelmed by good or bad consequences of great magnitude. For example, when told billions of people would die from a virus released into the atmosphere unless a man is killed, 68% of participants judged that such an action violates a moral rule. However, only 24% judged that the action was morally wrong, all things considered. Adding further detail to the dimensions of flexibility, Bartels (2008) found that people’s moral decisions depended on (i) the moral relevance ascribed to choices (i.e., whether or not they endorsed proscriptive rules for actions), (ii) evaluative focus (whether their attention was directed to rule-violating actions vs. the positive consequences these actions produced), and (iii) processing style (whether people were likely or unlikely to incorporate affective reactions to rule violations into their moral judgments of right and wrong)—

people who were more likely to “trust the gut” made more non-utilitarian judgments than people who were more skeptical of their own intuition. Consistent with the framework set out by Nichols and Mallon (2006), the results demonstrated that moral rules play an important, but context-sensitive role in moral cognition (see also Broeders et al., 2011). When violations of moral proscriptions were egregious, they generated affective reactions that overwhelmed consideration of the consequences favoring their violation. When attention was directed to the positive consequences of such actions, however, people reluctantly ignored these proscriptions.

#### *2.2.4 Moral dilemmas and moral flexibility*

Although there is an active debate among competing theories of moral dilemmas, there is also an underlying agreement concerning their significance: 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” (Cushman & Greene, 2012; Sinnott-Armstrong, 2008; Sloman, 1996). In a sense, this both compels and enables moral flexibility. It compels flexibility because certain circumstances require uncomfortable tradeoffs between competing moral values. But it also enables moral flexibility, because it leaves people with an array of potential bases for a favored judgment. It has been suggested that consequentialist reasoning is particularly susceptible to motivated moral reasoning: When people are opposed to an action, they may selectively focus on potential negative consequences of an action and disregard potential positive consequences (Ditto & Liu, 2011).

### **2.3. Judgments of moral blame and punishment**

Research into moral tradeoffs like the trolley problem revolves around tragically difficult choices that are as fantastical as they are gripping. The role of moral flexibility in situations like these is clear because of the inherent moral conflict that dilemmas engender. But a much more common feature of everyday life is the task of assigning responsibility, blame, and punishment to those around us for more minor infractions: A fender-bender, a broken promise, or a rude comment, for instance. In this section we describe the basic processes that translate our perception of an event (“The petty cash drawer was empty.”) to an assignment of responsibility (“Jim stole the money.”) to

a punishment (“Jim should be fired.”). Although one might suppose that such apparently simple judgments would leave little room for moral flexibility, in fact we find even the process of blame attribution is rife with conflict and compromise between competing moral principles.

Building on normative philosophical and legal traditions and classic work on the psychology of attribution (e.g., Heider, 1958; Kelley, 1967), psychologists have outlined the specific conditions necessary for arriving at a judgment that someone is blameworthy (e.g., Shaver, 1985; Weiner, 1995; Alicke, 2000; Malle, Guglielmo & Monroe, 2012). When presented with a possible moral infraction, perceivers ask themselves a series of questions about various features of the act, such as whether actors intended the outcome, had control over the outcome, and could foresee the results of their actions. Although the details differ between models, they all share some core features. Perceivers (i) assess whether there is a causally responsible agent, (ii) evaluate whether that agent intentionally caused the harm, and (iii) assign moral responsibility, blame and punishment. These steps are typically posited to occur in that temporal order. If all these conditions are met, perceivers conclude that the actor should be held responsible and blamed (or praised, for positive actions); however, several studies have identified instances when this attribution sequence “breaks.” By disrupting the ordinary causal relationship between an actor and a victim, researchers hope to further refine our understanding of how people decide whether a person is to blame and how much punishment (if any) they deserve.

### *2.3.1. Moral Luck*

Perhaps the simplest way to break the attribution sequence is with an accidental harm (“I thought I was putting sugar in your coffee, but it was rat poison!”) or an attempted harm (“I thought I was putting rat poison in your coffee, but it was sugar!”). In such cases, moral judgments are largely determined by a person’s intentions (Young, Cushman, Saxe & Hauser 2007; Young, Nichols & Saxe 2010). Along similar lines, people tend to discount moral blame when an agent does not act with control over their behavior (Shaver, 1985). For instance, relatives of people suffering from schizophrenia attenuate blame for actions that were undertaken as a direct result of the person’s (uncontrollable) hallucinations and delusions (Provencher & Fincham, 2000). Also, people are more likely to assign blame to AIDS patients if they contracted the disease through

controllable means (licentious sexual practices) than if through uncontrollable ones (receiving a tainted blood transfusion; Weiner, 1995).

There are some cases, however, in which accidental outcomes can make a surprising difference in our moral judgments. Consider, for example, two drunk drivers who were involved in accidents. One falls asleep, veers off the road, and strikes a tree, but the other falls asleep, veers off the road, and kills a pedestrian. The driver who kills the pedestrian faces a much stiffer punishment than the one who strikes the tree, a phenomenon is known as “moral luck” in philosophy and law (Hall, 1947; Hart & Honore, 1959; McLaughlin, 1925). In addition, many studies show moral luck effects in peoples’ intuitive judgments (Berg-Cross, 1975; Cushman, 2008; Cushman, Dreber, Wang, & Costa, 2009). According to one account of the phenomenon, intent-based moral judgment and outcome-based moral judgment operate in competition (Cushman, 2008). A competitive interaction between these two types of judgments may explain why people may feel caught in a dilemma in cases of moral luck. On the one hand, it seems wrong to treat the two drunk drivers differently given their identical behavior. On the other hand, it seems even more wrong to send one to prison for a DUI, or to let the other off with a ticket for killing a girl. In other words, intent to harm and causal responsibility for harm may not be fused into a single process of blame assignment, but rather exert independent influences on different categories of moral judgment (see also Buckholz et al., 2008).

### 2.3.2. *Causal Deviance*

Another example of when the standard attribution sequence breaks down also comes from the philosophical literature on cases of “causally deviant” actions (Searle, 1983). While traditional theories of responsibility specify that an agent should receive blame if she intended and caused an action (e.g., murder), people tend to reduce blame for an outcome if the intention and the cause are not linked tightly. Take this example, adapted from Chisholm (1966):

*Joe wants to kill his rich uncle, as he stands to inherit a large sum of money. He formulates his plan to murder his uncle, and begins the drive to his uncle’s home. Excited at the prospect of soon acquiring a lot of money, Joe is a bit careless at the wheel and hits and kills a pedestrian. This pedestrian turns out to have been his uncle.*

According to the standard descriptive theories, people should ascribe full blame to Joe for the death of his uncle, as he intended the outcome, and was its sole cause. Yet the “deviant” link between Joe’s intentions and the outcome cause people to find the actions less blameworthy. For instance, in one study researchers provided participants with a description of a woman who successfully murdered her husband by poisoning his favorite dish at a restaurant. Some participants received a modified version of this scenario, in which the husband’s death only came about because the poison made the dish taste bad, and the husband ordered a new dish to which he was deathly allergic. Across a variety of these cases, participants reduced blame (as well as praise for positive actions) in the scenarios that included a “deviant” causal chain (Pizarro, Uhlmann, & Bloom, 2003), demonstrating that even in cases where an act is intended and caused, disrupting the causal chain—even if the intentions remain—reduces the willingness of individuals to assign responsibility

### 2.3.3. *Backwards Inferences*

Ordinarily, attribution-based accounts of moral responsibility, blame, and punishment assume that people begin with a fixed representation of an event—what a person intended, how they acted, and what harm they caused—and then proceed to make a moral evaluation based on this representation. But several revealing cases show influences that work in the opposite direction—whether or not a person acted in a way that is morally wrong can influence perceptions of the actors’ intentions and causal responsibility. Consider, for example, a man speeding home in a rainstorm who gets into an accident and injures others. People are more likely to judge that the man had control over the car if he was speeding home to hide cocaine from his parents than if he was speeding home to hide an anniversary gift for his wife, irrespective of the fact that the factors that led to the accident were identical across both scenarios (Alicke, 1992). According to Alicke, our desire to blame the nefarious “cocaine driver” leads us to distort the criteria of controllability in a fashion that validates this blame.

Researchers also have demonstrated similar asymmetries in judgments of intentionality. People are more inclined to say that a consequence was produced intentionally when they regard its consequences as morally wrong than morally right (see Knobe, 2006, for a review). Suppose that the CEO of a company is told that implementing a new policy will have the side-effect of either harming or helping the environment. In both cases, the CEO explicitly states that he only cares

about increasing profits, not about the incidental side-effect of harming or helping the environment. Nonetheless, study participants assert that the CEO who harmed the environment did so intentionally, whereas the one who helped the environment did not (Knobe, 2003). Although the mechanisms underlying this effect, and the conditions under which it occurs are open questions (see, e.g., Guglielmo & Malle, 2009; Uttich & Lombrozo, 2010; Sloman, Fernbach, & Ewing, 2012), the basic phenomenon is easily replicable and can be observed in children (using simpler scenarios) young as six and seven years old (Leslie, Knobe & Cohen 2006).

#### *2.3.4. Blame and moral flexibility*

Many everyday cases of moral violations fit a standard sequence in which one person intentionally causes harm to another, and this sequence appears to be reflected in our ordinary attribution processes. However, the research reviewed in this section is suggestive of moral flexibility. Moral responsibility, blame, and punishment vary across contexts, which suggests that more nuanced influences are at play in these judgments. In addition to the studies reviewed above, a grab bag of other factors exacerbate or moderate judgments of deserved punishment, including but not limited to the induction of incidental anger (i.e., anger about something unrelated to the focal event; Lerner, Goldberg, & Tetlock, 1998), whether the actor apologizes to the victim, acknowledges guilt, and/or is granted forgiveness by the victim (Robinson, Jackowitz, & Bartels, 2012), and whether the action is described as happening in the past or in the future (Burns, Caruso, & Bartels, 2012), or cultural differences in the ways that perceived intentions, actions, and outcomes relate to judgments of responsibility and morality (Miller & Bersoff, 1992; Cohen & Rozin, 2001). Although no model explains all of the phenomena mentioned above, we can gain some understanding by (i) examining dissociations between the psychological processes that determine the culpability of acts based on intentions versus those that determine culpability based on outcomes and (ii) further probing the psychological processes that cause assessments of causation, control, and intent to be reciprocally influenced by the moral status of the action that a person performs.

## **2.4 Summary of evidence for moral flexibility**

If there is one undisputed fact about the human capacity for moral judgment, it is that the capacity itself comprises a diverse array of distinct psychological processes. These processes often operate in competition, and the moral dilemmas that result provide fertile ground for the work of philosophers and psychologists alike. They also provide people with the challenge of reconciling diverse moral concerns, and the opportunity to selectively recruit moral principles to support a favored judgment. This tremendous complexity gives rise to methodological and theoretical challenges for research in moral psychology. Next, we sketch some of the challenges and suggest a few promising paths forward for the field.

### 3. METHODOLOGICAL DESIDERATA

#### 3.1. Exercise caution when making comparisons to normative standards.

One research strategy that has been used in judgment and decision making research is to (i) identify a normative model, (ii) demonstrate ways that people's responses systematically diverge from the predictions of the normative model, and (iii) treat these "errors" as diagnostic of mental processes (e.g., Kahneman, 2000; Shafir & LeBoeuf, 2002). Sunstein (2005) uses this method to identify what he calls "moral heuristics." While adhering to this error-and-bias approach makes sense across many domains of choice where there is widespread agreement about the normative standard (such as probabilistic judgments), it is unclear that the approach is appropriate for ethical judgment given how little agreement there is among experts or lay judges about the "right" answer. For example, a survey of 73 professors with a PhD in philosophy and primary area of specialization in ethics revealed that 37% endorse deontological principles, 27% endorse utilitarian principles, 22% endorse virtue ethics, and 14% endorse none of the above (personal communication: reanalysis of data reported in Schwitzgebel & Cushman, 2012). In short, comparison to normative standards is more problematic for moral than for non-moral judgments and decisions due to the lack of consensus (or even a majority opinion) about which normative theory provides the right answer about how to act across situations.

Nonetheless, the error-and-bias approach to moral judgment is common in the literature. Several prominent researchers in the field adopt utilitarianism as the normative standard of comparison for ethical judgment and categorize deontological judgments as heuristics that can give

rise to “errors” when they conflict with the greater good. Sunstein (2005), for example, adopts this approach and argues that deontological “rules of thumb” give rise to “mistaken and even absurd moral judgments” (p.531). For Sunstein, the implication is that we should be wary of deontological intuitions, as they are likely to be “unreliable” and “unsound,” and that these intuitions ought to be deprecated when making decisions about law and public policy (for critiques of this approach, see Bauman & Skitka, 2009a; Mikhail, 2005; Pizarro & Uhlmann, 2005; Tetlock, 2005).

In addition to the problem of disagreement over normative standards, the standard manner used to assess moral judgments (i.e., through the use of sparsely-described moral tradeoff scenarios that pit a utilitarian option against a deontological option) might not identify the psychological mechanisms that give rise to the response. In the footbridge version of the trolley problem, for example, participants may choose to push the fat man because (i) they care so much about saving lives that they are reluctantly willing to do what would otherwise be a horrible thing, or (ii) because killing someone is not as aversive to them as it is to others. Pushing the man off the footbridge is the “optimal” response for a researcher adopting utilitarianism as the normative standard, but simply recording participants’ choices offers no way to distinguish between “real” utilitarians from potential psychopaths. In fact, some evidence suggests that these dilemmas are likely capturing the latter. People who score higher in antisocial personality traits, including psychopathy, Machiavellianism, and the perception that life is meaningless, are more likely to push the fat man and provide seemingly “utilitarian” responses in other similar dilemmas (Bartels & Pizarro, 2011).

Although identical responses to these dilemmas may reflect very different sets of intuitions, some have argued that any intuitions generated by these artificial problems are not trustworthy. For example, Hare (1981, p. 139) writes: “Undoubtedly, critics of utilitarianism will go on trying to produce examples which are both fleshed out and reasonably likely to occur, and also support their argument. I am prepared to bet, however, that the nearer they get to realism and specificity, and the further from playing trains—a sport which has had such a fascination for them—the more likely the audience is, on reflection, to accept the utilitarian solution.” These intuitions may also be a result of features that even respondents would agree are morally irrelevant (e.g., Ditto, Pizarro, & Tannenbaum, 2009; Uhlmann, Pizarro, Tannenbaum, & Ditto, 2009), and as such might not even derive from the application of a *moral* principle.

In short, there are good reasons to question some widespread methodological practices that are used in the study of moral judgment: the adoption of a normative standard to assess “errors” of moral judgment, and the reliance on sacrificial moral dilemmas. An overreliance on these methods may prevent us from uncovering the subtlety and complexity of our everyday moral psychology.

### **3.2. Measurement issues**

We noted earlier that both the world and people’s moral systems are complicated. One implication of moral flexibility is that study participants may look to the study stimuli for cues about how to affirm their values when making choices. Seemingly insignificant differences in study design may affect the responses that participants give and the resulting inferences that researchers draw.

#### *3.2.1. Answers depend on questions*

Discussing morality is difficult in the real world, and those difficulties extend to the laboratory. Participants respond to the question asked rather than to the question researchers think they asked. Questionnaire items need to be both accurate and precise; good items are interpreted in the same way by different people (accurate) and good items address the construct they were intended to assess (precise; see also Grice, 1975). Kahane & Shackel (2010) note that researchers use a wide array of items when studying moral dilemmas, often with unforeseen consequences. Different types of judgment—for instance, “wrong,” versus “inappropriate,” “forbidden,” or “blameworthy”—can produce systematic effects on responses (Cushman, 2008; O’Hara, Sinnott-Armstrong, Sinnott-Armstrong, 2010). Other differences that stem from differences in the way that questions are asked can be even more substantial. A large asymmetry can emerge across responses to prompts that differ in valence (e.g., “forbid” vs. “allow;” Holleman, 1999). Additionally, responses to questions about wrongness (e.g., “Is it wrong to X?”) and behavioral intentions (e.g., “Would you X?”) can differ dramatically (Borg, Hynes, Van Horn, Grafton, & Sinnott-Armstrong, 2006). Therefore, researchers must be as diligent when constructing measures as when creating stimuli and handling other methodological aspects of studies.

### 3.2.2. Response format

A wealth of studies demonstrate that the way researchers ask for a moral judgment influences the answers they will get from participants. Sometimes this is a nuisance, but at least as often, it turns out to be psychologically revealing. For instance, Bartels and Medin (2007) found that by refocusing participants' attention to the costs versus benefits of an environmental tradeoff (killing some fish on order to save others), they could vary the relationship between protected values and utilitarian choice. When benefits were highlighted, those who treated fish as protected values were more utilitarian; when costs were highlighted, those who treated fish as protected values were less so.

One of the most commonly used methods of manipulating participants' judgments is to use joint versus separate evaluation of alternatives, which represents a choice between within- versus between-subjects designs. Both designs have strengths and limitations, and the decision to use one over the other should depend on the particular question being asked (for an excellent discussion on this topic, see Charness, Gneezy, & Kuhn, 2012). Within-subjects designs provide greater statistical power and depend less on random assignment to minimize differences across conditions of the experiment (for further discussion of these issues, see Fisher, 1935). Between-subjects designs eliminate the possibility that an initial stimulus can influence how people perceive and respond to subsequent stimuli (i.e., carry-over effects; Davis, 1995), and they reduce the risk that participants may respond differently based on inferences they make about the purpose of the study (i.e., demand characteristics; Orne, 1962; Rosenthal, 1976). Standard stimulus ordering techniques, such as counterbalancing and Latin square designs, eliminate systematic carry-over effects, but they do not eliminate demand characteristics. The only way to truly know whether or how the experimental design of a study affects the results is to run both and compare them. Testing *every idea* using both within- and between-subjects designs is neither necessary nor practical, but it may be problematic to draw firm conclusions about some research questions based on evidence from only one type of design.

Within- and between-subjects designs represent different types contexts in which people make choices. Some have argued that most judgments occur in the context of single, isolated episodes; "much of life resembles a between-subjects experiment" (Kahneman, 2000, p. 682, see also Nagel & Waldmann, 2013). In other words, separate evaluation may be more reflective of the

kinds of judgments that people are likely to make under “ordinary” circumstances because people are not usually confronted with choices between well-controlled “stimuli”, one after another. However, joint evaluation may be a better model of some contexts, such as when philosophers, jurists or policy-makers explicitly consider hypothetical alternatives, when consumers face an explicit comparison between goods, or when doctors have to make decisions about assigning scarce organs or blood to patients (Goodwin & Landy, in press; Li, Veitri, Galvani, & Chapman, 2010). In short, between-subjects designs are especially appropriate means to model judgments about a single option or event, whereas within-subjects designs are particularly well-suited to model choices between salient alternatives.

Contrasting joint and separate evaluation can also be a useful technique for determining the psychological mechanisms underlying a judgment process. For instance, Hsee (1996) showed that people would pay more for a new dictionary with 10,000 words than a used dictionary with 20,000 words and a torn cover when the dictionaries were evaluated by separate groups of people (i.e., a between-subjects design). But, when the dictionaries were evaluated jointly (i.e., a within-subjects design), this preference reversed; people would pay more for the slightly tattered dictionary with more words than the new dictionary with fewer words. This experiment illustrates a general property of joint versus separate evaluation. Under separate evaluation, people tend to rely on heuristic processes and qualitative distinctions (“new is better, tattered is worse”) whereas under the enriched evaluative context of joint evaluation, they tend to rely on the controlled application of explicit rules and models (“more words is better for a dictionary”).

In the literature on moral judgment, joint versus separate evaluations have been contrasted several times, often with the purpose of discriminating automatic and controlled processes (e.g., Bartels, 2008; Kogut & Ritov, 2005; Lombrozo, 2009; Paharia, Kassam, Greene, & Bazerman, 2009). Bartels (2008), for example, found that people evaluated harmful actions as more unfavorable than harmful omissions when evaluating instances separately. For example, participants who read about administrators who caused some deaths by administering a vaccine to prevent more deaths from a contagious outbreak of a disease reported stronger condemnation than participants who read about administrators who did not administer a vaccine and accepted a greater number of deaths from the disease. When participants read both version of the scenario, however, the trend reversed; they perceived actions that caused some harm to be more favorably than omissions that

allowed a greater amount of harm to occur. Somewhat similarly, Cushman, Young, and Hauser (2006) found that people feel that it is worse to harm a person to save others (as in the footbridge version of the trolley problem) by physically touching them (e.g., push the fat man off the footbridge) than by mechanical mediation (e.g., pulling a level that drops the fat man through a trap door) when judging scenarios separately. However, people tended to deny the relevance of this distinction when the scenarios were evaluated side-by-side.

In summary, researchers must be mindful of the potential impact of response formats on participants' choices. Decisions about whether to use within- or between-subjects designs (or both) should depend on the particular research question under examination.

### **3.3. Moving beyond the laboratory and traditional stimuli to accommodate individual and cultural variation**

Theories of morality typically try to identify universal moral principles while also accommodating differences in judgments and choices across settings and people. A major challenge when conducting moral judgment and decision making research is creating compelling laboratory stimuli that capture the essence of moral judgment processes that people deploy in the real world. Doing so is essential to ensure that results are sufficiently generalizable to support the type of inferences researchers would like to make. This challenge is particularly difficult in the moral domain.

Whereas one can often create consumer choices, visual or acoustic environments, and ordinary social encounters that mirror reality, there are obvious constraints on the ability to create genuinely wrenching moral tradeoffs and/or opportunities for high-stakes (im)moral behavior in a laboratory environment.

Hypothetical cases such as the trolley problem have been widely used, and they come with both advantages and disadvantages. Trolley problems are attractive laboratory stimuli because they provide easily modifiable stimuli to examine moral phenomena (Greene, 2009; Mikhail, 2009). As the literature reviewed above indicates, research using trolley problems (and many other types of hypothetical vignettes) have inspired a large body of productive research. And, many findings involving trolley problems have *prima facie* connections to core phenomena in the moral domain.

For instance, judgments elicited by trolley problems are compromised among people high in psychopathy and Machiavellianism (Bartels & Pizarro, 2011) and people with brain trauma that affects social decision-making (Koenigs et al., 2007). Judgments elicited by trolley problems also correlate with levels of empathy (Gleichgerrcht & Young, 2013). However, there are also notable drawbacks to using trolley-type dilemmas for moral judgment research. For instance, people might not accept the closed-world assumptions of these scenarios—they might doubt whether a fat man's body can stop a trolley car or whether this is the only available solution (Bennis, Medin, & Bartels, 2010). Moreover, even if they accept the constraints of the problem, many people find trolley problems to be amusing rather than sobering, and some evidence suggests that trolley problems do not always engage the same psychological processes as other moral situations (Bauman, McGraw, Bartels, & Warren, 2013). Trolley problems also focus on just one dimension of moral concern—physical harm—to the exclusion of moral concerns such as fairness, purity, and so forth (Haidt & Joseph, 2004). Also, importantly, relying on any one set of stimuli is problematic; if most studies on a given topic use highly similar stimuli, researchers cannot determine whether or how common features of the stimuli influence the results (Campbell & Fiske, 1959; Wells & Windschitl, 1999).

An important complement to vignette-based work, wholly different in its approach, is anthropological research, such as Alan Fiske's (1991) observations of relational differences between Americans and Africans, which serve as complementary views by which to test the validity of models built in the laboratory. Ginges, Atran, and colleagues' field studies on sacred values in suicide/martyrdom attacks and peace negotiations similarly serve as case studies to contrast the predictions of decision models that assume people's choices aim to produce desirable outcomes (Ginges & Atran, 2009; Ginges, Atran, Sachdeva, & Medin, 2011).

Emerging evidence for moral flexibility presents another challenge for theorists. Both vignette-based studies (e.g., Haidt, Kohler & Dias 1993; Piazza & Landy, 2013) and anthropological evidence (e.g., Shweder et al., 1987) converge on the conclusion that there is substantial individual and cross-cultural variation, across and within countries, in moral norms (reviewed in Sachdeva, Singh, & Medin, 2011). Differences in what people see as right and wrong may be limited by a narrow selection of participants in laboratory experiments, reflecting a broader concern that applies across diverse areas of psychological research (Henrich, Heine & Noranzayan, 2010). For instance, Fiske's (1991) field work suggests that what is or is not sacred depends on culture. Whereas land is

freely bought and sold in the US, land in areas of rural Western Africa is considered communal and thus it is wrong to sell. And, what people deem as right and wrong can change rather quickly within a culture. Consider the rapid changes in people's opinion about tobacco (Rozin, 1999) or gay marriage in the United States. In less than 12 years, support went from 65% opposing gay marriage to over 50% favoring it (Pew, 2013).

One major area of recent research on individual and cultural diversity in moral values comes from moral foundations theory (Haidt & Joseph, 2004). Drawing from anthropology, evolutionary psychology, and moral psychology, moral foundations theory seeks to identify how people's conception of virtues gives rise to moral concerns that differ across situations and across people (Graham et al., 2013; Haidt & Joseph, 2004, 2007). In particular, much of the research on moral foundations theory has sought to understand differences in people's attitudes about contemporary issues (some of which correlate with political affiliation) rather than choices in experimenter-created scenarios. The theory proposes that foundations are innate, evolved psychological mechanisms that are malleable insofar as they are influenced by culture and refined through personal experience (Graham et al., 2013; Haidt, 2008). A finite number of foundations are expected to account for all possible moral considerations and functions. Initially, the theory identified four foundations, including: (i) Suffering, to nurture and protect vulnerable people, (ii) Reciprocity, to manage cooperation and exploitation in exchange relationships, (iii) Hierarchy, to encourage subordinates to be obedient and deferent and charging leaders to provide guidance and protection, and (iv) Purity, to govern material consumption, taboo ideas, and sexual behavior. It later expanded to include (v) Loyalty, to encourage group-sustaining behavior, including desires to reward individual sacrifice for the good of the group and punishment social loafers and traitors. Although these five foundations have received the most attention, the current theorizing acknowledges that other foundations likely exist. Theorists have only recently begun to specify the criteria that define whether something is a foundation (Graham et al., 2013), but a substantial body of evidence suggests that moral attitudes cohere according to a structure that does not correspond closely to consequentialism, deontology, and other organizing principles used in research on moral judgment and decision making. Therefore, this work illustrates why it is important to look beyond the laboratory—methods and research questions have a reciprocal influence on each other, and researchers who use different methods are able to “see” different parts of the big picture.

In summary, the field of moral psychology is presented with new challenges and opportunities at a methodological level. A central theme is the need to integrate research across diverse traditions, so that the precision and control of laboratory and field approaches can benefit from the ecological validity and richness of anthropological data, and vice-versa. Understanding individual and cultural differences in moral judgment is one area that may benefit from such a synthesis.

## 4. UNDERSTUDIED AREAS

In this chapter, we have reviewed several themes in the literature—those focusing on unique aspects of moral beliefs, moral tradeoffs, and moral judgments about others' actions. In this section, we address three areas of research that are understudied, but have recently received some attention in the literature.

### 4.1. Moral Praise

Moral judgment and decision making research has largely focused judgments of impermissibility, wrongness, blame, and punishment. However, there has been comparatively less attention devoted to the positive side of morality. What are the features of a moral decision context that lead people to view an action as being so morally good that people should be praised for the action and/or that features that make us feel morally obligated to perform the action? For instance, many accounts of how these judgments are made assume that judgments of responsibility for good actions should be symmetric, even though differences between good and bad actions have long been acknowledged (e.g., Kant's notion of perfect and imperfect duties). The rules for positive responsibility judgments (i.e., how we determine what is praiseworthy) are understudied. Some existing evidence suggests that differences exist between the way that negative actions elicit blame and positive actions elicit praise. For example, Pizarro, Uhlmann, & Salovey (2003) found that praise is not discounted for impulsive positive actions in the same way that blame is discounted for impulsive negative actions. Also, as noted earlier, Knobe (2003) found that that positive side-effects of an action are viewed as having been less intentional than negative side-effects of an action. Additionally, Goodwin and Darley (2012) have shown that people see the moral wrongness of negative actions as more objective than the moral rightness of positive actions after controlling for the extremity of goodness/badness judgments. These studies suggest that blame and praise are not mirror images of each other, but there is much more to learn about the sources of these asymmetries.

Even less research has been conducted to understand the judgmental processes by which people decide that some acts are supererogatory (those acts that go “above-and-beyond” what duty requires). While a great deal of moral heroism falls under this category, much more attention has been given to the question of which behaviors are obligatory, forbidden, or permissible (all of which

tend to be discussed with examples in the negative domain). In short, the literature on praise—how, when, and why people arrive at the judgment that an act is praiseworthy—is dwarfed by literature on the negative side of moral judgment.

#### **4.2. Moral obligation to help**

Closely allied with the study of praise, the moral obligation to help others has received a great deal of attention in normative ethics—particularly in utilitarian philosophy—but it has received relatively less attention in research on moral judgment and decision making. For example, philosophers have asked why most people feel that letting a child drown in a nearby pond is less permissible than letting a child die of malnutrition in a famine-struck country (Singer, 1972; Unger, 1996; Kamm, 2007). Studies of people’s moral intuitions about these cases suggest that people feel morally obligated to help others in dire need (and/or disparage others who do not help) when the victim’s needs are salient to the agent and the agent can help effectively (Nagel & Waldmann, 2013), and when others are in-group members (Baron & Miller, 2001). However, people feel less morally obligated to help (and judge not helping as less wrong) when they focus on the number or proportion of individuals they cannot help rather than focus on the individuals that they can help, a phenomenon dubbed “futility thinking” (Bartels & Burnett, 2010; Unger, 1996) or “pseudo-inefficacy” (Dickert et al., 2012). Interestingly, strict utilitarians may be less susceptible to futility thinking than people with more flexible moral systems. People who endorse utilitarian actions in sacrificial dilemmas (e.g., those who judge throwing the fat man off the footbridge to save the lives of five others to be the morally right action) are also more likely to judge actions that do not save lives in non-sacrificial contexts (e.g., not sending back an envelope to UNICEF with a donation to help needy children in a faraway country) as being morally wrong (Conway, Bartels, & Pizarro, 2013). The studies we discuss in this section begin to identify some phenomena of interest, but much more remains to be learned about people’s sense of moral obligation.

#### **4.3. Moral character**

Moral character and virtue represent a small share of theorizing in contemporary moral judgment and decision making research. This proportion is surprising given that virtue ethics comprises a

major branch of philosophical normative theory; recall that 22% of ethicists polled by Cushman and colleague endorsed virtue ethics. There is, however, a growing literature suggesting that the narrow focus on how people evaluate specific actions (such as whether a given action is right or wrong, or whether a person deserves blame or praise for performing the action) is insufficient for explaining the flexibility with which people assess blame or praise. Mounting evidence suggests that when making moral judgments, people often care about persons rather than about individual acts—they sometimes act like lay virtue theorists (see Pizarro & Tannenbaum, 2011 for a review). A character-based approach is consistent with an extended body of evidence in on the psychology of social evaluation pointing to the primacy of the good/bad dimension when evaluating people. For instance, people make these sorts of evaluations with great ease; they evaluate others as good or bad quickly and automatically (Bar, Neta, & Linz, 2006; Todorov, Said, Engell & Oosterhof, 2008), they appear to make these evaluations across many cultures (Fiske, Cuddy, & Glick, 2007), and the evaluation of others as good or bad emerges early in development (it can be seen in 6-month-olds; Hamlin, Wynn, & Bloom, 2007; Kuhlmeier, Wynn, & Bloom, 2003).

Given the primacy of character judgments in social evaluation, theorists should expect people to be sensitive to cues that might be informative about a person's underlying character. And, in fact, traits that are viewed as providing information about others' moral character (whether good or bad) appear to have a particularly large impact on the global impressions formed of people who possess those traits (Goodwin, Piazza, & Rozin, in press). One cue to a person's character is whether they perform actions while under a calm, rational state of mind versus act on impulse. For instance, Pizarro, Uhlmann, and Salovey (2003) found that participants tended to discount blame if a negative act was committed impulsively—a person who impulsively hit someone in a fit of anger seemed to participants to be less responsible than someone who deliberately decided to hit someone, but positive acts that were committed impulsively (e.g., impulsively donating money to charity because of a strong sympathetic reaction) were not judged to be less praiseworthy than the same actions performed after deliberation. Pizarro et al. argued that respondents reached inferences about the actors' character. In a similar vein, Woolfolk, Doris, and Darley (2006) demonstrated that actors can sometimes be judged as morally responsible for their actions even if such actions were completely constrained by external circumstances. Woolfolk and colleagues described a scenario where a man was under a clear situational constraint that forced him to murder a passenger on an airplane (he was forced by hijackers to kill the person or else he and 10 others would be killed).

Participants had no problem holding him responsible for the murder if it was something he wanted to do anyway (that is, if he “identified” with the act). On the other hand, if participants believed that, while under identical situational constraints, the agent did not identify with the action—that in some sense the behavior felt “alien” to him—participants then discounted their attributions of responsibility.

Another cue that a person possesses a bad character is if she takes pleasure in the suffering of others (Ames & Johar, 2009). Any cue about the hedonic responses a perpetrator experiences while committing a negative act should influence judgments of character. Consistent with this idea, Goodwin, Gromet, and Darley (2011) recently demonstrated that “hedonic markers” that suggested the perpetrator took pleasure in a killing (e.g., cutting the body into pieces) negatively impacted judgments of a person’s character, and led participants to favor the death penalty as punishment. A similar pattern of results emerged when participants were directly told the agent had experienced positive affect during the commission of the murder.

These cues are seen as informative of moral character because they are signals to a person’s underlying mental state—their intentions and desires, and these properties are viewed as diagnostic of more stable moral traits. For instance, even harmless actions are judged as blameworthy if they contain these cues, such as if the agent appears callous or insensitive (Tannenbaum, Uhlmann & Diermeier, 2011), or appears to possess “wicked” underlying desires (Inbar, Pizarro & Cushman, 2012). Even when there are no clear moral violations or harmful consequences, perceiving a person as having a bad moral character changes how their actions are judged. In short, the kind of person you are may impact the sanctions you receive just as much as the nature of the act you committed.

## 5. CONCLUDING REMARKS

Moral thinking pervades everyday decision making, and so understanding the psychological underpinnings of moral judgment and decision making is an important goal for the behavioral sciences. Research that focuses on rule-based models makes moral decisions appear straightforward and rigid, but our review suggests that they more complicated. Our attempt to document the state of the field reveals the diversity of approaches that (indirectly) reveals the flexibility of moral decision making systems. Whether they are study participants, policy makers, or the person on the street,

people are strongly motivated to adhere to and affirm their moral beliefs—they want to make the right judgments and choices, and do the right thing. But what is right and wrong, like many things, depends in part on the situation. So while moral judgments and choices can be accurately characterized as using moral rules, they are also characterized by a striking ability to adapt to situations that require flexibility.

Consistent with this theme, our review suggests that context strongly influences which moral principles people use to judge actions and actors and that apparent inconsistencies across situations need not be interpreted as evidence of moral bias, error, hypocrisy, weakness, or failure. One implication of the evidence for moral flexibility we have presented is that it might be difficult for any single framework to capture moral judgments and decisions (and this may help explain why no fully descriptive and consensus model of moral judgment and decision making exists despite decades of research). While several interesting puzzle pieces have been identified, the big picture remains unclear. We cannot even be certain that all of these pieces belong to just one puzzle. Fortunately for researchers interested in this area, there is much left to be learned, and we suspect that the coming decades will budge us closer to a complete understanding of moral judgment and decision making.

### **ACKNOWLEDGEMENTS**

We thank Gideon Keren and George Wu for the opportunity to write this chapter, their feedback, and their patience. We also thank the Moral Research Lab (MoRL), Rob St. Louis, participants in Pete McGraw's judgment and decision making seminar, and especially Doug Medin for helpful comments and suggestions.

## REFERENCES

- Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, 63, 368-378.
- Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin*, 126, 556-574.
- Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart, & Winston.
- Ames, D. R., & Johar, G. V. (2009). I'll know what you're like when I see how you feel: How and when affective displays influence behavior-based impressions. *Psychological Science*, 20(5), 586-593.
- Amir, O., & Ariely, D. (2007). Decisions by rules: The case of unwillingness to pay for beneficial delays. *Journal of Marketing Research*, 44, 142-152.
- Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means visual imagery and moral judgment. *Psychological Science*, 23(8), 861-868.
- Amir, O., Ariely, D., & Mazar, N. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45, 633-634.
- Arsenio, W. (1988). Children's conceptions of the situational affective consequences of sociomoral events. *Child Development*, 59, 1611-1622.
- Arsenio, W., & Lover, A. (1995). Children's conceptions of socio-moral affect: Happy victimizers, mixed emotions, and other expectancies. In M. Killen & D. Hart (Eds.), *Morality in everyday life* (pp. 87-130). New York: Cambridge University Press.
- Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion*, 6(2), 269.
- Baron, J. (1996). Do no harm. In D. M. Messick and A. E. Tenbrunsel (Eds.) *Codes of conduct: Behavioral research into business ethics*, (pp. 197-213). New York: Russell Sage.
- Baron, J. (2000). *Thinking and deciding (3rd ed.)*. New York: Cambridge University Press.
- Baron, J., Gürçay, B., Moore, A. B., & Starcke, K. (2012). Use of a Rasch model to predict response times to utilitarian moral dilemmas. *Synthese*, 189(1), 107-117.
- Baron, J., & Leshner, S. (2000). How serious are expressions of protected values?. *Journal of Experimental Psychology: Applied*, 6(3), 183-194.
- Baron, J., & Ritov, I. (2009). Protected values and omission bias as deontological judgments. In D.M. Bartels, C.W. Bauman, L.J. Skitka, & D.L. Medin, Eds. (2009). *Moral Judgment and Decision Making: The Psychology of Learning and Motivation, Vol. 50*, (pp. 133-167). San Diego: Elsevier.
- Baron, J., & Spranca, M. (1997). Protected values. *Organizational Behavior and Human Decision Processes*, 70(1), 1-16.
- Baron, J. & Miller, J.G. (2000). Limiting the scope of moral obligations to help: A cross-cultural investigation. *Journal of Cross-Cultural Psychology*, 31, 703-725.
- Bartels, D. M. (2008). Principled moral sentiment and the flexibility of moral judgment and decision making. *Cognition*, 108(2), 381-417.
- Bartels, D.M., Bauman, C.W., Skitka, L.J., & Medin, D.L., Eds. (2009). *Moral Judgment and Decision Making: The Psychology of Learning and Motivation, Vol. 50*. San Diego: Elsevier

- Bartels, D. M., & Burnett, R. C. (2011). A group construal account of drop-in-the-bucket thinking in policy preference and moral judgment. *Journal of Experimental Social Psychology, 47*(1), 50-57.
- Bartels, D. M., & Medin, D. L. (2007). Are morally motivated decision makers insensitive to the consequences of their choices?. *Psychological Science, 18*(1), 24-28.
- Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition, 121*, 154-161.
- Bauman, C.W., McGraw, A.P., Bartels, D.M., & Warren, C. (2013) Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. Working paper, University of California-Irvine.
- Bauman, C. W., & Skitka, L. J. (2009a). In the mind of the perceiver: Psychological implications of moral conviction. In D.M. Bartels, C.W. Bauman, L.J. Skitka, & D.L. Medin, Eds. (2009). *Moral Judgment and Decision Making: The Psychology of Learning and Motivation, Vol. 50*, (pp. 339-362). San Diego: Elsevier.
- Bauman, C. W., & Skitka, L. J. (2009b). Moral disagreement and procedural justice: Moral mandates as constraints to voice effects. *Australian Journal of Psychology, 61*, 40-49.
- Berg-Cross, L. G. (1975). Intentionality, degree of damage, and moral judgments. *Child Development, 46*, 970-974.
- Bennis, W. M., Medin, D. L., & Bartels, D. M. (2010). The costs and benefits of calculation and moral rules. *Perspectives on Psychological Science, 5*(2), 187-202.
- Blasi, A. (1984). Moral identity: Its role in moral functioning. In J. Gewirtz and W. Kurtines (Eds.) *Morality, moral behavior, and moral development* (pp 128-139). New York: Wiley.
- Broeders, R., van den Bos, K., Muller, P.A., & Ham, J. (2011). Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology, 47*, 923-934.
- Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., & Marois, R. (2008). The neural correlates of third-party punishment. *Neuron, 60*(5), 930-940.
- Burns, Z. C., Caruso, E. M., & Bartels, D. M. (2012), Predicting premeditation: Future behavior is seen as more intentional than past behavior. *Journal of Experimental Psychology: General, 141*, 227-232.
- Borg, J. S., Hynes, C., Van Horn, J., Grafton, S. and Sinnott-Armstrong, W. 2006: Consequences, Action, and Intention as Factors in Moral Judgments: An fMRI Investigation. *Journal of Cognitive Neuroscience, 18*, 803-17.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological bulletin, 56*, 81-105.
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior and Organization, 81*, 1-8.
- Chomsky, N. (1957). *Syntactic Structures*. Mouton, the Hague.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge.

- Chugh, D., Banaji, M., & Bazerman, M. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. In Moore, D., Cain, D., Loewenstein, G., & Bazerman, M. (Eds.), *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy* (pp.74-95). Cambridge: Cambridge University Press.
- Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social cognitive and affective neuroscience*, 2(2), 84-92.
- Cohen, A.B., & Rozin, P. (2001). Religion and the morality of mentality. *Journal of Personality and Social Psychology*, 81, 697-710.
- Converse, P. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and discontent* (pp. 206-261). New York: Free Press.
- Conway, P., Bartels, D.M., & Pizarro, D.A. (2013). An exploration of genuinely utilitarian inclinations: Clarifying the relations between moral judgments, antisocial traits, and the moral obligation to help others in need. Working paper, University of Cologne.
- Critcher, C., Inbar, Y. & Pizarro, D.A. (2012). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, 4, 308-315.
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences*, 107(40), 17433-17438.
- Cushman, F. A. (2008). Crime and Punishment: Differential reliance on causal and intentional information for different classes of moral judgment. *Cognition*, 108, 353-380.
- Cushman, F. A., Young, L., & Hauser, M. D. (2006b). The Role of Reasoning and Intuition in Moral Judgments: Testing three principles of harm. *Psychological Science*, 17, 1082-1089.
- Cushman, F., Dreber, A., Wang, Y., & Costa, J. (2009). Accidental outcomes guide punishment in a “trembling hand” game. *PloS one*, 4(8), e6699.
- Cushman, F. A., Murray, D., Gordon-McKeon, S., Wharton, S., & Greene, J. D. (2011). Judgment before principle: Engagement of the frontoparietal control network in condemning harms of omission. *Social Cognitive and Affective Neuroscience*, 7, 888-895.
- Cushman, F., Young, L., & Greene, J. D. (2010). Our multi-system moral psychology: Towards a consensus view. In J. Doris, G. Harman, S. Nichols, J. Prinz, and W. Sinnott-Armstrong (Eds.), *The Oxford Handbook of Moral Psychology* (pp. 47-71). Oxford: Oxford University Press.
- Cushman, F., & Greene, J. D. (2012). Finding faults: How moral dilemmas illuminate cognitive structure. *Social neuroscience*, 7(3), 269-279.
- Cushman, F. A. (2013). Action, outcome and value: A dual-system framework for morality. *Personality and Social Psychology Review*, 17(3), 273-292.
- Darwall, S. (1998). *Philosophical ethics*. Boulder, CO: Westview Press.
- Darley, J. M. & Shultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, 41, 525-556.

- Darwall, S. (1998). Empathy, sympathy, care. *Philosophical Studies*, 89(2), 261-282.
- Davis, A. (1995). The experimental methods in psychology. In G. Breakwell, S., Hammond, & C. Fife Shaw (Eds.), *Research methods in psychology* (pp. 50-68). London, UK: Sage.
- Deghani, M., Atran, S., Iliev, R., Sachdeva, S., Medin, D.L. , & Ginges, J. (2010). Sacred values and conflict over Iran's nuclear program. *Judgment and Decision Making*, 5, 540-546.
- Deghani, M., Iliev, R., Sachdeva, S., Atran, S., Ginges, J., & Medin, D.L. (2009). Emerging sacred values: Iran's nuclear program. *Judgment and Decision Making*, 4, 930-933.
- Dickert, S., Västfjäll, D., Kleber, J., & Slovic, P. (2012). Valuations of human lives: normative expectations and psychological mechanisms of (ir)rationality. *Synthese*, 189(1), 95-105.
- Ditto, P., Pizarro, D.A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D.M. Bartels, C.W. Bauman, L.J. Skitka, & D.L. Medin, Eds. (2009). *Moral Judgment and Decision Making: The Psychology of Learning and Motivation, Vol. 50* (pp.307-338). San Diego: Elsevier.
- Ditto, P. H., & Liu, B. (2011). Deontological dissonance and the consequentialist crutch. *The social psychology of morality: Exploring the causes of good and evil*, 51-70.
- Ehrich, K. R., & Irwin, J. R. (2005). Willful ignorance in the request for product attribute information. *Journal of Marketing Research*, 42, 266-277.
- Feltz, A., & Cokely, E. T. (2008). The fragmented folk: More evidence of stable individual differences in moral judgments and folk intuitions. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* (pp. 1771-1776). Austin, TX: Cognitive Science Society.
- Fiddick, L., Spampinato, M. V., & Grafman, J. (2005). Social contracts and precautions activate different neurological systems: An fMRI investigation of deontic reasoning. *NeuroImage*, 28(4), 778-786.
- Fisher, R. A. (1935). *The design of experiments*. New York: Hafner.
- Fiske, A. P. (1991). *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing*. New York: Free Press.
- Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological Review*, 99(4), 689-723.
- Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77-83.
- Fiske, A. P., & Tetlock, P. E. (1997). Taboo trade-offs: reactions to transactions that transgress the spheres of justice. *Political Psychology*, 18(2), 255-297.
- Foot, P. (1967). The problem of abortion and the doctrine of the double effect in virtues and vices. *Oxford Review*, 5, 5-15.
- Ginges, J. & Atran, S. (2009). Instrumental reasoning over sacred values: An Indonesian case study. In D.M. Bartels, C.W. Bauman, L.J. Skitka, & D.L. Medin, Eds. (2009). *Moral Judgment and Decision Making: The Psychology of Learning and Motivation, Vol. 50* (pp. 193-206). San Diego: Elsevier.

- Ginges, J., Atran, S., Medin, D., & Shikaki, K. (2007). Sacred bounds on rational resolution of violent political conflict. *Proceedings of the National Academy of Sciences*, *104*(18), 7357-7360.
- Ginges, J., & Atran, S. (2011). War as a moral imperative (not just practical politics by other means). *Proceedings of the Royal Society: Biological Sciences*.
- Ginges, J., Atran, S., Sachdeva, S., & Medin, D. (2011). Psychology out of the laboratory: The challenge of violent extremism. *American Psychologist*, *66*(6), 507.
- Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research*, *21*(2), 241-253.
- Gleichgerricht, E., & Young, L. (2013). Low levels of empathic concern predict utilitarian moral judgment. *PloS one*, *8*(4), e60418.
- Goodin, R. E. (1993). Utility and the good. In P. Singer (Ed.), *A companion to ethics* (pp. 241-248). Malden, MA: Blackwell Publishing.
- Goldstein, W. M., & Weber, E. U. (1995). Content and discontent: Indications and implications of domain specificity in preferential decision making. In J. Busemeyer, D. L. Medin & R. Hastie (Eds.), *The psychology of learning and motivation: Decision making from a cognitive perspective*, Vol. 32 (pp. 83-136). San Diego, CA: Academic Press.
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, *106*(3), 1339-1366.
- Goodwin, G. P., & Darley, J. (2012). Why are some moral beliefs seen as more objective than others? *Journal of Experimental Social Psychology*, *48*, 250-256.
- Goodwin, G.P., Gromet, D.G., & Darley, J.M. (2011). Pleasure at another's pain: The influence of hedonic states on attributions of evil. Paper presented at the meeting of the Society for Personality and Social Psychology, San Antonio, TX.
- Goodwin, G. P., & Landy, J. (in press). Valuing different human lives. *Journal of Experimental Psychology: General*.
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person perception and evaluation. *Journal of Personality and Social Psychology*, *106*, 148-168.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55-130.
- Gray, K., & Schein, C. (2012). Two minds vs. two philosophies: Mind perception defines morality and dissolves the debate between deontology and utilitarianism. *Review of Philosophy and Psychology*, *3*(3), 405-423.
- Gray, K., & Wegner, D. M. (2011). To escape blame, don't be a hero—Be a victim. *Journal of Experimental Social Psychology*, *47*(2), 516-519.
- Greene, J. D. (2007). Why are VMPFC patients more utilitarian? A dual-process theory of moral judgment explains. *Trends in cognitive sciences*, *11*(8), 322-323.

- Greene, J.D. (2009). Fruit flies of the moral mind. In M. Brockman (Ed.). *What's next: Dispatches from the future of science*. New York: Vintage.
- Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*, 364-371.
- Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44*(2), 389-400.
- 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.
- Grice, P. (1975). Logic of conversation. In P. Cole, & J.L. Morgan (Eds.), *Syntax and Semantics*\_(Vol. 3, Speech Acts, pp. 41-58). New York: Academic Press.
- Goodwin, G. P., Gromet, D. G., & Darley, J. M. (2011). Pleasure at another's pain: The influence of hedonic states on attributions of evil. Paper presented at the Twelfth Annual Conference of the Society for Personality and Social Psychology. San Antonio, USA.
- Guglielmo, S., Monroe, A. E., & Malle, B. F. (2009). At the heart of morality lies folk psychology. *Inquiry, 52*, 449-466.
- Guglielmo, S. & Malle, B. F. (2010). Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personality and Social Psychology Bulletin, 36*, 1635-1647.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814-834.
- Haidt, J. (2008). Morality. *Perspectives on Psychological Science, 3*, 65-72.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*, 55-66.
- Haidt, J., & Joseph, C. (2007). The moral mind: How 5 sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules. In P. Carruthers, S. Laurence & S. Stich (Eds.), *The Innate Mind, Vol. 3* (pp. 367-391). New York: Oxford.
- Haidt, J., Koller, S. H., & Dias, M. G. (1993). Affect, culture, and morality, or is it wrong to eat your dog?. *Journal of personality and social psychology, 65*, 613-628.
- Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology, 33*(1), 1-36.
- Hall, J. (1947). *General principles of criminal law*. Indianapolis, IN: Bobbs-Merrill Company
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal Infants. *Nature, 450*(7169), 557-559.
- Hare, R. M. (1981). *Moral thinking: Its levels, method, and point*. Oxford: Clarendon Press.
- Hart, H. L. A., & Honore, T. (1959). *Causation in the law*. Oxford, UK: Clarendon Press.
- Haslam, N. (Ed.) (2004). *Relational models theory: A contemporary overview*. Mahwah, NJ: Erlbaum.

- Hauser, M., Cushman, F., Young, L., Kang-Xing Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & language*, 22(1), 1-21.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Holleman, B. (1999). Wording effects in survey research using meta-analysis to explain the forbid/allow asymmetry. *Journal of Quantitative Linguistics*, 6(1), 29-40.
- Hume, D. (1888). *A treatise on human nature*. Oxford, England: Clarendon Press. (Original work published 1777)
- Hsee, C. K. (1996) The evaluability hypothesis: An explanation for preference-reversal between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes*, 67, 247-257.
- 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.
- Iliev, R., Sachdeva, S., Bartels, D. M., Joseph, C., Suzuki, S., & Medin, D. L. (2009). Attending to moral values. In D.M. Bartels, C.W. Bauman, L.J. Skitka, & D.L. Medin, Eds. (2009). *Moral Judgment and Decision Making: The Psychology of Learning and Motivation*, Vol. 50 (pp. 169–192). San Diego: Elsevier
- Iliev, R., Sachdeva, S., & Medin, D.L. (2012). Moral kinematics: The role of physical factors in moral judgments. *Memory & Cognition*, 40, 1387-1401
- Kagan, S. (1998). *Normative ethics*. Boulder, CO: Westview Press.
- Kahane, G., & Shackel, N. (2010). Methodological issues in the neuroscience of moral judgment. *Mind & Language*, 25(5), 561-582.
- 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, 393-402.
- Kahneman, D. (2000). A psychological point of view: Violations of rational rules as diagnostic of mental processes (Reply to Stanovich & West). *Behavioral and Brain Sciences*, 23, 681-683
- Kahneman, D., Knetsch, J.L., & Thaler, R.H. (1986). Fairness as a constraint on profit-seeking: Entitlements in the market. *The American Economic Review*, 76, 728-741.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263-291.
- Kamm, F. M. (2007). *Intricate ethics: Rights, responsibilities, and permissible harm*. Oxford, UK: Oxford University Press.
- Kant, I. (2002). *Groundwork for the metaphysics of morals*. (A. W. Wood, Trans.). Binghamton, NY: Vail-Ballou Press. (Original work published 1785)
- Kelly, D., Stich, S., Haley, S., Eng, S.J., & Fessler, D. (2007). Harm, affect, and the moral/conventional distinction. *Mind & Language*, 22, 117–131
- Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, 63, 190-193.
- Knobe, J. (2003). The concept of intentional action. *Philosophical Studies*, 130, 203-231.

- Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., Hauser, M., & Damasio, A. (2007). Damage to the prefrontal cortex increases utilitarian moral judgments. *Nature*, *446*(7138), 908-911.
- Kogut, T., & Ritov, I. (2005). The “identified victim” effect: An identified group, or just a single individual? *Journal of Behavioral Decision Making*, *18*(3), 157-167.
- Kohlberg, L. (1976). Moral stages and moralization: The cognitive developmental approach. In T. Lickona (Ed.), *Moral development and behavior: Theory, research and social issues* (pp. 31-53). New York: Holt, Rinehart, & Winston.
- Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of justice*. San Francisco: Harper & Row.
- Kristiansen, C. M., & Hotte, A. (1996). Morality and the self: Implications for the when and how of value-attitude-behavior relations. In C. Seligman, J. M. Olson & M. P. Zanna (Eds.), *The psychology of values: The Ontario symposium*. Hillsdale, NJ: Lawrence
- Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science*, *14*(5), 402-408.
- Kvaran, T., & Sanfey, A. G. (2010). Toward an integrated neuroscience of morality: the contribution of neuroeconomics to moral cognition. *Topics in Cognitive Science*, *2*(3), 579-595.
- Laupa, M. (1994). “Who’s in charge?” Preschool children’s concepts of authority. *Early Childhood Research Quarterly*, *9*(1), 1-17.
- Lerner, J.S., Goldberg, J.H., & Tetlock, P.E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin*, *24*, 563-574.
- Lewin, K. (1935). *A dynamic theory of personality*. New York, NY: McGraw-Hill.
- Lewin, K. (1951). *Field theory in social science*. New York, NY: McGraw-Hill.
- Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting Intentionally and the Side-Effect Effect Theory of Mind and Moral Judgment. *Psychological Science*, *17*(5), 421-427.
- Li, M., Vietri, J., Galvani, A. P., & Chapman, G. B. (2010). How do people value life? *Psychological Science*, *21*, 163-167.
- Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science*, *33*, 273-286.
- Luce, M. F., Payne, J. W., & Bettman, J. R. (1999). Emotional tradeoff difficulty and choice. *Journal of Marketing Research*, *36*, 143-159.
- Luce, M. F., Payne, J. W., & Bettman, J. R. (2000). Coping with unfavorable attribute values in choice. *Organizational Behavior and Human Decision Processes*, *81*(2), 274-299.
- Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*(3), 530-542.
- Malle, B. F., Guglielmo, S., & Monroe, A. E. (2012). Moral, cognitive, and social: The nature of blame. In J. Forgas, K. Fiedler, & C. Sedikides (Eds.), *Social Thinking and Interpersonal Behavior* (pp. 311-329). Philadelphia, PA: Psychology Press.
- March, J., & Heath, C. (1994). *A primer on decision making*. New York: Free Press.

- McGraw, A. P., Schwartz, J. A., & Tetlock, P. E. (2012). From the commercial to the communal: Reframing taboo tradeoffs in religious and pharmaceutical marketing. *Journal of Consumer Research*, 39(1), 157-173.
- McGraw, A. P., & Tetlock, P. (2005). Taboo tradeoffs, relational framing, and the acceptability of exchanges. *Journal of Consumer Psychology*, 15(1), 2-15.
- McGraw, A. P., Tetlock, P. E., & Kristel, O. V. (2003). The limits of fungibility: Relational schemata and the value of things. *Journal of Consumer Research*, 30(2), 219-229.
- McGraw, A.P. & Warren, C. (2010). Benign violations: Making immoral behavior funny. *Psychological Science*, 21, 1141-1149.
- McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. (2009). A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, 45(3), 577-580.
- McLaughlin, J. A. (1925). Proximate cause. *Harvard Law Review*, 39(2), 149-199.
- Mellers, B. A. (2000). Choice and the relative pleasure of consequences. *Psychological Bulletin*, 126(6), 910-924.
- Mikhail, J. (2000). Rawls' Linguistic Analogy: A Study of the 'Generative Grammar' Model of Moral Theory Described by John Rawls in "A Theory of Justice." Unpublished PhD Dissertation, Cornell University.
- Mikhail, J. (2005). Moral Heuristics or Moral Competence? Reflections on Sunstein. *Behavioral and Brain Sciences*, 28(4), 557-558.
- Mikhail, J. (2007). Universal Moral Grammar: Theory, Evidence, and the Future. *Trends in Cognitive Sciences*, 11, 143-152.
- Mikhail, J. (2009). Moral grammar and intuitive jurisprudence: a formal model of unconscious moral and legal knowledge. In D.M. Bartels, C.W. Bauman, L.J. Skitka, & D.L. Medin, Eds. (2009). *Moral Judgment and Decision Making: The Psychology of Learning and Motivation*, Vol. 50. (pp. 27-99). San Diego: Elsevier
- Miller, J.G. & Bersoff, D.M. (1992). Cultural and moral judgment: How are conflicts between justice and interpersonal responsibilities resolved? *Journal of Personality and Social Psychology*, 62, 541-554.
- Miller, R., Hannikainen, I., & Cushman, F. (2013). Bad actions or bad outcomes? Differentiating affective contributions to the moral condemnation of harm. *Manuscript submitted for publication*.
- Mischel, W. (1968). *Personality and assessment*. London: Wiley.
- Merritt, A. C., Effron, D. A., & Monin, B. (2010). Moral self-licensing: When being good frees us to be bad. *Social and Personality Psychology Compass*, 4(5), 344-357.
- Moll, J., De Oliveira-Souza, R., & Zahn, R. (2008) The neural basis of moral cognition. *Annals of the New York Academy of Sciences*, 1124, 161-180.
- Monin, B. & Merritt, A. (2012). Moral hypocrisy, moral inconsistency, and the struggle for moral integrity. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of*

- good and evil. Herzliya series on personality and social psychology.*, (pp. 167-184). Washington DC, US: American Psychological Association.
- Monin, B., Pizarro, D., & Beer, J. (2007). Deciding vs. reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology*, *11*, 99-111.
- 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*, 549–557.
- Moore, G. E. (1903). *Principia Ethica*. Cambridge University Press, New York: NY.
- Moretto, G., Làdavas, E., Mattioli, F., & Di Pellegrino, G. (2010). A psychophysiological investigation of moral judgment after ventromedial prefrontal damage. *Journal of Cognitive Neuroscience*, *22*(8), 1888-1899.
- Morgan, G. S., & Skitka, L. J. (2013). Universally and objectively true: Psychological foundations of moral conviction. Manuscript submitted for publication.
- Nagel, J., & Waldmann, M. R. (2013). Deconfounding distance effects in judgments of moral obligation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 237-252
- Nichols, S. (2002). Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, *84*, 221-236.
- Nichols, S., & Mallon, R. (2006). Moral dilemmas and moral rules. *Cognition*, *100*(3), 530-542.
- Nucci, L. P. (2001). *Education in the moral domain*. Cambridge University Press.
- Nucci, L. P., & Turiel, E. (1978). Social interactions and the development of social concepts in preschool children. *Child Development*, *49*(2), 400-407.
- Nucci, L. P., & Gingo, M. (2011). The development of moral reasoning. In U. Goswami (Ed.) *The Wiley-Blackwell Handbook of Childhood Cognitive Development* (pp. 420-445). Malden, MA: Wiley-Blackwell.
- O’Hara, R. E. O., Sinnott-Armstrong, W., Sinnott-Armstrong, N. A. (2010). Wording effects in moral judgments. *Judgment and Decision Making*, *5*, 547-554.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, *17*(11), 776–783.
- Paharia, N., Kassam, K. S., Greene, J. D., & Bazerman, M. H. (2009). Dirty work, clean hands: The moral psychology of indirect agency. *Organizational Behavior and Human Decision Processes*, *109*, 134-141.
- Petrinovich, L., O’Neill, P., & Jorgensen, M. (1993). An empirical study of moral intuitions: Toward an evolutionary ethics. *Journal of Personality and Social Psychology*, *64*, 467-478.
- Pew Research (2013). Gay marriage: Key data points from Pew Research, retrieved from <http://www.pewresearch.org/key-data-points/gay-marriage-key-data-points-from-pew-research/>
- Piaget, J. (1965). *The moral judgment of the child* (M. Gabain, Trans.). New York: Free Press. (Original work published 1932).

- Piazza, J., & Landy, J. F. (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.
- Piazza, J., Sousa, P., & Holbrook, C. (2013). Authority dependence and judgments of utilitarian harm. *Cognition*, 128, 261-270.
- Pincoffs, E. L. (1986). *Quandaries and virtues: Against reductivism in ethics*. Lawrence, KS: University Press of Kansas.
- Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P.R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 91–108). Washington, DC: American Psychological Association..
- Pizarro, D. A., & Uhlmann, E. L. (2005). Do normative standards advance our understanding of moral judgment?. *Behavioral and Brain Sciences*, 28(4), 558-558.
- Pizarro, D. A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, 39(6), 653-660.
- Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in Judgments of Moral Blame and Praise The Role of Perceived Metadesires. *Psychological Science*, 14(3), 267-272.
- Provencher, H. L., & Fincham, F. D. (2000). Attributions of causality, responsibility, and blame for positive and negative symptom behaviours in caregivers of persons with schizophrenia. *Psychological Medicine*, 30, 899-910.
- Rawls, J. (1999). *A Theory of Justice*, Revised Edition. Cambridge, MA: Harvard University Press.
- 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.
- Rettinger, D. A., & Hastie, R. (2001). Content effects on decision making. *Organizational Behavior and Human Decision Processes*, 85(2), 336-359.
- Robinson, P. H., Jackowitz, S.E., & Bartels, D.M. (2012), Extralegal punishment factors: A study of forgiveness, hardship, good deeds, apology, remorse, and other such discretionary factors in assessing criminal punishment. *Vanderbilt Law Review*, 65, 737-826.
- Rokeach, M. (1973). *The nature of human values*. New York: Free press.
- Rosenthal, R. (1976). *Experimenter effects in behavioral research*, 2nd ed. New York: Wiley.
- Royzman, E. B., & Baron, J. (2002). The preference for indirect harm. *Social Justice Research*, 15(2), 165-184.
- Royzman, E. B., Landy, J. F., & Leeman, R. F. (in press). Are thoughtful people more utilitarian? CRT as a unique predictor of moral minimalism in the dilemmatic context. *Cognitive Science*.
- Rozin, P. (1999). The process of moralization. *Psychological Science*, 10(3), 218-221.
- Ritov, I., & Baron, J. (1999). Protected values and omission bias. *Organizational Behavior and Human Decision Processes*, 79(2), 79-94.

- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners: The paradox of moral self-regulation. *Psychological Science*, 20, 523–528.
- Sachdeva, S., & Medin, D. L. (2008). Is it more wrong to care less? The effects of “more” and “less” on the quantity (in)sensitivity of protected values. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1239–1243). Austin, TX: Cognitive Science Society.
- Sachdeva, S., Singh, P., & Medin, D. (2011). Culture and the quest for universal principles in moral reasoning. *International Journal of Psychology*, 46(3), 161-176.
- Savage, L. J. (1954) *The foundations of statistics*. New York: Wiley.
- Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, 27(2), 135-153.
- Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge, UK: Cambridge University Press.
- Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual review of psychology*, 53(1), 491-517.
- Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.
- Shweder, R. A., Mahapatra, M., & Miller, J. G. (1987). Culture and moral development. In J. Kagan & S. Lamb (Eds.), *The emergence of morality in young children* (pp. 1-83) Chicago: University of Chicago Press.
- Shweder, R. (2004). Tuskegee re-examined. *Spiked*. Retrieved from <http://www.spiked-online.com/Articles/0000000CA34A.htm>.
- Singer, P. (1972). Famine, affluence, and morality. *Philosophy and Public Affairs*, 1(3), 229-243..
- Sinnott-Armstrong, W. (Ed.) (2008). *Moral Psychology, Volume 1: The Evolution of Morality*. Cambridge, MA: MIT Press,
- Sinnott-Armstrong, W. (Ed.) (2008). *Moral Psychology, Volume 2: The Cognitive Science of Morality*. Cambridge, MA: MIT Press,
- Sinnott-Armstrong, W. (Ed.) (2008). *Moral Psychology, Volume 3: The Neuroscience of Morality*. Cambridge, MA: MIT Press,
- Sinnott-Armstrong, W. (2008). Abstract + Concrete = Paradox. In S. Nichols & J. Knobe (Eds.), *Experimental Philosophy* (pp. 209-230). New York: Oxford University Press.
- Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, 4(4), 267-281.
- Skitka, L. J., & Bauman, C. W. (2008). Moral conviction as a political motivator: Does it lead to a conservative electoral advantage? *Political Psychology*, 29, 29-54.
- Skitka, L. J., Bauman, C. W., & Lytle, B. L. (2009). Limits on legitimacy: Moral and religious convictions as constraints on deference to authority. *Journal of Personality and Social Psychology*, 97, 567-578.

- Skitka, L. J., Bauman, C. W., & Sargis, E. G. (2005). Moral conviction: another contributor to attitude strength or something more?. *Journal of personality and social psychology*, 88(6), 895.
- Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.
- Sloman, S. A., Fernbach, P. M. & Ewing, S. (2012). A causal model of intentionality judgment. *Mind & Language*, 27(2), 154-180.
- Smetana, J. G. (1981). Preschool children's conceptions of moral and social rules. *Child Development*, 52, 1333-1336.
- Smetana, J. G. (1985). Preschool children's conceptions of transgressions: Effects of varying moral and conventional domain-related attributes. *Developmental Psychology*, 21(1), 18.
- Smith, M. (1995). *The moral problem*. Oxford: Blackwell.
- Sondak, H., and T. R. Tyler. (2001). What shouldn't money buy? The psychology of preferences for market solutions to allocation problems. Unpublished manuscript, University of Utah.
- Sousa, P., Holbrook, C., & Piazza, J. (2009). The morality of harm. *Cognition*, 113, 80-92.
- Spranca, M., Minsk, E., & Baron, J. (1991). Omission and commission in judgment and choice. *Journal of Experimental Social Psychology*, 27(1), 76-105.
- Sunstein, C. R. (2005). Moral heuristics. *Behavioral and Brain Sciences*, 28(4), 531-541.
- Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition*, 119(3), 454-458.
- Tannenbaum, D., Uhlmann, E. L., & Diermeier, D. (2011). Moral signals, public outrage, and immaterial harms. *Journal of Experimental Social Psychology*, 47(6), 1249-1254.
- Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly*, 44(4), 684-707.
- Tenbrunsel, A. E., & Smith-Crowe, K. (2008). Ethical decision making: Where we've been and where we're going. *The Academy of Management Annals*, 2(1), 545-607.
- Tenbrunsel, A. E., Diekmann, K. A., Wade-Benzoni, K. A., & Bazerman, M. H. (2010). The ethical mirage: A temporal explanation as to why we are not as ethical as we think we are. *Research in Organizational Behavior*, 30, 153-173
- Thomson, J. J. (1985). Double effect, triple effect and the trolley problem: Squaring the circle in looping cases. *Yale Law Journal*, 94, 1395-1415.
- Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of management*, 32(6), 951-990.
- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: intuitive politicians, theologians, and prosecutors. *Psychological Review*, 109(3), 451.
- Tetlock, P. E., Kristel, O. V., Elson, S. B., Green, M. C., & Lerner, J. S. (2000). The psychology of the unthinkable: taboo tradeoffs, forbidden base rates, and heretical counterfactuals. *Journal of personality and social psychology*, 78(5), 853.

- Tetlock, P. E., Peterson, R. S., & Lerner, J. S. (1996). Revising the value pluralism model: Incorporating social content and context postulates. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.), *The Ontario symposium: The psychology of values* (Vol. 8, pp. 5–51). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tetlock, P.E. (2005) Gauging the heuristic value of heuristics, reflections on Sunstein. *Behavioral and Brain Sciences*, 28, 562-563.
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences*, 12(12), 455-460.
- Trémolière, B., Neys, W. D., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition*, 124, 379-384.
- Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.
- Turiel, E. (2002). *The culture of morality: Social development, context, and conflict*. Cambridge, UK: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4), 297-323.
- Uhlmann, E. L., Pizarro, D. A., Tannenbaum, D., & Ditto, P. H. (2009). The motivated use of moral principles. *Judgment and Decision Making*, 4(6), 476-491.
- Unger, P. (1996). *Living high and letting die: Our illusion of innocence*. New York, NY: Oxford University Press.
- Uttich, K. & Lombrozo, T. (2010). Norms inform mental state ascriptions: A rational explanation for the side-effect effect. *Cognition*, 116, 87-100.
- Von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior*, 2<sup>nd</sup> edition. Princeton, NJ: Princeton University Press.
- Waldmann, M. R., Nagel, J., & Wiegmann, A. (2012). Moral judgment. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 364-389). New York: Oxford University Press.
- 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.
- Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, 25(9), 1115-1125.
- Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, 100(2), 283-301.
- Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Sciences*, 104(20), 8235-8240.

Young, L., Nichols, S., & Saxe, R. (2010). Investigating the neural and cognitive basis of moral luck: It's not what you do but what you know. *Review of Philosophy and Psychology*, 1(3), 333-349.