ANNUAL REVIEWS

Annual Review of Psychology

Moral Judgments

Bertram F. Malle

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Providence, Rhode Island 02912, USA; email: bfmalle@brown.edu

Annu. Rev. Psychol. 2021. 72:293–318

First published as a Review in Advance on September 4, 2020

The *Annual Review of Psychology* is online at psych.annualreviews.org

https://doi.org/10.1146/annurev-psych-072220-104358

Keywords

moral psychology, ethics, social psychology, moral cognition, social cognition

Abstract

Research on morality has increased rapidly over the past 10 years. At the center of this research are moral judgments—evaluative judgments that a perceiver makes in response to a moral norm violation. But there is substantial diversity in what has been called moral judgment. This article offers a framework that distinguishes, theoretically and empirically, four classes of moral judgment: evaluations, norm judgments, moral wrongness judgments, and blame judgments. These judgments differ in their typical objects, the information they process, their speed, and their social functions. The framework presented here organizes the extensive literature and provides fresh perspectives on measurement, the nature of moral intuitions, the status of moral dumbfounding, and the prospects of dual-process models of moral judgment. It also identifies omitted questions and sets the stage for a broader theory of moral judgment, which the coming decades may bring forth.

Contents

INTRODUCTION 294
WHAT IS (ARE) MORAL JUDGMENT(S)? 294
MORAL JUDGMENTS: AN ORGANIZING FRAMEWORK 295
Evaluations 295
Norm Judgments 296
Wrongness Judgments 297
Blame Judgments 299
Features of Four Moral Judgments in Summary 302
ALMOST MORAL JUDGMENTS 303
Moral Character 303
Punishment 304
APPLYING THE FRAMEWORK 306
Measurement 306
Moral Intuitions 307
Moral Dumbfounding 308
Dual-Process Models Reconsidered 309

Supplemental Material >

INTRODUCTION

Moral psychology has gained significant presence over the past decade, showing a fourfold increase in the number of published articles across numerous journals in psychology (see **Supplemental Appendix 1**). Several *Annual Review of Psychology* articles from the past decade have covered aspects of moral behavior such as altruism (Kurzban et al. 2015), prosociality (Hare 2017, Keltner et al. 2014), and cooperation (Tomasello & Vaish 2013), and two additional reviews have focused on moral justification of behavior (Mullen & Monin 2016) and on the role of religion in moral behavior (Bloom 2011). But moral behavior is only one side of moral psychology. Another is moral cognition, which encompasses the psychological processes that allow people to recognize, interpret, and evaluate moral and immoral behavior.

Perhaps the core concept of moral cognition is moral judgment. A remarkable variety of moral judgments have been investigated: from evaluating behaviors as right or wrong to making inferences about a person's moral character to endorsing values or policies. Understanding moral cognition requires understanding how these kinds of moral judgments both differ from one another and relate to other phenomena, such as norms, emotions, and punishment. This article offers a framework to organize the literature on moral judgment by distinguishing between four major classes of moral judgment that differ in their typical objects of judgment, the information they process, and their social functions. After reviewing research that supports such classes of moral judgment, I test the value of this framework in addressing thorny questions of how to measure moral judgment, what moral intuitions are, whether people suffer from moral dumbfounding, and the promise of a dual-process approach to moral judgments. I must omit many topics, including the controversial involvement of affect and emotion in moral judgment, which would warrant its own review.

WHAT IS (ARE) MORAL JUDGMENT(S)?

With as prominent a term as *moral judgment*, we might expect converging definitions and measurements. But what falls under this term is strikingly diverse. Consider a selection of measured variables below (with only one representative citation for each):

- how "negative or positive" (Cannon et al. 2011) or "bad" a behavior is (Zalla et al. 2011);
- whether one is "disapproving of" a behavior (Van Dillen et al. 2012);
- whether a behavior is "wrong" (Schnall et al. 2008), "morally wrong" (Wheatley & Haidt 2005), or "OK" to "extremely wrong" (Cheng et al. 2013);
- whether a behavior is "acceptable" versus "forbidden" (Young et al. 2012), "permissible," (Mikhail 2011), or "appropriate" (Greene et al. 2001);
- whether a described agent "should" act a certain way (Gold et al. 2015) or whether a choice is "obligatory" (Koralus & Alfano 2017);
- how "blameworthy" a behavior is (Siegel et al. 2017), how "morally blameworthy the agent is" (Young et al. 2010), or "[h]ow much blame [agent] deserves" (Cushman 2008); and
- how immoral the agent is (Royzman et al. 2011).

The above list omits positive moral judgments (far less frequently studied than negative ones) and value judgments, both of which are beyond the scope of this review. The list also omits measures of *responsibility*, a term that was common in the moral psychology literature of the 1970s to 1990s but, suffering from serious ambiguities (Gailey & Falk 2008, Guglielmo 2015), has been almost entirely abandoned in recent moral psychology research. Despite these omissions, the wide range of measures for allegedly the same phenomenon of moral judgment is remarkable. Some scholars have taken note of this diversity and treated it as a methodological challenge (Barbosa & Jiménez-Leal 2017, Bartels et al. 2015, Kahane & Shackel 2010, O'Hara et al. 2010). But it is a theoretical challenge as well, as the divergent measurements prevent us from generalizing findings from one study to the next and have led to substantial disagreement over whether there is anything unifying about moral judgment (for contrasting views, see Goodwin 2017, Gray et al. 2014, Sinnott-Armstrong & Wheatley 2012). In the absence of consensus, Sinnott-Armstrong (2016, pp. 350, 351) suggests that "moral science needs to shift towards taxonomic rigor" by "drawing distinctions among different kinds of moral judgments." This review offers a framework of such distinctions.

Responsibility: term with many meanings, including causality; obligation ("our responsibilities"); capacity to make moral decisions; deservingness of blame ("Who is responsible?")

Evaluative conditioning: Pavlovian learning in which evaluation of stimulus A is transferred to evaluation of B when A and B are paired

Electromyogram (EMG): recording of electrical activity produced by skeletal muscles, such as facial muscles

MORAL JUDGMENTS: AN ORGANIZING FRAMEWORK

The focus of this section is on research regarding judgments a perceiver makes about a morally significant event (e.g., an action or accident) and evidence for the varied psychological responses that flow from this event and generate distinct moral judgments. Four major classes of such judgments have received substantial attention in moral psychology research: evaluations, norm judgments, wrongness judgments, and blame judgments. **Figure 1** organizes these four classes of judgment, hinting at a potential hierarchy from simple to complex information processing.

Evaluations

The first class of judgment to consider consists of evaluations of good and bad, positive and negative. Evaluations represent one of the most basic human responses, can be easily transferred from one object to another through evaluative conditioning (De Houwer et al. 2001), and can be made about virtually anything: from written characters, sounds, and objects all the way to human decisions, unintended outcomes, persons, and groups. Evaluations of moral stimuli in particular exhibit very fast onset—within 300–600 ms when measuring the earliest electrophysiological responses (Leuthold et al. 2015, Yoder & Decety 2014) and within 1,000 ms when measuring facial electromyogram (EMG) responses ('t Hart et al. 2018). It seems clear that the brain can quickly, under favorable conditions, distinguish bad from good stimuli. Whether these early evaluation

Figure 1

Four major classes of moral judgment prominent in current moral psychology research.

stages are genuinely moral may be doubted, but even if the earliest evaluative processes do not yet specifically code for moral valence, they quickly prepare the organism to process the stimulus for potential moral significance. Likewise, whether fast (moral) evaluations constitute genuine judgments may be doubted, but they enable people to make subsequent conscious judgments of moral badness, which are themselves quite fast, taking approximately 1,600 ms (Cusimano et al. 2017).

Evaluation is typically considered fundamentally affective, but that need not always be the case. If affect encompasses both valence and arousal, consciously experienced (Russell & Barrett 1999), then evaluation based solely on valence may not yet constitute affect and preconscious evaluation would not be an affective experience. Indeed, evaluative priming can occur without feelings (Niedenthal et al. 2003); very early markers of evaluation seem to emerge faster than markers of emotional arousal (Gui et al. 2016); and people seem to make badness judgments faster than they report feelings or specific emotions (Cusimano et al. 2017).

Fast moral evaluation cannot take all morally relevant information into account. When people witness a prototypical norm-violating behavior such as a shove to the ground, they may rapidly evaluate it as bad, but their judgment may change once they process the person's goal (e.g., to hurt or save the other) or assess the amount of damage done. Morally relevant information itself unfolds over time—whether as live behavior or as words in a narrative—and further information integration does, too. Fast evaluation of negative events may therefore initiate additional information processing, enabling more elaborate moral judgments (Guglielmo 2015).

Evaluative priming: response to a target stimulus is faster if preceded by another stimulus (prime) of the same valence as the target

Trolley dilemma: a train is destined to kill five people; should one switch it onto another track where it kills only one person?

Norm Judgments

A second large class of moral judgment may be termed norm judgments—whether something is permissible, required, forbidden, and so forth. Such judgments were made famous in moral psychology by research into the trolley dilemma and related moral dilemmas (Christensen & Gomila 2012, Greene et al. 2001, Mikhail 2011). In most of these dilemmas, the protagonist contemplates an action that sacrifices one life but saves multiple other lives. In such dilemmas, moral

perceivers are placed into the moment before the protagonist makes the deeply conflictual decision and are asked to make a norm judgment, of which numerous variants exist (Christensen & Gomila 2012, pp. 1257–58): "Is it morally permissible to. . . . ," "Is it appropriate [for you] to . . . ," or "Should [he]. . .?" Some studies have asked for first-person predictions (such as "Would you perform the described action?"), effectively measuring imagined moral decision making rather than moral judgment. Decision-making questions, however, do not seem to trigger the same processing (Schaich Borg et al. 2006) and do not lead to the same patterns of results (Gold et al. 2015, Tassy et al. 2013), supporting the conceptual division between moral judgments and moral decisions, which is also supported in virtual reality experiments (Francis et al. 2016).

Deontic logic: refers to logical systems that formalize reasoning about obligations, prohibitions, and permissions

Most of the norm judgment probes (e.g., permissible, forbidden) refer to concepts investigated in a long tradition of work on deontic logic (McNamara 2006) but rarely examined for their distinct psychological characteristics (but see Janoff-Bulman et al. 2009). The major categories are permissions (*acceptable*, *permissible*), prescriptions (*appropriate*, *should*), and prohibitions (*forbidden*). In deontic logic, these categories are strictly related—for example, *A is permissible* $\leftrightarrow$ *A is not forbidden*. However, a well-studied wording effect in survey research shows that this equivalence does not hold in ordinary people's judgments (Holleman 1999). A new avenue of research would be to carefully examine people's psychological interpretations of the various types of norm judgments.

The major point here is that norm judgments are rather different from moral evaluations. Norm judgments invoke the standards against which evaluations are measured and thus set the context for any judgments that are to be called moral (Nichols & Mallon 2006). Moreover, people will explain or defend an evaluation by referring to norms, and sometimes to more abstract concepts such as values and virtues.

Whereas badness has a continuous range, most norm judgments appear to be categorical (something is forbidden or not, permitted or not). A linguistic analysis using the one-billion-word Corpus of Contemporary American English (COCA; Davies 2008) allows us to compare occurrences of *bad* and *permissible* in phrases that indicate continuous degrees, such as "how bad/permissible . . . ," "pretty bad/permissible," or "worse/more permissible." Whereas *bad* occurs 216 times more often in continuous than categorical use, for *permissible* and other norm judgments this ratio is between 0.2 and 1.5 (see **Supplemental Appendix 2** for details).

Supplemental Material >

In addition to this tendency toward categorical use, norm judgments seem to differ from evaluations in two other respects. First, whereas evaluations can be about any kind of event, norm judgments typically take intentional actions as their objects. In fact, as instructions to (not) act a certain way, norm judgments presuppose that the agent can intentionally initiate (or avoid) said action. Second, norm judgments are instructions that guide action, so they are often invoked before the action is performed. In line with these two characteristics, almost all moral dilemma studies have probed norm judgments before the protagonist's actual decision. Their future-directed focus on intentional action makes norm judgments ideal vehicles for social acts of warning, advising, teaching, and persuading.

Wrongness Judgments

May (2018, p. 52) calls the phrase *That's just wrong* the "paradigm moral judgment." Wrongness judgments have certainly been heavily used in moral psychology. One reason for such prominence is that Haidt (2001), in his seminal conceptualization of moral judgments, characterized wrongness judgments as intuitions. His proposal was that, when encountering a moral violation, "many people say something like, 'I don't know, I can't explain it, I just know it's wrong'" (Haidt 2001, p. 814). Subsequently, almost all studies on violations of purity norms relied on judgments of (*morally*)

wrong to examine how people respond to such violations (for reviews, see Giner-Sorolla et al. 2018, Landy & Goodwin 2015).

False alarm: in a task of discriminating a signal from noise, falsely designating a noise stimulus to be signal

Miss: in a task of discriminating a signal from noise, failing to detect a signal, believing it to be noise

To understand what makes wrongness judgments the measure of choice for so much of moral psychology research, we need to identify the features that make these judgments distinct. First, judgments of moral wrongness specifically flag intentional violations (Cushman 2015a, Malle et al. 2014, Patil et al. 2017). In a randomly drawn sample of 100 uses of the phrase "morally wrong" in COCA, a single one referred to an arguably unintentional behavior (see **Supplemental Appendix 3** for details). In line with this pattern, the subset of moral psychology studies that most frequently use moral wrongness measures—studies on impurity and disgust—almost exclusively contain intentional violations.

While many events can be "bad," such as breaking one's leg or telling an unfunny joke, only those that violate a moral norm are morally wrong (Cameron et al. 2017), such as breaking someone else's leg or telling a racist joke. Combining this with the intentionality feature, wrongness judgments typically convey that the perceiver negatively evaluates an intentional action that violated a moral norm (Malle et al. 2014, Patil et al. 2017; see **Supplemental Appendix 4** for lay definitions of wrongness that support this conception). When the act's intentionality is easily detectable and the connection to a moral norm is clear, wrongness judgments can arise quickly and implicitly. Cameron et al. (2017) presented participants with words denoting morally wrong actions or neutral actions and asked them to decide whether each word "represents an action that is morally wrong" or not. Almost 80% of people were able to make wrongness judgments in under 500 ms on average. At this speed, however, they showed false-alarm rates of 32% and misses of 30%. These errors reduced to 10% and 12%, respectively, when the response window was extended to 800 ms, but people were still able to make wrongness judgments in 555 ms on average (see **Supplemental Appendix 5** for details).

Supplemental Material >

Is there empirical evidence for a dissociation between wrongness judgments and evaluations? Linke (2012) asked people to make judgments of a hypothetical norm violation (theft of clothing valued at $1,000) committed by one's mother, a classmate, or a foreigner. Among several dependent variables, the researcher asked how bad the behavior was and how morally wrong it was. Only badness judgments were sensitive to variations in the perpetrators' social closeness (e.g., mother versus stranger), whereas wrongness judgments were constant across closeness. This finding may indicate that wrongness judgments capture action types (largely independent of who performs the action), whereas badness judgments respond to the specifics of the given agent, behavior, and outcome; as such, badness judgments may combine facts and feelings, including more positive feelings toward close others.

Even though wrongness judgments may be less sensitive to who performed a given action, they are sensitive to why the agent did it—probing the offender's mental states (Cushman 2015a). These mental states help determine the seriousness of the transgression (Young et al. 2010), but more importantly they tell us the agent's reasons for transgressing. Some reasons might actually justify the transgression (Riordan et al. 1983): It is morally wrong if a father hurts his daughter, but not so when he is hurting her because he is plucking sea urchin spines out of her foot to prevent an infection. It is morally wrong to injure or kill another person, but not if it is in genuine self-defense. Consideration of specific and potentially justifying reasons for acting may differentiate wrongness judgments from most norm judgments. When a norm is declared (e.g., "Attendance is required"), the agent's reasons are typically not part of the action description. In fact, some actions may be identified as generally forbidden but not morally wrong, if, in a particular case, they are performed for justified reasons (Nichols & Mallon 2006).

Though systematic studies contrasting the elements of norm judgments and wrongness judgments do not exist, there are initial indications for a dissociation between the two. Malle et al.

(2015) presented participants a "switch" trolley dilemma and asked some of them whether the act of switching (saving many but with the unintended consequence of letting one person die) was permissible or not; 65% affirmed that it was. Other participants were told the protagonist's decision (either to switch or to withhold action) and asked to indicate whether the decision was morally wrong. If *morally wrong* were the complement of *permissible* (i.e., impermissible), we would expect approximately 35% of participants to consider switching morally wrong, but 49% did (and only 15% considered not switching morally wrong). Furthermore, Voiklis et al. (2016), using the same data set, content-coded participants' answers to the question "Why does it seem [permissible | morally wrong] (or not) to you?" The patterns of people's explanations were significantly different across judgments: Permissibility judgments were explained predominantly by good or bad consequences (which may ground norms), whereas wrongness judgments were explained less by consequences and more by referring directly to norms and mental states.

Moral luck: when actors A and B perform exactly the same action but A's action causes a morally negative outcome while B's action does not

Other studies did not provide clear evidence for separability. O'Hara et al. (2010) found highly similar ratings for whether a variety of violations were *wrong*, *inappropriate*, *forbidden*, or *blameworthy*. Similarly, Barbosa & Jiménez-Leal (2017) found no mean differences among ratings of *wrong*, *blame*, *impermissible*, *unacceptable*, and *should*. And Kneer & Machery (2019) found that ratings of *permissible*, *wrong*, and *blameworthy* all differentiated similarly (and modestly) between cases of moral luck, though the effect size for permissibility was overall smaller than the effect size for blame and wrongness judgments.

These results cast some doubt on the separability of permissibility and wrongness judgments, but specific features in these studies may have inhibited such separability. First, judgments were made about actions that the person had already performed, which occurs more frequently for uses of *wrong* in English than for *permissible* (see **Supplemental Appendix 6**). Thus, the consistent backward-looking formulations may have encouraged people to interpret the permissibility probes just like the wrongness probes. Second, participants were asked to use rating scales for all judgments, even though *wrongness* and especially *permissibility* are often treated as categorical judgments (see **Supplemental Appendix 2**). Better than speculation, of course, would be a comprehensive comparison between these two types of judgments across experimentally manipulated natural linguistic contexts (preaction versus postaction), question format (dichotomous versus ratings), and various information inputs.

Supplemental Material >

Finally, what are the social functions of moral wrongness judgments? Whereas norm judgments will often be used to announce a norm and to warn, persuade, or teach another person, the present analysis suggests that wrongness judgments announce the detection of a norm violation, declare the violation to be intentional, and may presuppose the absence of a justification. Research to examine these hypothesized functions is clearly needed.

Blame Judgments

Moral wrongness judgments merge evaluations and norm judgments of intentional actions. Blame judgments[1] build on all three processes. An initial blame value is hypothesized to be formed from evaluations and wrongness judgments in light of the seriousness of the violated norm (Alicke 2000, Malle et al. 2014). But blame judgments provide significant extensions: One is to fully incorporate the notion of justification; a second is to handle unintentional norm violations. Blame achieves

[1]I set aside a simpler use of *blame* that refers to "who is to blame?" when multiple candidate agents are considered as the cause of a violation. I also set aside nonagentic uses of *blame*, such as when "four lives were blamed on the hurricane."

these extensions by processing multiple sources of information, including the agent's causal contributions to the event, reasons and their potential justification, and counterfactuals about what the agent could and should have done differently (Cushman 2008, Laurent et al. 2016, Malle et al. 2014). Being sensitive to all this information enables blame to be a graded moral judgment that can express fine-grained moral criticism of the norm violator. A closer look at these extensions and their basis in information processing is warranted.

Counterfactuals: thoughts about events or actions that did not occur but could have occurred or could have been performed

Justifications. When people assess a norm-violating behavior, they determine the behavior's intentionality by considering whether the person wanted to bring about the outcome and had the requisite beliefs to do so. But people also try to infer exactly what the person wanted and believed. These motives or reasons allow people to understand why the person acted and to evaluate how morally justified the action was. When unjustified, reasons can amplify negative moral judgments; when justified, they can mitigate such judgments (Monroe & Malle 2019, Young et al. 2010).

Self-justifications for one's actions are sometimes distorted and can facilitate unethical behavior (Shalvi et al. 2015). But if an agent's justifications are to mitigate other people's blame, the justifications must be acceptable to the community, which means they must uphold important community and legal norms or at least have credible positive consequences for others.

Is the power of justifications limited to blame judgments? Wrongness judgments appear to be sensitive to justifications, but no direct evidence is currently available. Norm judgments, by contrast, rarely take justifications into account, except perhaps for established permission norms such as self-defense. Recent studies suggest that blame and norm judgments indeed dissociate. Whereas blame judgments tracked the distinct justifications people granted different agents (human versus machine agents facing a moral dilemma), norm judgments (what the agent should do) did not (Malle et al. 2019, Scheutz & Malle 2021).

Unintentional violations. In the classic social psychology of morality, responsibility and blame were treated as the culmination of moral judgments and applied to both intentional and unintentional violations (Heider 1958, Shaver 1985, Weiner 1995). In the past two decades, the literature has focused heavily on intentional violations of purity and harm and on decisions in moral dilemmas. Even so, studies on moral (un)luck and on the joint consideration of intentions and outcomes have provided clear evidence that people form moral judgments of unintentional behavior and do so in differentiated ways (Martin & Cushman 2016, Patil et al. 2017, Young et al. 2007).

Several experiments have independently manipulated an agent's intention (neutral versus negative) and the resulting outcome (neutral versus negative), leading to four event types that people systematically order in their judgments of wrongness and blame:

> *no violation* [= neutral intention & neutral outcome] < *accident* [= neutral intention & negative outcome] < *attempted violation* [= negative intention & neutral outcome] < *intentional violation* [= negative intention & negative outcome].

Importantly, in studies where different moral judgments were compared, patterns of blame were more responsive to outcome variations than were wrongness and permissibility judgments (Cushman 2008, Patil et al. 2017). What causes this difference in outcome responsiveness? One possibility is that the objects of judgment are distinct. In most studies, wrongness and permissibility probes target the agent's behavior ("How wrong was [agent]'s behavior?" or "[Agent]'s behavior was... Permissible to Forbidden"), whereas blame probes target the agent ("How much blame does [agent] deserve?"). These different formulations are not a linguistic confound. If wrongness and permissibility judgments normally take intentions and actions as their objects, then the probe

must focus on those objects. Blame is broader, inviting consideration of the agent's mind, the agent's action, and its outcomes. It would be natural to ask, "How much blame does the person deserve for what happened?" But it would be less natural to ask, "How wrong was what happened?" In fact, the latter phrase occurred 0 times in COCA and 0 times in the first 50 entries of a Google search for "How wrong was what. . . . "

Anchoring effect: the biasing influence of an initial representation on subsequent judgments (e.g., initial number representation on subsequent number estimates)

A second, related possibility is that the different judgments have different social functions. Whereas people use norm judgments to declare and affirm a community's moral standards and use wrongness judgments to mark their violations, people use blame to regulate one another's behavior (Przepiorka & Berger 2016, Voiklis & Malle 2018). Such regulation must be responsive not only to intentional actions and their underlying motives but also to unintentional outcomes, since the perpetrator, and other community members, must be encouraged to prevent such unintentional violations in the future.

To achieve the regulation of unintentional violations in a fair manner, people take preventability information into account—whether the agent could have and should have prevented the outcome (Catellani et al. 2004, Malle et al. 2014). The *could have* aspect is people's assessment that the agent had the capacity to prevent the outcome, and it modulates moral judgments (Martin & Cushman 2016, Monroe & Malle 2019). This capacity can be cognitive (could the agent have foreseen the negative outcome?) or physical (could the agent have physically altered the outcome?). The *should have* aspect is the assessment that the agent had an obligation to prevent the outcome, and it too modulates judgments (Kneer & Machery 2019). If both of these counterfactuals are affirmed, then blame, even for accidents, can be substantial. In this way, blame regulates unintentional behavior by criticizing what agents did or failed to do and by motivating them to do better next time.

Blame as sophisticated information processing. Of all moral judgments, blame appears to be the most flexible, complex, and sophisticated. It is most flexible because it can be applied to intentional and unintentional behaviors, actions, mental states, and outcomes. Blame is most complex because, as mentioned above, people integrate morally relevant information from multiple sources: features of the norm-violating event (e.g., degree of harm), the agent's causal involvement, intentionality, the agent's reasons for acting (if the violation is deemed intentional), and counterfactual preventability (if the violation is deemed unintentional). One example of both flexible and complex processing is that people differentiate between agents who think about performing a norm-violating action, want to perform it, or intend to perform it; their blame judgments linearly increase across these three levels, over and above general evaluation ratings (Guglielmo & Malle 2019).

The proposal that blame is the most sophisticated moral judgment may be surprising, given its somewhat tainted reputation. In the "blame game," people accuse others of wrongdoing while deflecting or denying their own wrongdoing. Furthermore, consistently unwarranted blaming has been considered a sign of defective relationships (Fincham et al. 1987), and blaming can become an expression of hate and destruction (Furlong & Young 1996). But blaming badly in this way is itself a norm violation—an unjust accusation, a baseless condemnation. People recognize and spurn such acts of deflection, denial, and hate precisely because they are deficient moral judgments, failing to consider the information that make blame judgments complex: Was the agent causally involved? Did he act intentionally? Could she have prevented the outcome?

When engaging in sophisticated information processing, people usually seek the morally relevant information for blame in an orderly way, from causality to intentionality to either reasons or preventability (Guglielmo & Malle 2017, Mikula 2003). People update their blame judgments systematically as soon as information elements change (Monroe & Malle 2017), and they show no discernible anchoring effect (Monroe & Malle 2019)—an effect that is so often found in other

human judgments (Furnham & Boo 2011). Blame judgments are also surprisingly fast (considering their complexity). It takes people only 2.5 s to make updated blame judgments after they receive relevant new information (Monroe & Malle 2017), and across numerous recent studies we have found that people can make blame judgments within approximately 1,600 ms.

Outcome effect: the influence of the severity of a behavior's outcome (even if unintended) on people's moral judgments of that behavior

Alicke (2000) proposed a more pessimistic view of people's blame judgments. In this model, initial spontaneous evaluations of the norm-violating event influence causal and mental information processing, which then guides blame judgments. These evaluations can also directly affect blame, and information processing is conducted afterward so as to confirm the initial evaluations. The evidence for this model has typically been indirect, showing that experimental manipulations of outcome or character information (both assumed to trigger spontaneous evaluations) affect causal and mental processing and/or blame. Accumulated and recent evidence suggests that outcome effects are weak (Robbennolt 2000) and rely on substantial inferential activity (Kneer & Machery 2019). Character information, too, triggers a range of inferences (Nadler & McDonnell 2012, Siegel et al. 2017). If those inferences are not painstakingly measured, experimental results are often consistent with multiple competing models (Royzman & Hagan 2017). Mediation analyses sometimes suggest a direct effect of outcome or character manipulations on blame and show blame predicting causal inferences (Alicke et al. 2011), but the patterns are not decisive in favor of the overall model (Guglielmo 2015). The most direct evidence would come from subtle manipulations or actual measurements of early spontaneous evaluations (as distinct from blame) and careful tracking of the time course and accuracy of various ensuing causal and mental inferences, as well as of blame itself.

Even if we assume, optimistically, that people are able to engage in careful processing en route to blame judgments, a complex process that considers numerous pieces of information is susceptible to a variety of biases, including motivated cognition (Ditto et al. 2009). For example, people take life history and biology into account when blaming a person for present-day violations, but life history is ambiguous and therefore leaves room for motive-serving interpretations (Gill & Ungson 2018). Similarly, people take character information into account when making blame judgments, and while there is debate over whether this amounts to a bias (Nadler & McDonnell 2012), it is clear that the information itself—the personality ascription inferred from behavior or other sources—can be unreliable and distorted. Finally, when people take counterfactual preventability information into account, they must construct those counterfactuals, and there is obviously considerable leeway in inferring what someone could have done or known.

Nonetheless, to serve their critical role in social regulation, blame judgments must be communicated to the transgressor or the community. At that point, the judgments are open to others' scrutiny—to correct false information, identify stereotypical assumptions, or point to flawed conclusions. Blaming is costly, for both the blamer and the blamee, and these costs will heighten the level of scrutiny and the demand for warrant. Arguably, the best warrant for blame judgments lies in presenting the kind of information that is normally processed to form the blame judgment: causality, intentionality, reasons, preventability, and the evidence that supports those inferences. Preparing such warrant and offering verifiable explanations of one's judgments may sharpen information processing (Lerner & Tetlock 1999), so the social demand to offer warrant may be partially responsible for the sophisticated information processing underlying many blame judgments (Voiklis & Malle 2018).

Features of Four Moral Judgments in Summary

We have seen that moral judgments can be categorized into four distinct yet systematically related classes. **Table 1** summarizes the features of these classes, with some features grounded in

Table 1 Features of four major classes of moral judgments in response to a norm violation

Class of moral judgment	Typical object	Primary information input[a]	Estimated speed	Predominant social functions
Evaluation	Anything	Behavior, outcome	300–600 ms	Tracking of violations, initiation of information search
Norm judgment	Intentional action (not yet performed)	Action	Unknown	Teaching, persuading, affirming of norms
Wrongness judgment	Intentional action performed	Action, mental states	800 ms	Violation declaration, norm reminder
Blame judgment	Person's intentional action or unintentional behavior plus outcome	Outcome, causality, intentionality, mental states, preventability	1,600 ms	If public, moral criticism (second or third person), regulation of future behavior

[a]Processing for all judgments presupposes at least implicit comparison to a norm system.

extant evidence and others more speculative. One might consider the four classes as standing in a hierarchical relationship, such that the more complex ones build on the simpler, faster ones. However, very often the information processing flow will respond to a norm violation, in which case norm judgments have already been implicitly made when the other judgments are formed. When moral judgments are expressed in social settings, their functions seem to differ as well (though research in this area is scant). Norm judgments serve to persuade others to (not) take certain actions ("That's not allowed!"), declare applicable norms ("We don't approve of this here"), and teach others ("The appropriate thing to do is. . ."). Wrongness serves mainly to mark a norm-violating intentional action and perhaps to reject insufficient justifications ("It's still wrong!"), while blame criticizes, influences reputation, and regulates relationships.

Judgments of moral character and punishment have been omitted from the discussion so far. They may not be fundamentally different from moral judgments, but they are different enough to warrant a separate discussion, to which I turn next.

ALMOST MORAL JUDGMENTS

Moral Character

Character judgments are not assessments of morally significant events (e.g., norm violations); rather, they are inferences from such events. In contrast to the four moral judgments discussed above, moral character assessments are a form of personality judgment, central to person perception (Goodwin 2015). Character assessments therefore come with all the standard features of dispositional inference, including relative slowness (Malle & Holbrook 2012) and risk of overattribution bias under limited evidence (Ross et al. 1977). Though we do not know how frequent spontaneous moral character inferences are, we do know that, when prompted, people's inferences are sensitive to the agent's reasons for acting (Martin & Cushman 2016, Reeder et al. 2002). People's character inferences seem to be cautious: They are often close to the scale midpoint (Tannenbaum et al. 2011) and significantly below that midpoint for unintentional violations (Martin & Cushman 2016). Moreover, character inferences often dissociate from moral judgments. For example, stealing a dead chicken was seen as more immoral than having sex with a dead chicken, but the latter was judged as more indicative of the person's bad or abnormal character (Uhlmann & Zhu 2013).

Dispositional inference: inference of a stable characteristic (e.g., attitude, personality trait) from a person's behavior

Some studies suggest that moral character inferences can have an impact on blame, potentially mediated by emotions (Nadler & McDonnell 2012). The relationships among these processes

may be more complex, however, because the experimental information participants receive about a person's character routinely invites a variety of other inferences (e.g., about mental states or obligations), which then might mediate the effect of perceived character both on emotion and on blame (Malle et al. 2014, Royzman & Hagan 2017). Indeed, people often integrate character information with increased mental state inference into moral judgments that conform to Bayesian rationality (Kim et al. 2020).

Bayesian rationality: refers to changing one's beliefs in light of new evidence in accordance with principles of probability theory

In general, the significance of moral character inferences will be predominant in encounters with strangers and acquaintances, for whom people lack dispositional information; novel character inferences can then guide people to interact with those persons in the future (Martin & Cushman 2015). In everyday interactions with familiar social partners, character inferences will be infrequent (though character knowledge may still be influential), whereas the major moral judgments of wrongness and blame are impactful for both strangers and close others.

Punishment

In the psychology literature, we are once more confronted with a variety of ways in which punishment is conceptualized or measured:

1. answering "How much punishment does [agent] deserve?" (Kneer & Machery 2019) or indicating to what extent one wants to personally punish the offender (Hofmann et al. 2018);
2. recommending formal sanctions, such as jail sentences and fines (Laurent et al. 2014) or suspensions and demotions (Bauman et al. 2016);
3. assigning "points of disapproval" to another player in an economic game (Dugar 2010);
4. reprimanding a person who littered (Balafoutas et al. 2016) or asking someone to silence their loud music on a train (Przepiorka & Berger 2016);
5. reducing another player's monetary payoff in an economic game (Yamagishi et al. 2009);
6. mixing hot chili powder into another person's drink (Gollwitzer & Bushman 2012) or blasting an unpleasant sound into another person's headphones (Pedersen et al. 2018);
7. physical abuse, denunciation, or abandonment in close relationships (Fitness 2001);
8. publicly shaming others on the internet (Klonick 2015); and
9. countless forms of formal penalties, discipline, and institutional punishment.

These examples fall into three groups. The first two exemplify punishment recommendations, the dominant way to measure punishment in the moral psychology literature. Some have a well-defined currency (example 2), but for others, participants are free to imagine various forms of punishment, including fines, jail time, shaming, or social exclusion. Given the wide range of possibilities, studies using such measures are difficult to compare. Examples 3 and 4, sometimes labeled social punishment, are hardly punishing but rather communicate moral criticism (akin to social acts of blaming) and open the door to a process of correction and reconciliation. Examples 5 and 6 are mild laboratory analogs of destructive punishment, whereas examples 7 to 9 exemplify the increasingly brutal damage humans sometimes inflict on one another (Farrington 1996), with little room for reconciliation or negotiation.

When we contrast the above measures of punishment with those of moral judgment listed in the introductory section, it becomes clear that punishment cannot simply be considered another moral judgment. Even so, punishment recommendations and blame are often treated as parallel (Ames & Fiske 2013, Cushman 2008) or are averaged as correlated measures (e.g., Rothschild et al. 2012). There is no doubt that both social blame and punishment have regulatory and pedagogical functions (Cushman 2015b, Malle et al. 2014). Also, like blame, punishment is flexible in taking into account outcomes, actions, and mental states (Kneer & Machery 2019). However,

classic attribution research and recent studies (Bauman et al. 2016, Buckholtz et al. 2015, Cramer et al. 2014) have shown that punishment recommendations are mediated, not just accompanied, by blame judgments and that they are more influenced by prior wrongness judgments than the other way around (Leloup et al. 2018). Thus, punishment, as a recommended or actual moral sanction, is grounded in moral judgments.

Numerous experimental economics studies have examined punishment behavior in monetary games, allowing participants to punish other players by reducing their payoffs. For example, when the responder in an Ultimatum Game rejects the proposer's offer to selfishly split the money (e.g., 8:2 ratio), neither receives any money, and this rejection is interpreted as costly punishment. It is punishment because the proposer loses out and costly because the responder loses out.

Approximately half of "unfair" offers (8:2 ratio or worse) are rejected, and research has examined numerous predictors of these rejection or punishment rates. Here I highlight one pattern of findings: People will punish the proposer if that is the only option available, but when alternatives are given, people prefer them over punishment (FeldmanHall et al. 2018). Furthermore, granted an opportunity to communicate their disapproval (e.g., by sending a note), people reduce their accompanying monetary punishment (Xiao & Houser 2005), and people are generally less harsh and punitive when communicated disapproval is the designated response option, compared with when monetary punishment is the only option (Leibbrandt & López-Pérez 2014). Thus, it appears that people care about expressing their moral disapproval one way or another. Punishment is one way, but people often prefer other ways.

In the Ultimatum Game, the victim of the injustice performs the punishment, sometimes called second-party punishment. Researchers have also investigated the conditions of third-party punishment, imposed by unaffected witnesses. Some scholars have argued that there is a strong incentive in cooperative communities for a witness to punish transgressors (Fehr & Gächter 2000, Gintis 2000). Despite initial support, this hypothesis has been criticized on methodological (Pedersen et al. 2013), empirical (Kiyonari & Barclay 2008), and theoretical grounds (Baumard 2010, Krasnow et al. 2012). Results show that people (or 15–60% of them, depending on game structure) choose to punish primarily when no other options are available. When given a choice between punishing the perpetrator and compensating the victim, they prefer the latter (Chavez & Bicchieri 2013). When given no chance to punish or compensate but to warn others who will interact with the transgressor, people select that path (Feinberg et al. 2012). And when offered an opportunity to later punish an unfair player, approximately 40% of people declare an intention to do so, but far fewer go through with the intended punishment if they can wriggle their way out of it (Kriss et al. 2016). Perhaps this reluctance to actually punish is for the better, because Dugar (2010) showed that, in a group coordination game, groups that had monetary punishment opportunities (and enacted them) showed worse coordination than groups that had an opportunity to merely express criticism. Punishment may not be the healthiest path to cooperation.

Outside the boundaries of economics games, Pedersen et al. (2018) tested whether people would punish (with an unpleasant sound blast) another person who insulted either them or a friend or stranger. Across multiple studies, victims of the insult punished the insulter, and friends of the victim did too, albeit to a lesser extent; strangers, however, did not engage in punishment. When we look outside the laboratory, punishment is generally rare in the ethnographic record and believed to be rarer yet in presettlement human communities (Guala 2012). Similarly, in today's public sphere, even norm enforcement of the milder kind is relatively infrequent. Fewer than 15% of people ask a stranger to pick up the litter they dropped (Balafoutas & Nikiforakis 2012). In the safety of bystanders, enforcement rates can increase to 50%, such as for enforcing the silence rule in train cars (Przepiorka & Berger 2016). Also, people in principle endorse intervening with a restaurant customer who mistreats waitstaff, but in reality most of them express their moral

Ultimatum Game: a proposer offers a responder a cut of some amount of money, (e.g., $4/$10); both receive money only if responder accepts the offer

Third-party punishment: when a witness of a norm violation (e.g., A hurts B) is unaffected by the violation but punishes the norm violator

judgment to the victim rather than directly confront the perpetrator (Hershcovis & Bhatnagar 2017).

One conclusion we might draw is that monetary games, where costs are low and precisely calculable, may not be an adequate model for social punishment in real life. There, punishing a stranger comes with high cost and high uncertainty and is therefore quite rare. Still rare, but naturally preferred, is a more civil form of social regulation, namely social blame, expressed to the offender directly or to other community members. Such regulation sends the important signal of one's commitment to moral norms and gives the offender an immediate opportunity to repair the breach, reducing the costs to both blamer and blamee compared with those of literal punishment.

APPLYING THE FRAMEWORK

Measurement

Given the diversity of moral judgments, some researchers are concerned that "we do not even know what type of moral judgement subjects are making" (Kahane & Shackel 2010, p. 567). The present framework begins to address this concern by identifying classes of moral judgment and their differential features (**Table 1**). But the next step is to develop measurement standards that ensure the judgments people make are in fact the judgments we hope to elicit.

One challenge is that, currently, the types of violations researchers present covary with the kinds of judgments they probe. Moral dilemmas are almost always paired with norm judgments (*permissible*, *acceptable*), purity violations with moral wrongness, and harm and injustice (and most unintentional violations) most often with blame and punishment. Too rarely do researchers measure multiple judgments, and when they do, they often average them. Future research needs to complete the missing cells, measure blame for impurity violations and dilemmas (e.g., Malle et al. 2015), and compare multiple judgments for the same violations (Kneer & Machery 2019, Leloup et al. 2018, Patil et al. 2017).

But we need to be cautious. Asking multiple questions in the same study does not guarantee that people will make multiple distinct judgments. Nor does asking a face-valid question guarantee that participants will provide the specific judgment the researcher had in mind. Below I present samples of interpretational challenges in past findings in which the intended and actual judgments may have come apart.

First, some evaluation probes can be problematic. Evaluation is such a general process that probing valence without indication of the intended moral meaning may not actually measure a moral judgment. For example, measures of (dis)approval (e.g., Van Dillen et al. 2012) may pick up a good deal of general discomfort or weirdness (Gray & Keeney 2015) rather than a genuine moral judgment.

Second, questions of interpretation arise when norm judgments are used for unintentional behaviors (e.g., Kneer & Machery 2019, Young et al. 2012). In one study, psychopaths and control participants judged a large number of events on a scale from "permissible" to "forbidden" (Young et al. 2012). However, a quarter of the events contained bad accidents, and these were the events on which the two groups differed: Psychopaths rated the accidents as more permissible. It is unclear what it means that an accident is (im)permissible, so people may have evaluated not the accident itself but rather the person's action that preceded the negative event. Control participants may have "transformed" (Royzman & Hagan 2017) the permissibility scale into a blame judgment, which can assess unintentional outcomes. Psychopaths did not seem to perform such a transformation, but they might make similar judgments when directly asked for blame judgments.

Third, wrongness judgments, too, can be ambiguous. Margoni et al. (2019) found that older adults gave harsher wrongness judgments than younger adults for stimuli introduced as "actions,"

but half of these stimuli were unintentional outcomes. If wrongness is not well suited for judging unintentional events, participants may have transformed the wrongness probe into blame judgments, and the age difference may have arisen from differential preventability considerations for blame. This is actually consistent with Margoni et al.'s hypothesis and the results of their study 2, in which negligence inferences (that the agent could and should have prevented the outcome but did not) were the primary driver of the age difference. The suspicion that participants translated wrongness into blame judgments is also consistent with the fact that their wrongness judgments showed the same pattern as their punishment judgments, which are typically parallel to blame, whereas wrongness normally diverges from both punishment and blame (Cushman 2008).

It would be false to conclude that these examples reflect the unreliability of people's moral judgments. Instead, they should encourage us to consistently and clearly formulate questions that elicit just the judgments we theorize about. Study participants are adaptive and eager to interpret our questions within the communicative context we place them in. But if the question or response options do not fit the task or stimuli, people will reinterpret the question (Laurent et al. 2019) or select the least objectionable option (Guglielmo & Malle 2010). Moreover, to make sense of the presented narratives, participants actively infer the agents' mental states that are neither mentioned nor intended to be inferred (Royzman & Hagan 2017), just like they do for any narrative (Graesser et al. 1994). This is the reality of complex human social and moral cognition, and our methods must fully live up to this complexity (**Supplemental Appendix 7** offers some recommendations for sharpening our current methods).

Supplemental Material >

Moral Intuitions

If there are multiple classes of moral judgments, then what are moral intuitions? For some scholars, all moral judgments originate in intuitions: "Moral judgments appear in consciousness automatically and effortlessly as the result of moral intuitions" (Haidt 2001, p. 818); "we must build our moral judgments and arguments from the raw materials of our moral intuitions" (Clark & Winegard 2019, p. 13). Mikhail (2011, p. 113) challenged such a conception of intuition, for it does not specify how a morally significant event causes an intuition. Now that we have identified an array of judgments from basic evaluations to complex blame, we can consider the following candidate specifications of moral intuitions. Either such intuitions precede all moral judgments (i.e., they are prior to evaluations), or they are one of these judgments.

Considering the first possibility, it is difficult to see what morally relevant process could precede 300-ms fast evaluations. Instead, more plausible is the second possibility—that intuitions are themselves evaluations. Haidt (2001, p. 818) characterized evaluations as examples of moral intuition: "One sees or hears about a social event and one instantly feels approval or disapproval." Similarly, Clark & Winegard (2019, p. 14) cite "breeding dogs is bad" as a moral intuition. However, because evaluations are often not moral, a constraint would have to be added to the notion of moral intuitions as evaluations: Relevant norms must be invoked to morally evaluate an event (Cameron et al. 2017), so moral intuitions must, at a minimum, evaluate something as morally bad in light of some relevant norm (otherwise, the intuition could simply be a dislike). Such moral evaluations would be likely to arise effortlessly, automatically, and fast—properties typically ascribed to intuitions, moral or otherwise (Haidt 2001, Kahneman 2013).

We may then ask what information moral intuitions as evaluations could or could not deliver. At this stage of processing, they might code for simple causality (De Freitas & Alvarez 2018) and for intentionality of visibly performed behaviors (Decety & Cacioppo 2012), thus conforming to Mikhail's (2011) suggestion that moral intuitions must provide some structural representation of a behavior. But qua evaluations they do not incorporate the agent's specific reasons and do

not provide counterfactual assessments of preventability; they provide a starting point for such higher-level judgments, which require additional information processing.

A second possibility is that moral intuitions are norm judgments—such that one simply knows what one should or should not do. Some prohibitions may be intuitive in this sense, such as in the case of action aversion (Cushman et al. 2012), where certain actions are felt to be aversive and impermissible because of a strong reinforcement history (Crockett 2013). However, it is more difficult to see prescriptions as intuitive; instead, they may often counteract intuition in that they replace an intuitively preferred behavior with a socially accepted one (e.g., standing in line instead of walking up to the counter; keeping a promise that one would rather break).

The third possibility is that moral intuitions are wrongness judgments. This covers the most frequent use of the term, as most studies on impurity violations probed wrongness judgments and often assumed the intuitive quality of those judgments. However, if the earlier analysis of moral wrongness judgments is correct, then we must grant such wrongness "intuitions" at least three cognitive processes: activating the relevant norm, assessing that the person acted intentionally, and judging that the action violated the norm. If wrongness judgments also incorporate the agent's reasons, then moral intuitions would, too; and if wrongness judgments do not apply to unintentional violations, moral intuitions would not, either.

Finally, if we designated blame judgments to be intuitions, then the term *intuition* would lose its intended meaning, because blame judgments, as we have seen, are grounded in a great deal of information processing about intentionality, justification, preventability, and so on—just the kind of processing typically excluded from *intuition*.

In summary, if intuitions are to be fast and largely automatic, then they appear to resemble evaluations. Such intuitions do not provide very rich information, and substantially more processing is needed to arrive at further moral judgments. Intuitions would then be the beginning of moral judgments but would not constitute all of moral judgment. If we grant intuitions more information processing (by equating them with norm, wrongness, or blame judgments), then they become increasingly powerful but also figure to be less automatic and arguably to involve considerable reasoning—precisely the construct they were meant to replace (Haidt 2001).

But perhaps it is counterproductive to ask where intuitions are located within the layering of moral judgments. If we instead identify properties that intuitions are expected to have (e.g., speed, inaccessible origin, affective feel, automaticity), then we can empirically determine which class of moral judgment has which intuitive properties. We need not even expect that each class has a fixed profile of these properties; instead, each can show some properties but not others, depending on moral domain, stimulus prototypicality, or social demands. Intuition then becomes a process characterization, applicable to various moral judgments rather than reified to be one of them.

Moral Dumbfounding

The nature of moral intuitions is often tied to the hypothesis that people display "moral dumbfounding" (Haidt et al. 2000), usually defined as the "inability to justify" (Haidt & Björklund 2008, p. 197) one's moral judgment or the tendency to maintain such a judgment in the "absence of supporting reasons" (McHugh et al. 2017, p. 1). In the original procedure to test this hypothesis, an experimenter interviewed participants, exposed them to strongly norm-violating scenarios, probed for people's moral responses, and asked them to explain their responses (Haidt et al. 2000). The researchers coded the interviews for several variables, such as participants' hesitations, so-called unsupported declarations (e.g., "It's just wrong to do that!" or "That's terrible!"; Haidt et al. 2000, p. 9), and their self-declared inability to explain their responses. The two initial studies did not provide data on the actual frequency of such statements, but a recent replication did. The authors

of this study (McHugh et al. 2017) classified two of the above dumbfounding responses: unsupported declarations ("it's just wrong") and explicit admissions of having no reasons. Aggregating across several studies (see **Supplemental Appendix 8**), 32% of participants who called a given scenario wrong provided one or more dumbfounding responses for this scenario.

Supplemental Material >

This rate of dumbfounding is surprisingly low, given that the experimental procedure requires the interviewer "to undermine whatever reason the participant put forth in support of his/her judgment or action" (Haidt et al. 2000, p. 7). Thus, participants' initial explanations of why, say, incest between siblings is wrong were always challenged, and answers that referred to harm were explicitly rejected. Critics of this procedure note that people's concerns about actual or potential harm in fact dominate their justifications, so the researchers' decision to undermine or reject these concerns distorts the results (Gray et al. 2014, Royzman et al. 2015, Stanley et al. 2019). In particular, participants in studies by Royzman et al. (2015) and Stanley et al. (2019) very much believed that there was potential harm in the violation scenarios (including incest), and the degree of these beliefs predicted people's wrongness judgments. Furthermore, directly manipulating the salience of such harm potential increased wrongness judgments (Stanley et al. 2019).

An important question here is what counts as a reason or justification for one's moral judgment. Haidt, McHugh, and colleagues do not accept a norm statement ("because it's incest") as a reason but deem it an unsupported declaration, whereas Stanley et al. (2019) firmly consider it a reason. Statements such as "it's just wrong" are arguably merely restated wrongness judgments; but identifying the actual norm that the behavior violated is a fitting candidate for justifying the judgment. It roots the wrongness judgment in one of its constitutive components: a norm judgment, in addition to the obvious evaluation (see **Table 1**).

Considering all classes of moral judgment, from evaluation to blame, we can explore each judgment's potential justifications. Little justification is available for evaluations, except by pointing to salient positive or negative features of the stimulus. But this is true for all evaluations, not just moral ones. Liking a painting or disliking beets can hardly be justified, by laypersons or by scientists. Norm judgments, too, are difficult to justify, but sometimes harmful consequences or underlying values will be salient ("it's a sign of respect"; Stohr 2012). It would seem to be an unreasonable standard, however, to expect people to justify norms in terms of philosophical principles like utilitarianism or the doctrine of double effect, or to give a cultural history of how certain norms emerged in their community.

Wrongness judgments, if cast as evaluations of an intentional action that violates a norm, can be justified by reference to intentionality or the violated norm. By contrast, the social demands on justifying blame judgments, and the justifications available for them, go much further (Malle et al. 2014, Voiklis & Malle 2018). If a person cannot point to the causal, intentional, mental, or counterfactual information that grounded their blame judgment, their conversation partner would rightly consider the judgment unjustified and challenge it with specific information arguments—"but she didn't do it intentionally" or "I couldn't possibly have anticipated this." Consistent with this hypothesized readiness to justify one's blame judgments, Bucciarelli et al. (2008, study 3) found that people had no trouble explicating their blame judgments in a think-aloud protocol, and neither did participants in Voiklis et al.'s (2016) study of postjudgment explanations.

Dual-Process Models Reconsidered

The year 2001 marked the publication of two highly influential articles in the new moral psychology literature. One (Haidt 2001) inspired the questions in the previous two sections; the other (Greene et al. 2001) inspires the question in this section: whether moral judgment stems, like many other psychological responses, from the reconciliation of two competing processes—an

automatic emotional one and a controlled cognitive one. Greene and colleagues answered this question in the affirmative and proposed a model in which early and fast emotional processes lead to deontological moral judgments (presumed to be based on inflexible rules) and slower controlled processes lead to utilitarian moral judgments (presumed to be based on considerations of consequences).

Deontological: describes an ethical theory in which actions are inherently right or wrong according to moral rules, independent of their outcomes

Utilitarian: describes an ethical theory in which actions are right or wrong because of the positive or negative consequences they bring about

Greene et al.'s (2001) specific model had an enormous impact on moral psychology research, but it has increasingly been criticized for (*a*) an invalid identification of deontological and utilitarian assessments with, respectively, automatic and controlled processes (Kahane 2012, Rosas & Aguilar-Pardo 2019); (*b*) the reduction of deontological/utilitarian assessments to action/inaction judgments in moral dilemmas (Gawronski et al. 2017); and (*c*) model-disconfirming evidence in neural data (Demaree-Cotton & Kahane 2018, Klein 2011), reaction time data (Gürçay & Baron 2017), and cognitive load data (Bago & De Neys 2019, Sauer 2012). Though the sheer volume and consistency of this recent critique are compelling, most critics left two foundational premises untouched: that moral judgment is a choice between deontology and utilitarianism in the first place and that a dual-process model can be tested on permissibility probes of moral dilemma scenarios. These premises, I suggest, have impeded an evaluation of process considerations in moral judgment.

A general dual-process model of moral judgment would propose that, when a moral judgment is made, two processes are triggered that compute the available information and that can be characterized by the bundle of features associated with the general two-systems view in psychology (Sloman 1996). The negative evidence cited above disconfirms the two-systems predictions, but only for permissibility judgments, and only in moral dilemmas tailored to the deontological/utilitarian dichotomy. The results say nothing about the full breadth of human moral judgments. The framework presented here moves the deontological/utilitarian dichotomy to the side (as it concerns normative commitments, not the psychological processes underlying moral judgments) and clears the way for investigating which classes of moral judgments—from evaluations to blame judgments—are formed by which combinations of processes.

The processes to consider for such an investigation may be grouped within a two-systems frame, but we may want to delay talk of two systems until more comprehensive data about the moral domain become available. Most fruitful would be to ask first to what extent the classes of moral judgments are characterized by the range of core process properties (automaticity, speed, sensitivity to new information, involvement of core affect, etc.) and then assess how these properties cluster.

The key advance would be to examine each of these process properties for each of the classes of moral judgments. We would examine, separately and jointly, evaluations, norm judgments, wrongness judgments, and blame judgments (and perhaps more), using all available methodological tools and a full range of stimuli. Methods would include verbal judgments, reaction times, cognitive load, process dissociation, physiology, brain imaging, emotion induction, and more. Stimuli would range from pictures to videos, from verb phrases to narratives, from observed to experienced violations. Stimulus content would vary violations of specific norms to broader values, illegal and nonillegal, mild to severe, from first-person and observer perspectives. And morally relevant information would be manipulated or its inferences measured, including causality, intentionality, mental states and their justification, and the obligation and capacity to prevent unintentional negative outcomes.

I must leave it to the research community to apply these rich methodological tools to the distinctions among moral judgments reviewed here—and, more broadly, to the fundamental question of what out there in the world gives rise to which morally relevant mental processes, in what order, and with what sensitivity, flexibility, and social consequences. The framework I presented here is not a complete theory but rather draws the outlines of multiple small theories about each of the

judgments and their constituent processes, eventually merging into one broader theory of moral judgment that the next decade may bring forth.

SUMMARY POINTS

1. Prototypical moral judgments are evaluative responses to a norm-violating event accompanied by varied information processing about the event. However, the diversity of what has been called moral judgment suggests that it is not just one phenomenon but many.
2. The four main classes of moral judgments are evaluations, norm judgments, moral wrongness judgments, and blame judgments. Building on these judgments, moral perceivers also make inferences about a person's moral character and dole out punishment.
3. Evaluations are fast, graded assessments of how good or bad an event is, but they take only limited moral information into account. Norm judgments designate a behavior (often future and intentional) as permissible, prescribed, or forbidden. Moral wrongness judgments determine that a norm-violating intentional action was performed that lacked justification.
4. Blame judgments are flexibly applied to both intentional and unintentional violations. They are graded assessments that take numerous pieces of information into account: the violated norm, the agent's causal contribution, and the agent's intentionality and possible justifications (reasons) for the chosen action or, in the case of unintentional violations, counterfactual beliefs about how the person could have and should have prevented an unintentional negative outcome. Because of the costs of blame, the community may scrutinize blame judgments for their accuracy more than any other judgment class.
5. Punishment, not only in its legal but also in its social form, is a moral sanction that builds on blame but is distinct from it. Research suggests that punishment is not as frequent as is sometimes claimed, and people punish primarily when they are the victim of a violation and when the costs are low.
6. This framework helps guide the proper measurement of moral judgments, the nature of moral intuitions, the status of moral dumbfounding, and the prospects of a dual- (or multi-) process model of moral judgments.
7. A successful investigation of the full range of moral judgments will require a commitment to use a variety of methodological tools, a wide array of stimuli, and systematic variations of their content features. Grounded in such research, a comprehensive theory of moral judgments is likely to emerge.

FUTURE ISSUES

1. One unexplored area is the diversity of norm judgments people make: not only what is permissible but also what is expected, obligatory, forbidden, and so on. Little research has addressed the psychological interpretations of these norm judgments (e.g., in contrast to axioms of classic deontic logic).
2. The social functions of the different classes of moral judgment are not well understood, in part because moral judgments are rarely studied in their communicative and

interactive contexts (e.g., teaching someone a norm, expressing one's outrage, negotiating blame).

3. Researchers must overcome the strong habits of studying permissibility judgments for moral dilemmas, wrongness judgments for purity violations, and blame and punishment for most other violations. Different types of moral stimuli must be crossed with the different classes of moral judgments, enabling generalizability as well as insights into the unique use and functions of these different judgments.
4. Innovative measurement approaches are needed to probe multiple classes of moral judgment without allowing the probes to collapse into a single blended judgment. We need measurement tools that keep apart judgments that are semantically similar but psychologically distinct.
5. A comprehensive theory of moral judgment must build on the accumulating knowledge of different classes of moral judgment and integrate them into a theory of dynamic information processing, flowing from detecting and evaluating a moral event all the way (and in loops) to expressing blame, initiating repair, or inflicting punishment. The vague concept of moral intuition may then be replaced by a process account of multiple distinct moral judgments.
6. Another rich set of research questions derives from the hypothesis that the major classes of moral judgment stand in a hierarchical relationship. In addition to questions of processing depth and speed, a possible developmental order would be intriguing to study. Evaluations may develop very early; in light of learned norm representations, wrongness judgments are possible; and once processing of intentions for wrongness is mastered, blame for unintentional violations can be refined.

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

I am deeply grateful to Fiery Cushman, Corey Cusimano, Geoffrey Goodwin, Walter Sinnott-Armstrong, and Liane Young for their thoughtful and challenging comments on a previous version of this article; to Daryl Cameron for providing raw data for analysis; to Vincent Rice for helping me manage some of the literature; and to my collaborators, Steve Guglielmo, Boyoung Kim, Andrew Monroe, Matthias Scheutz, and John Voiklis, for thinking together about matters of moral judgment.

LITERATURE CITED

Alicke MD. 2000. Culpable control and the psychology of blame. *Psychol. Bull.* 126(4):556–74

Alicke MD, Rose D, Bloom D. 2011. Causation, norm violation, and culpable control. *J. Philos.* 108(12):670–96

Ames DL, Fiske ST. 2013. Intentional harms are worse, even when they're not. *Psychol. Sci.* 24(9):1755–62

Bago B, De Neys W. 2019. The intuitive greater good: testing the corrective dual process model of moral cognition. *J. Exp. Psychol. Gen.* 148(10):1782–801

Balafoutas L, Nikiforakis N. 2012. Norm enforcement in the city: a natural field experiment. *Eur. Econ. Rev.* 56(8):1773–85
Balafoutas L, Nikiforakis N, Rockenbach B. 2016. Altruistic punishment does not increase with the severity of norm violations in the field. *Nat. Commun.* 7:13327
Barbosa S, Jiménez-Leal W. 2017. It's not right but it's permitted: wording effects in moral judgement. *Judgm. Decis. Mak.* 12(3):308–13
Bartels DM, Bauman CW, Cushman FA, Pizarro DA, McGraw AP. 2015. Moral judgment and decision making. In *The Wiley Blackwell Handbook of Judgment and Decision Making*, ed. G Keren, G Wu, pp. 478–515. New York: Wiley
Bauman CW, Tost LP, Ong M. 2016. Blame the shepherd not the sheep: Imitating higher-ranking transgressors mitigates punishment for unethical behavior. *Organ. Behav. Hum. Decis. Process.* 137:123–41
Baumard N. 2010. Has punishment played a role in the evolution of cooperation? A critical review. *Mind Soc.* 9:171–92
Bloom P. 2011. Religion, morality, evolution. *Annu. Rev. Psychol.* 63:179–99
Bucciarelli M, Khemlani S, Johnson-Laird PN. 2008. The psychology of moral reasoning. *Judgm. Decis. Mak.* 3(2):121–39
Buckholtz JW, Martin JW, Treadway MT, Jan K, Zald DH, et al. 2015. From blame to punishment: Disrupting prefrontal cortex activity reveals norm enforcement mechanisms. *Neuron* 87(6):1369–80
Cameron CD, Payne BK, Sinnott-Armstrong W, Scheffer JA, Inzlicht M. 2017. Implicit moral evaluations: a multinomial modeling approach. *Cognition* 158:224–41
Cannon PR, Schnall S, White M. 2011. Transgressions and expressions: affective facial muscle activity predicts moral judgments. *Soc. Psychol. Personal. Sci.* 2(3):325–31
Catellani P, Alberici AI, Milesi P. 2004. Counterfactual thinking and stereotypes: the nonconformity effect. *Eur. J. Soc. Psychol.* 34(4):421–36
Chavez AK, Bicchieri C. 2013. Third-party sanctioning and compensation behavior: findings from the Ultimatum Game. *J. Econ. Psychol.* 39:268–77
Cheng JS, Ottati VC, Price ED. 2013. The arousal model of moral condemnation. *J. Exp. Soc. Psychol.* 49(6):1012–18
Christensen JF, Gomila A. 2012. Moral dilemmas in cognitive neuroscience of moral decision-making: a principled review. *Neurosci. Biobehav. Rev.* 36(4):1249–64
Clark CJ, Winegard BM. 2019. Optimism in unconscious, intuitive morality. *Behav. Brain Sci.* 42:e150
Cramer RJ, Clark JWI, Kehn A, Burks AC, Wechsler HJ. 2014. A mock juror investigation of blame attribution in the punishment of hate crime perpetrators. *Int. J. Law Psychiatry* 37(6):551–57
Crockett MJ. 2013. Models of morality. *Trends Cogn. Sci.* 17(8):363–66
Cushman F. 2008. Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108(2):353–80
Cushman F. 2015a. Deconstructing intent to reconstruct morality. *Curr. Opin. Psychol.* 6:97–103
Cushman F. 2015b. Punishment in humans: from intuitions to institutions. *Philos. Compass* 10(2):117–33
Cushman F, Gray K, Gaffey A, Mendes WB. 2012. Simulating murder: the aversion to harmful action. *Emotion* 12(1):2–7
Cusimano C, Thapa S, Malle BF. 2017. Judgment before emotion: People access moral evaluations faster than affective states. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, ed. G Gunzelmann, A Howes, T Tenbrink, EJ Davelaar, pp. 1848–53. Austin, TX: Cogn. Sci. Soc.
Davies M. 2008. *The Corpus of Contemporary American English* (*COCA*): *one billion words*, *1990–2019*. **https://www.english-corpora.org**
De Freitas J, Alvarez GA. 2018. Your visual system provides all the information you need to make moral judgments about generic visual events. *Cognition* 178:133–46
De Houwer J, Thomas S, Baeyens F. 2001. Association learning of likes and dislikes: a review of 25 years of research on human evaluative conditioning. *Psychol. Bull.* 127(6):853–69
Decety J, Cacioppo S. 2012. The speed of morality: a high-density electrical neuroimaging study. *J. Neurophysiol.* 108(11):3068–72
Demaree-Cotton J, Kahane G. 2018. The neuroscience of moral judgment. In *The Routledge Handbook of Moral Epistemology*, ed. A Zimmerman, K Jones, M Timmons, pp. 84–104. New York: Routledge

Ditto PH, Pizarro DA, Tannenbaum D. 2009. Motivated moral reasoning. In *Moral Judgment and Decision Making*, Vol. 50, ed. DM Bartels, CW Bauman, LJ Skitka, DL Medin, pp. 307–38. San Diego, CA: Academic

Dugar S. 2010. Nonmonetary sanctions and rewards in an experimental coordination game. *J. Econ. Behav. Organ.* 73(3):377–86

Farrington K. 1996. *Dark Justice: A History of Punishment and Torture*. New York: Smithmark

Fehr E, Gächter S. 2000. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* 90(4):980–94

Feinberg M, Willer R, Stellar J, Keltner D. 2012. The virtues of gossip: reputational information sharing as prosocial behavior. *J. Personal. Soc. Psychol.* 102(5):1015–30

FeldmanHall O, Otto AR, Phelps EA. 2018. Learning moral values: Another's desire to punish enhances one's own punitive behavior. *J. Exp. Psychol. Gen.* 147(8):1211–24

Fincham FD, Beach S, Nelson G. 1987. Attribution processes in distressed and nondistressed couples. 3. Causal and responsibility attributions for spouse behavior. *Cogn. Ther. Res.* 11(1):71–86

Fitness J. 2001. Betrayal, rejection, revenge and forgiveness: an interpersonal script approach. In *Interpersonal Rejection*, ed. MR Leary, pp. 73–103. New York: Oxford Univ. Press

Francis KB, Howard C, Howard IS, Gummerum M, Ganis G, et al. 2016. Virtual morality: transitioning from moral judgment to moral action? *PLOS ONE* 11(10):e0164374

Furlong M, Young J. 1996. Talking about blame. *Aust. N. Z. J. Fam. Ther.* 17(4):191–200

Furnham A, Boo HC. 2011. A literature review of the anchoring effect. *J. Socio-Econ.* 40(1):35–42

Gailey JA, Falk RF. 2008. Attribution of responsibility as a multidimensional concept. *Sociol. Spectr.* 28(6):659–80

Gawronski B, Armstrong J, Conway P, Friesdorf R, Hütter M. 2017. Consequences, norms, and generalized inaction in moral dilemmas: the CNI model of moral decision-making. *J. Personal. Soc. Psychol.* 113(3):343–76

Gill MJ, Ungson ND. 2018. How much blame does he truly deserve? Historicist narratives engender uncertainty about blameworthiness, facilitating motivated cognition in moral judgment. *J. Exp. Soc. Psychol.* 77:11–23

Giner-Sorolla R, Kupfer T, Sabo J. 2018. What makes moral disgust special? An integrative functional review. In *Advances in Experimental Social Psychology*, Vol. 57, ed. JM Olson, pp. 223–89. San Diego, CA: Elsevier

Gintis H. 2000. Strong reciprocity and human sociality. *J. Theor. Biol.* 206(2):169–79

Gold N, Pulford BD, Colman AM. 2015. Do as I say, don't do as I do: Differences in moral judgments do not translate into differences in decisions in real-life trolley problems. *J. Econ. Psychol.* 47:50–61

Gollwitzer M, Bushman BJ. 2012. Do victims of injustice punish to improve their mood? *Soc. Psychol. Personal. Sci.* 3(5):572–80

Goodwin GP. 2015. Moral character in person perception. *Curr. Dir. Psychol. Sci.* 24(1):38–44

Goodwin GP. 2017. Is morality unified, and does this matter for moral reasoning? In *Moral Inferences*, ed. J-F Bonnefon, B Trémolière, pp. 9–36. New York: Routledge/Taylor & Francis

Graesser AC, Singer M, Trabasso T. 1994. Constructing inferences during narrative text comprehension. *Psychol. Rev.* 101(3):371–95

Gray K, Keeney JE. 2015. Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Soc. Psychol. Personal. Sci.* 6(8):859–68

Gray K, Schein C, Ward AF. 2014. The myth of harmless wrongs in moral cognition: automatic dyadic completion from sin to suffering. *J. Exp. Psychol. Gen.* 143(4):1600–15

Greene JD, Sommerville RB, Nystrom LE, Darley JM, Cohen JD. 2001. An fMRI investigation of emotional engagement in moral judgment. *Science* 293(5537):2105–8

Guala F. 2012. Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behav. Brain Sci.* 35(1):1–15

Guglielmo S. 2015. Moral judgment as information processing: an integrative review. *Front. Psychol.* 6:1637

Guglielmo S, Malle BF. 2010. Can unintended side effects be intentional? Resolving a controversy over intentionality and morality. *Personal. Soc. Psychol. Bull.* 36(12):1635–47

Guglielmo S, Malle BF. 2017. Information-acquisition processes in moral judgments of blame. *Personal. Soc. Psychol. Bull.* 43(7):957–71

Guglielmo S, Malle BF. 2019. Asymmetric morality: Blame is more differentiated and more extreme than praise. *PLOS ONE* 14(3):e0213544
Gui D-Y, Gan T, Liu C. 2016. Neural evidence for moral intuition and the temporal dynamics of interactions between emotional processes and moral cognition. *Soc. Neurosci.* 11(4):380–94
Gürçay B, Baron J. 2017. Challenges for the sequential two-system model of moral judgement. *Think. Reason.* 23(1):49–80
Haidt J. 2001. The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychol. Rev.* 108(4):814–34
Haidt J, Björklund F. 2008. Social intuitionists answer six questions about moral psychology. In *Moral Psychology*, Vol 2: *The Cognitive Science of Morality: Intuition and Diversity*, ed. W Sinnott-Armstrong, pp. 181–217. Cambridge, MA: MIT Press
Haidt J, Björklund F, Murphy S. 2000. *Moral dumbfounding: when intuition finds no reason*. Work. Pap., Univ. Va., Charlottesville. **https://pdfs.semanticscholar.org/d415/e7fa2c2df922dac194441516a509ba5eb7ec.pdf**
Hare B. 2017. Survival of the friendliest: *Homo sapiens* evolved via selection for prosociality. *Annu. Rev. Psychol.* 68:155–86
Heider F. 1958. *The Psychology of Interpersonal Relations*. New York: Wiley
Hershcovis MS, Bhatnagar N. 2017. When fellow customers behave badly: witness reactions to employee mistreatment by customers. *J. Appl. Psychol.* 102(11):1528–44
Hofmann W, Brandt MJ, Wisneski DC, Rockenbach B, Skitka LJ. 2018. Moral punishment in everyday life. *Personal. Soc. Psychol. Bull.* 44(12):1697–711
Holleman B. 1999. Wording effects in survey research: using meta-analysis to explain the forbid/allow asymmetry. *J. Quant. Linguist.* 6(1):29–40
Janoff-Bulman RJ, Sheikh S, Hepp S. 2009. Proscriptive versus prescriptive morality: two faces of moral regulation. *J. Personal. Soc. Psychol.* 96(3):521–37
Kahane G. 2012. On the wrong track: process and content in moral psychology. *Mind Lang.* 27(5):519–45
Kahane G, Shackel N. 2010. Methodological issues in the neuroscience of moral judgement. *Mind Lang.* 25(5):561–82
Kahneman D. 2013. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux
Keltner D, Kogan A, Piff PK, Saturn SR. 2014. The sociocultural appraisals, values, and emotions (SAVE) framework of prosociality: core processes from gene to meme. *Annu. Rev. Psychol.* 65:425–60
Kim M, Park B, Young L. 2020. The psychology of motivated versus rational impression updating. *Trends Cogn. Sci.* 24(2):101–11
Kiyonari T, Barclay P. 2008. Cooperation in social dilemmas: Free riding may be thwarted by second-order reward rather than by punishment. *J. Personal. Soc. Psychol.* 95(4):826–42
Klein C. 2011. The dual track theory of moral decision-making: a critique of the neuroimaging evidence. *Neuroethics* 4(2):143–62
Klonick K. 2015. Re-shaming the debate: social norms, shame, and regulation in an internet age. *Md. Law Rev.* 75(4):1029–65
Kneer M, Machery E. 2019. No luck for moral luck. *Cognition* 182:331–48
Koralus P, Alfano M. 2017. Reasons-based moral judgment and the erotetic theory. In *Moral Inferences*, ed. J-F Bonnefon, B Trémolière, pp. 77–106. New York: Routledge/Taylor & Francis
Krasnow MM, Cosmides L, Pedersen EJ, Tooby J. 2012. What are punishment and reputation for? *PLOS ONE* 7(9):e45662
Kriss PH, Weber RA, Xiao E. 2016. Turning a blind eye, but not the other cheek: on the robustness of costly punishment. *J. Econ. Behav. Organ.* 128:159–77
Kurzban R, Burton-Chellew MN, West SA. 2015. The evolution of altruism in humans. *Annu. Rev. Psychol.* 66:575–99
Landy JF, Goodwin GP. 2015. Does incidental disgust amplify moral judgment? A meta-analytic review of experimental evidence. *Perspect. Psychol. Sci. J. Assoc. Psychol. Sci.* 10(4):518–36
Laurent SM, Clark BAM, Walker S, Wiseman KD. 2014. Punishing hypocrisy: the roles of hypocrisy and moral emotions in deciding culpability and punishment of criminal and civil moral transgressors. *Cogn. Emot.* 28(1):59–83

Laurent SM, Nuñez NL, Schweitzer KA. 2016. Unintended, but still blameworthy: the roles of awareness, desire, and anger in negligence, restitution, and punishment. *Cogn. Emot.* 30(7):1271–88
Laurent SM, Reich BJ, Skorinko JLM. 2019. Reconstructing the side-effect effect: a new way of understanding how moral considerations drive intentionality asymmetries. *J. Exp. Psychol. Gen.* 148(10):1747–66
Leibbrandt A, López-Pérez R. 2014. Different carrots and different sticks: Do we reward and punish differently than we approve and disapprove? *Theory Decis.* 76(1):95–118
Leloup L, Meert G, Samson D. 2018. Moral judgments depend on information presentation: evidence for recency and transfer effects. *Psychol. Belg.* 58(1):256–75
Lerner JS, Tetlock PE. 1999. Accounting for the effects of accountability. *Psychol. Bull.* 125(2):255–75
Leuthold H, Kunkel A, Mackenzie IG, Filik R. 2015. Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Soc. Cogn. Affect. Neurosci.* 10(8):1021–29
Linke LH. 2012. Social closeness and decision making: moral, attributive and emotional reactions to third party transgressions. *Curr. Psychol.* 31(3):291–312
Malle BF, Guglielmo S, Monroe AE. 2014. A theory of blame. *Psychol. Inq.* 25(2):147–86
Malle BF, Holbrook J. 2012. Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *J. Personal. Soc. Psychol.* 102(4):661–84
Malle BF, Scheutz M, Arnold T, Voiklis J, Cusimano C. 2015. Sacrifice one for the good of many? People apply different moral norms to human and robot agents. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction* (*HRI'15*), pp. 117–24. New York: ACM
Malle BF, Thapa S, Scheutz M. 2019. AI in the sky: how people morally evaluate human and machine decisions in a lethal strike dilemma. In *Robotics and Well-Being*, ed. MI Aldinhas Ferreira, J Silva Sequeira, G Singh Virk, MO Tokhi, EE Kadar, pp. 111–33. Cham, Switz.: Springer
Margoni F, Geipel J, Hadjichristidis C, Surian L. 2019. The influence of agents' negligence in shaping younger and older adults' moral judgment. *Cogn. Dev.* 49:116–26
Martin JW, Cushman F. 2015. To punish or to leave: Distinct cognitive processes underlie partner control and partner choice behaviors. *PLOS ONE* 10(4):e0125193
Martin JW, Cushman F. 2016. Why we forgive what can't be controlled. *Cognition* 147:133–43
May J. 2018. *Regard for Reason in the Moral Mind*. Oxford, UK: Oxford Univ. Press
McHugh C, McGann M, Igou ER, Kinsella EL. 2017. Searching for moral dumbfounding: identifying measurable indicators of moral dumbfounding. *Collabra Psychol.* 3(1):23
McNamara P. 2006. Deontic logic. In *Handbook of the History of Logic*, Vol. 7, ed. DM Gabbay, J Woods, pp. 197–288. Amsterdam: North-Holland
Mikhail J. 2011. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. New York: Cambridge Univ. Press
Mikula G. 2003. Testing an attribution-of-blame model of judgments of injustice. *Eur. J. Soc. Psychol.* 33(6):793–811
Monroe AE, Malle BF. 2017. Two paths to blame: Intentionality directs moral information processing along two distinct tracks. *J. Exp. Psychol. Gen.* 146(1):123–33
Monroe AE, Malle BF. 2019. People systematically update moral judgments of blame. *J. Personal. Soc. Psychol.* 116(2):215–36
Mullen E, Monin B. 2016. Consistency versus licensing effects of past moral behavior. *Annu. Rev. Psychol.* 67:363–85
Nadler J, McDonnell M-H. 2012. Moral character, motive, and the psychology of blame. *Cornell Law Rev.* 97:255–304
Nichols S, Mallon R. 2006. Moral dilemmas and moral rules. *Cognition* 100(3):530–42
Niedenthal PM, Rohmann A, Dalle N. 2003. What is primed by emotion concepts and emotion words? In *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*, ed. J Musch, KC Klauer, pp. 307–33. Mahwah, NJ: Erlbaum
O'Hara RE, Sinnott-Armstrong W, Sinnott-Armstrong NA. 2010. Wording effects in moral judgments. *Judgm. Decis. Mak.* 5(7):547–54
Patil I, Calò M, Fornasier F, Cushman F, Silani G. 2017. The behavioral and neural basis of empathic blame. *Sci. Rep.* 7:5200

Pedersen EJ, Kurzban R, McCullough ME. 2013. Do humans really punish altruistically? A closer look. *Proc. Biol. Sci.* 280(1758):20122723
Pedersen EJ, McAuliffe WHB, McCullough ME. 2018. The unresponsive avenger: more evidence that disinterested third parties do not punish altruistically. *J. Exp. Psychol. Gen.* 147(4):514–44
Przepiorka W, Berger J. 2016. The sanctioning dilemma: a quasi-experiment on social norm enforcement in the train. *Eur. Sociol. Rev.* 32(3):439–51
Reeder GD, Kumar S, Hesson-McInnis MS, Trafimow D. 2002. Inferences about the morality of an aggressor: the role of perceived motive. *J. Personal. Soc. Psychol.* 83(4):789–803
Riordan CA, Marlin NA, Kellogg RT. 1983. The effectiveness of accounts following transgression. *Soc. Psychol. Q.* 46(3):213–19
Robbennolt JK. 2000. Outcome severity and judgments of "responsibility": a meta-analytic review. *J. Appl. Soc. Psychol.* 30(12):2575–609
Rosas A, Aguilar-Pardo D. 2019. Extreme time-pressure reveals utilitarian intuitions in sacrificial dilemmas. *Think. Reason*. In press. **https://doi.org/10.1080/13546783.2019.1679665**
Ross L, Amabile TM, Steinmetz JL. 1977. Social roles, social control, and biases in social-perception processes. *J. Personal. Soc. Psychol.* 35:485–94
Rothschild ZK, Landau MJ, Sullivan D, Keefer LA. 2012. A dual-motive model of scapegoating: displacing blame to reduce guilt or increase control. *J. Personal. Soc. Psychol.* 102(6):1148–63
Royzman EB, Goodwin GP, Leeman RF. 2011. When sentimental rules collide: "norms with feelings" in the dilemmatic context. *Cognition* 121:101–14
Royzman EB, Hagan JP. 2017. The shadow and the tree: inference and transformation of cognitive content in psychology of moral judgment. In *Moral Inferences*, ed. J-F Bonnefon, B Trémolière, pp. 56–74. New York: Routledge/Taylor & Francis
Royzman EB, Kim K, Leeman RF. 2015. The curious tale of Julie and Mark: unraveling the moral dumbfounding effect. *Judgm. Decis. Mak.* 10(4):296–313
Russell JA, Barrett LF. 1999. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *J. Personal. Soc. Psychol.* 76(5):805–19
Sauer H. 2012. Morally irrelevant factors: What's left of the dual process model of moral cognition? *Philos. Psychol.* 25(6):783–811
Schaich Borg J, Hynes C, Van Horn J, Grafton S, Sinnott-Armstrong W. 2006. Consequences, action, and intention as factors in moral judgments: an fMRI investigation. *J. Cogn. Neurosci.* 18(5):803–17
Scheutz M, Malle BF. 2021. May machines take lives to save lives? Human perceptions of autonomous robots (with the capacity to kill). In *Lethal Autonomous Weapons: Re-Examining the Law and Ethics of Robotic Warfare*, ed. J Gaillot. Oxford, UK: Oxford Univ. Press. In press
Schnall S, Haidt J, Clore GL, Jordan AH. 2008. Disgust as embodied moral judgment. *Personal. Soc. Psychol. Bull.* 34(8):1096–109
Shalvi S, Gino F, Barkan R, Ayal S. 2015. Self-serving justifications: doing wrong and feeling moral. *Curr. Dir. Psychol. Sci.* 24(2):125–30
Shaver KG. 1985. *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. New York: Springer
Siegel JZ, Crockett MJ, Dolan RJ. 2017. Inferences about moral character moderate the impact of consequences on blame and praise. *Cognition* 167:201–11
Sinnott-Armstrong W. 2016. The disunity of morality. In *Moral Brains*, ed. SM Liao, pp. 331–54. Oxford, UK: Oxford Univ. Press
Sinnott-Armstrong W, Wheatley T. 2012. The disunity of morality and why it matters to philosophy. *Monist* 95(3):355–77
Sloman SA. 1996. The empirical case for two systems of reasoning. *Psychol. Bull.* 119(1):3–22
Stanley ML, Yin S, Sinnott-Armstrong W. 2019. A reason-based explanation for moral dumbfounding. *Judgm. Decis. Mak.* 14(2):120–29
Stohr K. 2012. *On Manners*. New York: Routledge
't Hart B, Struiksma ME, van Boxtel A, van Berkum JJA. 2018. Emotion in stories: facial EMG evidence for both mental simulation and moral evaluation. *Front. Psychol.* 9:613
Tannenbaum D, Uhlmann EL, Diermeier D. 2011. Moral signals, public outrage, and immaterial harms. *J. Exp. Soc. Psychol.* 47(6):1249–54

Tassy S, Deruelle C, Mancini J, Leistedt S, Wicker B. 2013. High levels of psychopathic traits alters moral choice but not moral judgment. *Front. Hum. Neurosci.* 7:229

Tomasello M, Vaish A. 2013. Origins of human cooperation and morality. *Annu. Rev. Psychol.* 64:231–55

Uhlmann EL, Zhu L. 2013. Acts, persons, and intuitions: person-centered cues and gut reactions to harmless transgressions. *Soc. Psychol. Personal. Sci.* 5(3):279–85

Van Dillen LF, van der Wal RC, van den Bos K. 2012. On the role of attention and emotion in morality: Attentional control modulates unrelated disgust in moral judgments. *Personal. Soc. Psychol. Bull.* 38(9):1222–31

Voiklis J, Kim B, Cusimano C, Malle BF. 2016. Moral judgments of human versus robot agents. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication* (*RO-MAN*), pp. 486–91. Piscataway, NJ: IEEE

Voiklis J, Malle BF. 2018. Moral cognition and its basis in social cognition and social regulation. In *Atlas of Moral Psychology*, ed. K Gray, J Graham, pp. 108–20. New York: Guilford

Weiner B. 1995. *Judgments of Responsibility: A Foundation for a Theory of Social Conduct*. New York: Guilford

Wheatley T, Haidt J. 2005. Hypnotic disgust makes moral judgments more severe. *Psychol. Sci.* 16(10):780–84

Xiao E, Houser D. 2005. Emotion expression in human punishment behavior. *PNAS* 102(20):7398–401

Yamagishi T, Horita Y, Takagishi H, Shinada M, Tanida S, Cook KS. 2009. The private rejection of unfair offers and emotional commitment. *PNAS* 106(28):11520–23

Yoder KJ, Decety J. 2014. Spatiotemporal neural dynamics of moral judgment: a high-density ERP study. *Neuropsychologia* 60:39–45

Young L, Cushman F, Hauser M, Saxe R. 2007. The neural basis of the interaction between theory of mind and moral judgment. *PNAS* 104(20):8235–40

Young L, Koenigs M, Kruepke M, Newman JP. 2012. Psychopathy increases perceived moral permissibility of accidents. *J. Abnorm. Psychol.* 121(3):659–67

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. *Rev. Philos. Psychol.* 1(3):333–49

Zalla T, Barlassina L, Buon M, Leboyer M. 2011. Moral judgment in adults with autism spectrum disorders. *Cognition* 121:115–26

Supplementary Material to *Moral Judgments*

Annual Review of Psychology, 72 (2021)

Bertram F. Malle

Link	Content
Appendix 1	Rise of articles featuring the term *moral* in their titles since 2000
Appendix 2	Terms of moral judgment in continuous vs categorical use (COCA analysis)
Appendix 3	The term *morally wrong* takes almost exclusively intentional actions as its object
Appendix 4	Lay definitions of *morally wrong*
Appendix 5	Aggregated results from Cameron et al. (2017)
Appendix 6	Comparison of present-tense and past-tense uses of classes of moral judgments
Appendix 7	Recommendations for measuring the moral judgments we intend to measure
Appendix 8	Aggregated results from McHugh et al (2017)

Supplemental Material: Annu. Rev. Psychol. 2021. 72.
https://doi.org/10.1146/annurev-psych-072220-104358

Appendix 1. Rise of articles featuring the term *moral* in their titles since 2000

Method: I searched PsycINFO for journal title (e.g., JN "Emotion") and TI "moral*", vs. no restriction on title, with Publication Dates of 2000-2009 vs. 2010-2019.
Results: Compared to a 1.5-times increase in nonmoral articles between the decades, moral-related articles increased by a factor of 4.0, which is 2.7 times larger.

Table 1.1 Counts of articles with "moral" in their titles, in 15 psychological journals during 2000-2009 and 2010-2019, compared to all articles and nonmoral ones

	Articles with "moral" in title			All articles in time span			Nonmoral articles		
	2000-2009	2010-2019	2000-2019	2000-2009	2010-2019	2000-2019	2000-2010	2011-2020	2000-2020
British Journal of Psychology	1	5	6	390	583	973	389	578	967
British Journal of Social Psychology	7	20	27	381	500	881	374	480	854
Cognition	15	67	82	1,044	1,945	2,989	1,029	1,878	2,907
Cognition and Emotion	6	19	25	645	1,334	1,979	639	1,315	1,954
Cognitive Science	1	13	14	407	814	1,221	406	801	1,207
Current Directions in Psychological Science	3	17	20	627	768	1,395	624	751	1,375
Emotion	3	19	22	530	1,314	1,844	527	1,295	1,822
European Journal of Social Psychology	4	35	39	646	892	1,538	642	857	1,499
Journal of Experimental Psychology: General	1	29	29	373	1,069	1,442	372	1,040	1,413
Journal of Experimental Social Psychology	12	73	85	918	1,499	2,417	906	1,426	2,332
Journal of Personality and Social Psychology	25	36	61	1,514	1,324	2,838	1,489	1,288	2,777
Judgment and Decision Making	6	24	30	176	517	693	170	493	663
Personality and Social Psychology Bulletin	9	56	65	1,327	1,238	2,565	1,318	1,182	2,500
Psychological Science	17	36	53	1,530	2,296	3,826	1,513	2,260	3,773
Social Cognition	4	11	15	318	351	669	314	340	654
All Journals	114	*460*	*573*	*10,826*	*16,444*	*27,270*	*10,713*	*15,984*	*26,697*

Table 1.2 Factor of increase in articles from 2000-2009 to 2010-2019

Journal name	Moral	Nonmoral	Abbrev.
British Journal of Psychology	5.0	1.5	BJP
British Journal of Social Psychology	2.9	1.3	BJSP
Cognition	4.5	1.8	Cog
Cognition and Emotion	3.2	2.1	CogEmo
Cognitive Science	13.0	2.0	CogSci
Current Directions in Psychological Science	5.7	1.2	CD
Emotion	6.3	2.5	Emo
European Journal of Social Psychology	8.8	1.3	EJSP
Journal of Experimental Psychology: General	29.0	2.8	JEP:G
Journal of Experimental Social Psychology	6.1	1.6	JESP
Journal of Personality and Social Psychology	1.4	0.9	JPSP
Judgment and Decision Making	4.0	2.9	JDM
Personality and Social Psychology Bulletin	6.2	0.9	PSPB
Psychological Science	2.1	1.5	PS
Social Cognition	2.8	1.1	SocCog
All Journals	*4.0*	*1.5*	

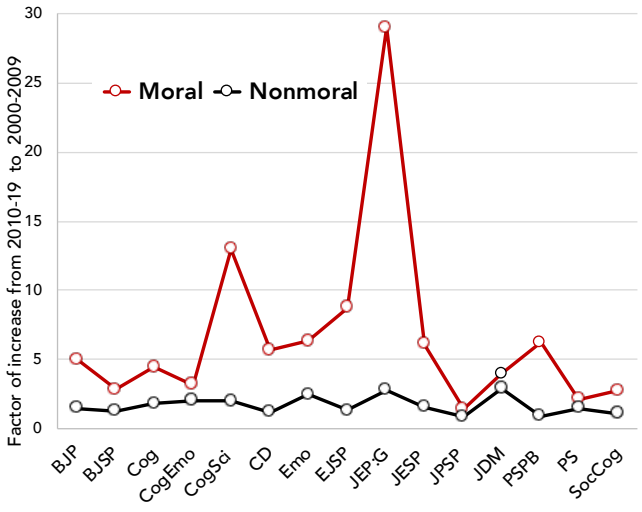


Table 1.3 Percentage of articles with "moral" in title out of all articles

Journal name	2000-2009	2010-2019
British Journal of Psychology	0.3%	0.9%
British Journal of Social Psychology	1.8%	4.0%
Cognition	1.4%	3.4%
Cognition and Emotion	0.9%	1.4%
Cognitive Science	0.2%	1.6%
Current Directions in Psychological Science	0.5%	2.2%
Emotion	0.6%	1.4%
European Journal of Social Psychology	0.6%	3.9%
Journal of Experimental Psychology: General	0.0%	2.7%
Journal of Experimental Social Psychology	1.3%	4.9%
Journal of Personality and Social Psychology	1.7%	2.7%
Judgment and Decision Making	3.4%	4.6%
Personality and Social Psychology Bulletin	0.7%	4.5%
Psychological Science	1.1%	1.6%
Social Cognition	1.3%	3.1%
All Journals	*1.1%*	*2.8%*

Appendix 2. Terms of moral judgment in continuous or categorical use (COCA analysis)

Method: To determine what it means to show continuous vs. categorical use, I first selected two comparison terms — one that is arguably continuous, namely *tall*, the other that is arguably categorical, namely *impossible*. I inspected their most frequent pre-collocates (adverbs and other phrases that directly precede the term) and thus accumulated a first set of continuous vs. categorical collocates. These included on the continuous side: *more x (than)*, *very x*, *as x as*, *how x*; on the categorical side: *almost x*, *become x*, *x or not*. Then I inspected frequent collocates of the moral judgment target terms and added ones that seemed meaningfully to imply continuous use (e.g., *pretty x*, *extremely x*) vs. categorical use (e.g., *completely x*, *simply x*). In the 1-billion Corpus of Contemporary American English (COCA) I then looked up the raw frequencies of each collocate-target pair and computed the percent of continuous out of total use and percent of categorical out of total use. Because these percentages differ as a function of other dominant uses of the terms I also computed a ratio of continuous over categorical comparisons.

Results: The data patterns suggest that (a) *bad* is used as a continuous concept similar to the comparison standard of *tall*; (b) most norm judgments (e.g., *required, forbidden, permissible*) are used as categorical concepts similar to the comparison standard of *impossible,* though *acceptable* has a notable continuous use pattern as well; (c) *wrong* shows both continuous and categorical uses but lies much closer to categorical norm judgments than to continuous evaluations. The reader is invited to delete the most frequent collocates for each moral judgment term to see that the results are robust over particular selection decisions. However, a more systematic analysis is needed that is grounded in linguistic theory and establishes inter-rater agreement for the classification of a given collocate as indicating continuous vs. categorical use.

Table 2.1 Summary of continuous and categorical use contexts for different classes of moral judgment and two comparisons terms

	Evaluation	Norm judgments						Wrongness	Comparisons	
	bad	*required*	*mandatory*	*prohibited*	*forbidden*	*permissible*	*acceptable*	*wrong*	*tall*	*impossible*
% Continuous out of total	38.37%	0.09%	0.28%	0.10%	0.37%	0.80%	5.42%	1.30%	24.03%	1.11%
% Categorical out of total	0.18%	0.27%	1.77%	0.63%	1.14%	2.43%	3.74%	1.37%	0.05%	8.75%
Continuous/Categorical Ratio	**216.3**	**0.34**	**0.16**	**0.16**	**0.33**	**0.33**	**1.45**	**0.95**	**476.3**	**0.13**

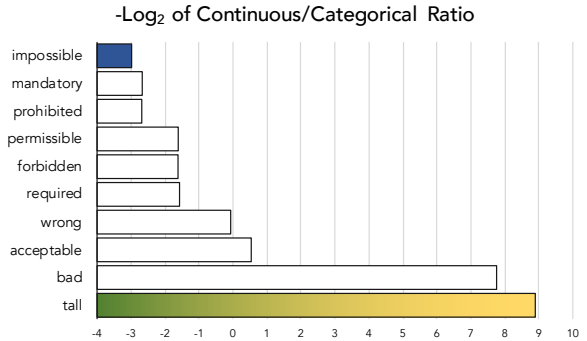
Table 2.2 Counts of continuous and categorical uses, broken down by collocates, for different classes of moral judgment and two comparisons terms

		Evaluation	Norm judgments						Wrongness	Comparisons	
	Phrase	*bad*	*required*	*mandatory*	*prohibited*	*forbidden*	*permissible*	*acceptable*	*wrong*	*tall*	*impossible*
Phrases that cast target adjective as continuous	more x (than)	88,025	27	15	1	9	9	779	420	6,918	103
	less x (than)	132	10	5	0	0	3	91	56	13	25
	as x as	5,707	3	3	0	4	1	15	152	948	71
	equally x	142	3	0	3	3	1	52	19	12	34
	very x	6,083	1	2	1	2	0	53	1,330	775	9
	how x	5,998	2	4	0	3	1	18	787	540	150
	pretty x	2,724	0	0	0	0	0	6	13	68	34
	particularly x	414	1	0	0	3	0	1	24	22	1
	especially x	220	2	0	1	3	0	0	7	16	1
	extremely x	172	0	0	0	0	0	0	18	40	1
	quite x	125	2	0	0	0	3	89	167	71	164
	incredibly x	109	0	0	0	0	0	0	21	11	2
	rather x	65	2	0	0	0	0	0	5	35	9
	exceptionally x	41	0	0	0	0	0	0	0	19	0
	fairly x	40	1	0	0	0	0	4	0	22	9
	unbelievably x	36	0	0	0	0	0	0	9	0	0
	extraordinarily x	26	0	0	0	0	0	0	7	12	0
	increasingly x	28	27	1	1	2	0	12	5	3	29
Phrases that cast target adjective as categorical	x or not	25	12	0	2	3	3	20	477	4	7
	almost x	4	25	30	1	6	0	6	4	3	3,483
	become x (+ lemmas)	81	31	129	9	11	9	156	20	6	620
	practically x	1	12	6	0	1	0	3	1	0	291
	simply x	76	31	0	4	7	7	0	397	1	211
	absolutely x	12	76	24	18	40	2	4	301	0	189
	completely x	16	2	0	9	19	3	42	922	0	143
	fundamentally x	12	2	0	1	0	0	0	255	0	23
	inherently x	117	1	0	0	0	0	0	213	0	20
	entirely x	44	0	2	0	0	7	21	200	0	28
	obviously	51	18	0	0	0	0	1	168	0	39
	definitely x	29	13	0	1	0	0	4	115	1	5
	intrinsically x	17	0	0	0	0	0	0	50	0	3
	certainly x	19	18	0	0	1	8	12	41	4	10
	unequivocally x	4	0	0	0	1	0	0	8	1	0
	perfectly x	1	0	0	0	0	16	503	15	0	3
Sums	CONTINUOUS	110,087	81	30	7	29	18	1,120	3,040	9,525	642
	CATEGORICAL	509	241	191	45	89	55	772	3,187	20	5,075
	Total # in COCA	286,875	90,304	10,809	7,161	7,833	2,263	20,648	233,477	39,643	57,981

Table 2.3 Continuous/Categorical Ratio from highest to lowest

tall	476.3
bad	216.3
acceptable	1.45
wrong	0.95
required	0.34
forbidden	0.33
permissible	0.33
prohibited	0.16
mandatory	0.16
impossible	0.13

Note: Reference terms in bold face



Appendix 3. The term *morally wrong* takes almost exclusively intentional actions as its object

Method: I searched for instances of the term "morally wrong" in COCA (https://www.english-corpora.org/coca/) in 2016. Of the 580 instances, I requested a random sample of 100, and these are shown below. I then classified each object of moral wrongness (what was designated "morally wrong") as intentional, unintentional, or undeterminable. When the excerpted text around the search term was too short to clarify what the object of wrongness was, I inspected the wider text in the original source and included it below. For three instances, no determination of intentionality was possible (e.g., mainly because the use was abstract or underspecified).

Result: 1 our of 100 instances binds "morally wrong" to an unintentional behavior (shown in red below). 96 are arguably intentional (shown in green), 3 are undeterminable (grey).

Table 3.1 One hundred randomly selected instances of *morally wrong* in COCA and the intentionality of the object of wrongness

	197N	EWS9	UATNo9	daN9o9yls di9ti	dUToo99yls di9ti	msi9siIAs7rli
e	eaar	TypS	oSuc Pdi	ngtxYtI ldNNitml sBtH ltdtm Ids **to have an abortion** fffYqdYtNOdPgtY tYqOt MgtPOdYqtmxYqxstHxROtit)	ldNNitml sBt tw<5t(tT YtI ldNNitml sBt0F5t(tEO9OsPgt%5t(tT	✓
F	eaav	uuE	OsOldYx sg	)gftxsOg **is using**)lgftNdltdsPtYqxgtxgtdNmdigt	ldNNitml sBt tdhh lPxsBtY tYqOtOYqxhdNt9lxshx9NOt HtlOg9OhYtH lt9Olg sg	✓
w	eaaw	)u	uIOlxhd	gYdYOgt Ht9l 9 lYx sdNxgItYqdYtxYtdYYOI9YgtY	ldNNitml sBt t**actions by a good intention** ftcqxgtntlOBlOYtY	✓
v	eaaw	)u	uIOlxhd	gYxHxOPtitdt9l 9 lYx sdYOtlOdg stYqOitdlOts YtgdixsBtYqdYt	ldNNitml sBt **actions (ex objecto) can be justified** titYqOtOsP	✓
0	eaaw	TypS	SdsIdsql s	**lives of unborn babies to be a choice** tOhdgOtxYtxgt	ldNNitml sBt tdsPtOhdgOtYqOtsdYx stYqdYtPOgYl igtxYgti sBtPOgYl ig	✓
%	eaa%	)u	cn)y	9dihqOhftusPtYqOsthdlOtYqOthNxshqOlt	ldNNitml sBt tdsPtxsPOHOsgxN **not to pay federal workers.** tntP tsMY	✓
r	Fe0	TypS	S) sxY l	pqith sYlxOgtdlOtmdNNxsBtYqOIgONROgtxsttdsPt YqOIgt YtSOhlxYitqdmgt YlI9OYtYqOtxI9dhYt H **walls** tt st **stopping or detecting infiltrators** tthlxYxh gOOtYqOItdg	ldNNitml sBt tdsPtxsOHHOhYxROtgYtPxROlYxsBt9O 9NOtY tdNYOlsdYxRO	✓
<	F	TypS	qxhdB	(tutN Yt Ht9NdiOlgtHOONtdlNOt **has done nothing**	ldNNitml sBt tdsPtxgtOxsBt9sxgqOPtitYqOtTuutdsPtYqOtIOPxd	✓
a	eaae	S	TTTggHxlO	cqOgOt9O 9NOtdlOthONOldYxsBts Yt sNitYqOxlt9l hNxRxYitxYtgOOIgtY tI dlOthONOIdYxsBtmqdYtYqOitP ftntIOdsti tqdPtxsti Ith NIstdYt m Idstmq tqdPtHdNNOstxstN ROtmxYqtds YqOltm IdtttT mtdBdxstYqOtqIhqt YOdhqOgtYqdY	sOt ldNNitml sBt tdsPtxYtgOOIgtmqOsti tHNdsYtYqdYtxstHl sYt HtYqOI	sxsYOsYx sdNt
e	eaa	TypS	SdsIdsql s	mdxYftpdltxgts YtYqOtdsgmOlft(tnYtxgt	ldNNitml sBt tdsPt9 NxYxhdNNitxsO9OPxOs **sacrifice the lives o** H	✓
ee	Fa	TypS	ugg hlOgg	YqO **use of embryos for cell research** tdl sYtY tYqOtPOgYlhYx st HtqIdst NxHOtdsPtd	ldNNitml sBt tdsPtgq NPtOtdssOPtitNdmft(tcqOth sYldlitdlBIOsY	✓
eF	Fr	S	T)IsxsB	nHtg IOYqxsBtxgtI ldNNitml sBtNOYMgt 99 gOtxYftcqOtq sOgYtdlBIOsYtxgtntl YqxstY **abortion** gt	ldNNitml sBt tdsPtg IO sOth NPtYdOtYqdYt9 gxYx stdsPtYqOstgYxHi	✓
ew	Fw	uuE	nsgYlgihq	9ldhYxhOgtYqO **discriminated against European Americans** tdsPtYqOlOH mOlOtYqt	ldNNitml sBt tdsPtsh sgYxYYx sdNftcqxgtNOBdNtgqxHYtdmditHI ItI ms	✓
ev	F0	S	Tue	m ltdBdxsgYt **the death penalty** tmqxhqtgqOtONxOROgtxgtYq	ldNNitml sBt tdsPtsh sgYxYYx sdNftlOOdstxgtdtldstdYq Nxhtss	✓
e0	F<	uuE	cqO N BSYP	**Sexual acts** ttYqdYttdlOts Yt9dlYt HtdstdhYt HtldlxYdNth xYgtdlOt	ldNNitml sBt tOhdgOtYqOitdlOtOggOsYxdNNithN gOPtY tlO9l PhYx st	✓
e%	eaa0	)u	uYNdsYxh	dYYdht st9sxgqlOsYftOtlOOhYOPt **punishment** ts YtOhdgOtxYgt	ldNNitml sBt tYtOhdgOtxYtP OgtsMYtm lftpftfuPOs	✓
er	F%	uuE	IdsxgY	t(tnt 99 gOtYqOt **death penalty** s YtOhdgOtxYtxgt	ldNNitml sBt tYtOhdgOtxYtxgtxsOHHOhYxROtdsPtPdsBOl gftlYqOlI IO	✓
e<	Fe0	TypS	EOsROl	dt**mass shooting for terrorist reasons** tm NPtOth sgxPOlOPt	ldNNitml sBt titg hxOYitxsBtgdxPftYtYqOtYOlI lxgYtIditYqxs	✓
ea	Fev	S	TT	IOdglOtY PditgYtw0t9OlhOsYtgdxP **smoking marijuana** txgt	ldNNitml sBt th I9dlOPtmxYqtr9OlhOsYtdhtxstea<rftROltYqOtsOY	✓
F	eaar	TypS	oSuc Pdi	9lOBsdsYtm IOstsPOlB t9lOsdYdNtYOgYxsBt HtYqOxltHOYqOgtItY	ldNNitml sBt tH ltdtm IdstY t **have an abortion** txHtYOgYxsBtxsPxhdYOgtdsi	✓
Fe	F0	TypS	gY s	dstOIdxNtY tYqOtgY stql sxhNOftnYtm NPtOt	ldNNitml sBt tH l **the execution to go forward** txHtYqOtlitPxPtsMY	✓
FF	eaav	S	nsPxIdBq	xgtsdhhO9YdNOtdI sBtYOOsgt ltdsi PitONgOft **Homosexuality** txgt	ldNNitml sBt tqOtgdigtfdNNgt)xhqdONtdhgstgxggxHxOPftOtdPI sxgqOg	✓
Fw	Fv	TypS	S) sxY l	uIOlxhdtMgt9OlhO9Yx st HtBdINxsBftmth NPt **gambling** tOt	ldNNitml sBt txHtgYdYOtB ROlsIOsYtmdgtP xsBtxYt(tOlY	✓
Fv	eaav	uuE	OsOldYx sg	I PONtxgtY tlOg9 sPtY tYqOtxsYxYxROtgOsgOtYqdYtg IOYqxsBtxg	ldNNitml sBt txstdtBxROstgxYdYx sftyYqxhdNt9l NOIgtP ts Yth IOtmxYq	sPOYOlIxsdNO
F0	eaae	TypS	pdgqgY	gYdsPxsBtxstYqOtdhBl sPtdsPtmdYhqxsB **racist practices** tmdgtNxOtmdYhqxs dt9NxhtgY sxsBftntsPOlgY PtYqdYtmqdYtm NPtOt	ldNNitml sBt txstEdNNdgt ltxROl9 Ntmdgts tPxHHOlOsYtqOlOftntqdP	✓
F%	eaar	TypS	gY s	tntONxOROtY **intimacy that goes on between two consenting adults** xg YqOtgxsOggt HtYq gOtdPNYgtdsPtgq NPts YtOtgqdlOPtmxYqtYqOtm lNPftusl Og9OhxdNNits YtmxYqt Iti Yqft(tqdgYxYitstqdgtYqOtlxBqYtY t9l I YOtYqO Bdith IIsxYitYql BqtYqOtI RxOtdsPtYONORxgx stsPgYlitYts YtY tYqOt OYOsYt HtgqdlxsBtxsYxIdhitmxYqtRxOmOlgtmq tqdROts th sYl Nt ROltmqO OxsBt9l PhOPft(tcqOlOtxgtg IOYqxsBt	ldNNitml sBt txst NNim PtdsPtnt9lditxYthdstOth llOhYOPtOhdgOtYqO	✓
Fr	eaa<	uuE	S hxdNOgIhq	tntmdsYtY t9l 9 gOtdtBOsOldNtYOgYqYtH ltPOhxPxsBtmqOstdsPtmqOYqOltYqOtY HtdtONxOHtgigYOItY thqxNPlOstxgtI ldNNitPOHOsgxNOftnHtxYtxgtOROlt YqO **teaching this system to children** tmxNNtIOdstYqdYtNdYOltxstNxHOtY Y tq NPtONxOHgtYqdYtYqOitm NPtI gYtNxONits YtqdROthq gOstH ltYqOIgO YqOitqdPtdhhOggtY tdsitdNYOlsdYxROgtYqOs	ldNNitml sBt t Htmq OROlt9lOgIOgtY txI9 gOtYqxgtgigYOItdsPtY thq gOtH l	✓
F<	eaaw	)u	uIOlxhd	ftcqOitdlOtgdixsBtYqdY **an action** thdsts YtOtPBOPt	ldNNitml sBt tgxI9NititN xsBtdYtYqOtldYOlxdNtqd99OsxsBt ltdYtxYg	✓
Fa	eaa<	uuE	qIhqSYdYO	gq mstYqdYtYqOtSOROsYqPdituPROsYxgYtqIhqtYOdhqOgtYqdYtxYtxg	ldNNitml sBt tY **to be a member of or pay dues to a labor**	✓
w	Fee	n	usnsRxYdYx sSxs	ftussdtqdPtY tdPIxYtdgtmONNtYqdYtxYtmdgt9l dNxt	ldNNitml sBt tY **creep about someone 's house** tNxOtYqxgt9OO9xsBtdsP	✓
we	Fev	S		YqtgxPOgtnYgtNxOti th NPtgditxYgt	ldNNitml sBt tY **deprive school choice or hurt small business** tmxYqtmdBO	✓
wF	FF	S	TTTggHxlO	ftngtxYtI ldNNitml sBtTxsOYit9OlhOsYtgditxYtxgt	ldNNitml sBt tY **do what you claim to have done** tT m	✓
ww	Fev	)u	TdYO B	gNxYtYqOthdl YxPtdsPtBNdlftcqOtONxOHtYqdYtxYtxg	ldNNitml sBt tY **eat animals** xgtd99OdNxsBtdsPtIdiOtdgtdtg9OhxOg	✓
wv	eaa0	uuE	OdNYqS hxdNp	9l RxPOtgHHxhxOsYtNOBdNt9l YOhYx sY tq g9xYdNgt ltYqOtNxHOtxgt	ldNNitml sBt tY **ignore a family 's objections to donation** tROlhdgY	✓
w0	eaaw	TypS	oSuc Pdi	ftnYtqdgtY tP tmxYqtYqOtHdhYtYqdYtxYtxgt	ldNNitml sBt tY **oppress people** dgOPt sthqdldhYOlxgYxhgt ltOqdRx l	✓
w%	eaaF	S	SSs) Is	xYtI ldNNitml sBtY **withdraw their aid** tdsPtxgtxYt	ldNNitml sBt tY tmxYqPldmtdxPtHI It9O 9NOtOdBOltY tm ltmqOstYqOlO	✓
wr	F0	uuE	uIOlnsPxds	ft(tugxPOtHl ItYqOgOth shOlsgtYqOlOtxgtdNg tg IOYqxsBt	ldNNitml sBt tmqOstnsPxdsg **are prevented from maintaining**	✓
w<	eaa	TypS	pdgqgY	Sq lYYOlItgdhlxHxhOgtdlOtsOhOggdlit **to preserve** tYqOtuIOlxhdstN sBYOlIt gYdsPdlPt HtNxRxsBtYtYqOlOtxgts Yq	ldNNitml sBt tmxYqtuIOlxhdft(tcqOt9NxhtdNg tONxOROgtYqdYtNOgg	✓
wa	eaae	uuE	IOxBsuHHdxIg	T tghq **vast plan of quick conquest** tqdPtOROltOOstH lINdYOPtxstI POlst qxgY liftu99dlOsYNitxYtsOROlt hhllOPtY tYqOtdROldBOtd9dsOgOtYqdYtY mdgtdsiYqxsB	ldNNitml sBt tmxYqtxYftd9dstMgtPxRxsOtIxggx stY tlOdNxOtdfnhqx	✓
v	Fw	)u	S9 lYgnNN	OsYhftuPPOPt)xYhqONNtntP tsMYtgOOtdsiYqxsBt	ldNNitml sBt tmxYq **playing against Negroes, Indians, Russians** t ltdsi	✓
ve	FF	uuE	cqO N BSYP	t**homosexual genital acts** tdlOtdNmdigt OhYxRONitg9OdxsBt	I ldNNitml sBt H ltqI gOdNtdhYgtdlOtdNmdigt OhYxRONitORxNt	✓
vF	Fee	TypS	S) sxY l	ltdOltq 9Ogft(t)dlxYit **Gay behavior** tgYxNNt	ldNNitml sBt t(tcqOtYlYqtxgtYqdYtYqOtRdgYtldIxYit HtORdsBONxhdNgt	✓
vw	eaa<	uuE	qIhqSYdYO	pqOstYqOtYqO hldYtxIdBxsOgtYqdYtgqOth NPtdtRxhYxIt H **religious persecution** tgqOtgq NPtgOOtYqdYtYqxgtIxgH lYsOtm NPtOts YtIOlONi 9dxsHNtYtdNg tYqdYtxYtm NPtO	ldNNitml sBt tdhh lPxsBtY tqOlt msth I9lOqOsgxROtP hYlxsOftSqO	✓
vv	FeF	TypS	pdgqgY	YqdYtmqxNOt sNitFwt9OlhOsYt HtuIOlxhdsgtONxOROt **divorce** tY tOt	ldNNitml sBt taFt9OlhOsYtONxOROtxYtxgtxII ldNtH ltdtldllxOPtlds	✓
v0	eaa<	S		YqOitONxOROthOlYdxstYqxsBgtdlOtgxstlOHOlOsh tY **abortion** tdsPthOlYdxstYqxsBg dlOt	ldNNitml sBt tdsPtOhdgOtYqOitO9IOggtYqOxltRxOm9 xsYgtYqOitdlO	✓
v%	eaar	S	TcdNTdYx s	YqdYtmOtY t HtldhOtmdgtgxI9NitYqxgtxYtmdgt	tdsPtxYtOhdIOtgOOsYNitxNNOBdNtY **discriminate against people** t stYqOtdgxgt HtldhO ldNNitml sBt	✓
vr	F<	uuE	dmNxh N	YNdmxsBt **abortion** txgtY t9OlgdPOtYqOtuIOlxhdst9O 9NOtYqdYtxYtxgt	ldNNitml sBt tdsPtYqOtHxlgYtgYO9tY mdlPtdhh I9NxgqxsBtYqdYtxgtdtIdggxRO	✓
v<	F0	uuE	dmNxh N	YqOlOtdlOtg IOtgOlx gtI ldNtml sBgtYqdYtqdlIts t sOtdsPtYqdYtYqgtIdit Ot9l qxxYOPtdsPtYqOlOtdlOtg IOt **serious harms to others** tmqxhqtdlOts Y	ldNNitml sBt tdsPtYqgtgq NPts YtOt9l qxxYOPftsOth NP	✓
va	eaaw	)u	uIOlxhd	lth sPhYtP Ogts YtitYqdYtROlitHdhYtIdOtYqOt **action**	tdgthOlYdxstYldPxYx sdNtH lINdYx sgtg99 gOPftcqOtdhYx stOh IOgt I ldNNitml sBtmqOstdNNtYqxsBgth sgxPOlOPtYqOlOtxgts Ytdt ldNNitml sBt 9l 9 lYx sdYOtlOdg stxstYqOtdhYtgYxHxsBtYqOtPxgRdNOf	✓
0	FF	S	TTNxYxhg	xstYqOtxIYqt HtdtqIdstOxsBtONxORxsBtxYtxgt	ldNNitml sBt tYtdtsdll mtIdlxYitd99l ROt Ht **cloning** tY tdxPtIOPxhdN	✓
0e	eaa	TypS	TcxIOg	sdYx stYqdYts mtPdlOgtY **assert its freedom** txgts YtIOlONit	ldNNitml sBt tYtBO 9 NxYxhdNNitsgsPftnYtlOdgtYqOtPOI hldYxh	✓
0F	eaaF	S	uIxsNOi	dtB PtlOh lPt HtgdixsBtYqdYt **children out of wedlock** txgt	ldNNitml sBt tH ltYqOtlxhqtdgtmONNtdgtH ltYqOt9 l	✓

53	2011	SPOK	Fox_Five	[corporations…] would often **move to a state where they don't deal with unions**. That is against the law. It's something it's wrong. It's	morally wrong , historically wrong. In this case, I will give	✓
54	2001	MAG	Jet	in hope. Not only is **profiling** by race or religion	morally wrong , it distracts us from bringing the perpetrators to	✓
55	1998	SPOK	NPR_Morning	seduce us with Seahaven 's charms, too. It's	morally wrong , it's baldly deceptive, it's fabulously false,	undeterminable
56	1998	SPOK	Fox_Sunday	[Clinton] was less than candid about this **private act** [with Monica Lewinski]. It was wrong,	morally wrong , perhaps legally wrong. JACKSON: It did not come	✓
57	1993	SPOK	CBS_Morning	that **the homosexual lifestyle** is not acceptable, that it's	morally wrong , that homosexuals, indeed, are not entitled to	✓
58	1991	NEWS	WashPost	government judges and told them enough was enough: It's	morally wrong , the lawyers said, for the state to **try to**	✓
59	2010	SPOK	Fox_Beck	to a bill that federally funded **abortions**, because it's	morally wrong , they thought. They were called the "Stupak 13	✓
60	2001	ACAD	TheologStud	**policy** may therefore prove to be as unworkable as it is	morally wrong , ultimately serving only those whose goal is unlimited	✓
61	2015	SPOK	ABC	What **happened to Rachel and Logan** [being used by police as "confidential informants"] was wrong. It's	morally wrong . And the system needs to change. ELIZABETH-VARGAS#	✓
62	2000	FIC	Bk:HighFive	Ranger leaned closer and lowered his voice. " Let me explain my work ethic to you. I **do n't do things** I feel are	morally wrong . But sometimes my moral code strays from the norm.	✓
63	1993	NEWS	USAToday	Church congregation. The church says **homosexuality** is	morally wrong . CONNECTICUT HARTFORD - Four girls, ages 9-14, have	✓
64	2000	SPOK	NPR_TalkNation	56 percent of Americans believe that **homosexual behavior** is	morally wrong . Earlier this month, in an emotional and divisive vote	✓
65	1994	MAG	ChristCentury	When I asked Butler if his department engaged in teaching values through any of its activities, he told me that in some of their youth education efforts they stressed that some things are	. For example: Smith and Wessons are **[= using guns]** not the morally wrong way to solve problems.	✓
66	1995	SPOK	NPR_Weekend	adamantly opposed to **sex before marriage** because he thinks it is	morally wrong . He also believes cohabitation has contributed to the	✓
67	2008	MAG	America	in this way be eliminated. "Such **euthanasia** is always	morally wrong . Here the church insists on the important distinction	✓
68	2010	MAG	Ms	into submission by **threatening their sacramental life** is	morally wrong . However, the Catholic Church 's history of opposing	✓
69	1999	ACAD	SocialResrch	We do not have **chattel slavery** because the precipitate of our historical experience has determined it to be	morally wrong . However, we have religious toleration because we have	✓
70	1995	MAG	TIME	Opposing sides, each consisting of trusty comrades, have lined up and fired. The salvo: You 're wrong-and not only that, you 're	morally wrong . If Washington plays the naked power game, if Los	undeterminable
71	1992	NEWS	SanFranChron	share by the end of 1992. This political-economic **bullying** is	morally wrong . In fact, under worldwide agreements the United States	✓
72	2008	ACAD	TheologStud	moral teaching that **homosexual** acts are intrinsically	morally wrong . In recent years, however, some have challenged this	✓
73	2003	MAG	NatlReview	not their responsibility, and therefore **punishment** of them is	morally wrong . It is the kind of argument that liberals use because	✓
74	1993	MAG	America	obvious if the object is **characterized** in advance as	morally wrong . No theologian would or could contest the papal statement	✓
75	1998	ACAD	AfricaToday	that **racism, segregation, and second-class citizenship** were	morally wrong . Nyerere espoused a citizenship based on the principle of	✓
76	2005	MAG	USCatholic	ethic does not in any way explicitly condemn **all war** as	morally wrong . One of the heroes of the Old Testament is Joshua	✓
77	1995	ACAD	AcademicQs	**homosexual conduct to be (like adultery and dishonesty)**	morally wrong . Professor Nussbaum 's testimony was, however, by her	✓
78	2001	SPOK	CBS_Sixty	To just proceed to **kill somebody who has no concept of whats happening**, who does nt understand, who isnt being punished by it -- I mean, theres a line thats just	morally wrong . ! STAHL: Carla Ryan vows to press	✓
79	2000	ACAD	AcademicQs	, but will instead sometimes label their **actions** as being	morally wrong . Suppose, for example, a father **beats** his children	✓
80	2000	SPOK	CNN_LiveSun	40 percent of the 1,200 people surveyed said [engaging in/conducting] **the Human Genome Project** would be harmful, but 47 percent said it was not	morally wrong . The practical use of genome research could come sooner	✓
81	1993	SPOK	CBS_Morning	than half of the public thinks that **the homosexual lifestyle** is	morally wrong . They do n't believe even in, sort of,	✓
82	2010	SPOK	CBS_NewsEve	[talking about a law that will give Joe Arpaio "the tools to step up his efforts to combat the flood coming across Arizonas border with Mexico."] MAYOR-PHIL-GORDON: It -- its just	morally wrong . We -- this country isnt about having people wear armbands	✓
83	2012	ACAD	Futurist	says Whitby. "**Prohibition** would, on balance, be	morally wrong . What is morally right is building and employing such	✓
84	1997	SPOK	Ind_Springer	I can understand freedom of speech and things like that, but that is so wrong, and that is	. You do n't need to **use a little child** that lost her life […] for morally wrong entertainment. It's wrong.	✓
85	1994	MAG	Ms	I didn't want to rush in with too little evidence, but if any women had reported rapes it **[not doing anything]** would have been	morally wrong . "As the final meeting neared, Gabriel, Hayn…	✓
86	1991	NEWS	WashPost	John Gutfreund **would do anything** at the edge of legality or	morally wrong . "# Such feelings are widely held by people who	✓
87	2007	NEWS	Chicago	can urge them to disobey **segregation ordinances**, for they are	morally wrong . "# That's not what you hear from conservatives	✓
88	2003	NEWS	WashPost	all equally suffering. To **ignore great groups of children** is	morally wrong . "# The recommendation to leave Head Start as is	✓
89	1991	MAG	NatlParks	for a passionate belief -- that **owning another human being** was	morally wrong . # The best evidence of this dedication is the fact	✓
90	1991	NEWS	Atlanta	We believe that this discriminatory policy is contrary to the best interests of Westminster. We further believe that this **policy**, no matter how pure its religious purpose, is in practice anti-Semitic and	morally wrong . # The policy is not essential to its mission as	✓
91	2006	SPOK	NPR_Morning	**issuing demands** to the American people. It? s just	morally wrong . LUDDEN: Krikorian says the marches also created a	✓
92	1993	MAG	America	What would make a whole marriage ruled by the intention to avoid children through N.F.P. morally flawed is exactly what makes any **act of intercourse** that has been rendered sterile by some artificial birth control	morally wrong : the same anti-life **intention**. Surely, when N.F.P.	✓
93	2012	NEWS	CSMonitor	government ; most disagreed that **homosexual relationships** are	morally wrong ; and few agreed that basic health insurance is a right	✓
94	1992	MAG	Futurist	zero. **Fattening** pigs and cattle on grains was viewed as	morally wrong ; grains are for people, said "green "consumers	✓
95	1993	SPOK	ABC_DayOne	Part of what was wrong about that night was because you had to **beat Rodney King as much as you did**. Is that true? Mr. POWELL: We did. That wasn't wrong, as far as criminally wrong, but as far as society wrong, there should have been a better way SCHADLER What you're're saying is that you were technically right, but perhaps	morally wrong ? Mr. POWELL: Maybe. Maybe SCHADLER That's a	✓
96	1991	SPOK	CNN_Crossfire		? Mr. WOLTZ: What's morally wrong? Mr. BAUER: Is it morally wrong for 13 and 14-year-olds to **be having multiple sexual** moral dimension. Mr. BAUER: Do you think it's morally wrong **partners**?	✓
97	2002	SPOK	CNN_Crossfire	Here are the results [of a poll]. Is it morally acceptable to clone a human being ? Seven percent say yes. Is it morally wrong	? Ninety percent say it is morally wrong to **do what you claim to have done.** Is it	✓
98	2010	SPOK	NPR_TellMore	white friends **make racial jokes**, and I think that's morally wrong	? So, what would she say then? Do n't	✓
99	2008	MAG	Esquire	use it. "Is **what was done to Jose Padilla** morally wrong	? "I really can not talk about that, however	✓
100	1996	SPOK	CBS_Sixty	Did you ever, for one second, think **you might be doing something** wrong in any way -- not just legally -- that you might -- you might be doing something morally wrong	? Ms-CONNATY: No, I did not. [**not letting father of child help raise the child**]	✓

Appendix 4. Lay definitions of *morally wrong*

Method: We prompted participants who had made judgments of illegal and nonillegal violations this way:
"In the question you answered —*"Do you think what the person did is morally wrong?"*—what did *morally wrong* mean to you?"
Results: The lay definition of morally wrong combines evaluation (bad, harmful) with reference to norms for intentional behaviors. Evaluations are directed either at the action itself or its consequences; combining the two facets into one leads to a count of 17 (57%) for evaluations.

Table 4.1 Summary of categorized lay definitions of *morally wrong*

Coded category	How many of 30 participants mentioned something falling under the category	
Norms	17	57%
Intentionality	12	40%
Harmful consequences	11	37%
Bad	7	23%

Table 4.2 All lay definitions *morally wrong and the categories into which they were classified*

Verbal response to the question "what did morally wrong mean to you?"	Intentionality	Norms	Bad	Harmful consequences	Other unclassified
If the action was **intentional** and **caused harm** to someone else.	1			1	
Anything that would generally be seen as **frowned upon**		1			
Against the will of God		1			will of God
Moralizing is when you view an action as **right or wrong**. Morally wrong means it is something considered "wrong" by **societal expectations** of what moralization is.		1			
Doing something that **society dictates** is **bad**.		1	1		
Bad or not of a good nature or consideration of others.			1		
If it **hurts** someone else, sometimes even accidentally, it is morally wrong.				1	even accidentally
It means in the **morals an values** of your beliefs how would you rate the person's actions accordingly.		1			
Morally wrong to me means something that someone did that is completely wrong, horrifying, or **damaging to others.**				1	horrifying
Morally wrong to me means that the action is considered **bad** in **society** and is **looked down upon**.		1	1		
It means **bad** behavior, bad character.			1		character
That it was **not right** that the person **intentionally** did what they did	1	1			
It **hurt** somebody with malice **intent** for the most part. If it hurt a lot of people by accident, then it's **still wrong**.	1			1	even accidental
Something that violates societal **norms** of ethical behavior		1			
Something **bad** that was done **intentionally** or irresponsibly.	1		1		evil
When the person action was just downright **mean spirited** or evil	1				
Something that we as humans can all **universally** agree on as **bad**.		1	1		
Was it wrong to do by **hurting** someone else.				1	
If a person did something that **harmed** another **purposely** it's definitely morally wrong.	1			1	
that what the person did was wrong, un-ethical, **not** the **good** or **right** thing to do.		1			
Morally wrong took factors like **intentionality** and degree of **harm** into account. That's how I defined it.	1			1	
If it went against **all that we believe**, if it personally **hurt** someone, was very very **bad**.		1	1	1	
Knowingly committing a **moral** indescretion	1	1			
Doing something **God** would disapprove of, **not accidents** unless the accidents were people being really careless.	1	1			God disapproves
IIf **others wouldn't** do it or **if it wasn't accident**	1	1			

Whether it was **appropriate** to do or not, and if it **caused harm** to someone else.		1		1	
Morally wrong meant **socially unacceptable** and/or unlawful behavior.		1			
Deliberately causing **real harm** is definitely a confident morally wrong. **Accidental** harm is more gray, and causing no one harm isn't morally wrong.	1			1	accidental is gray
When someone did it **on purpose** and it endangered or **negatively effected** other people.	1			1	
It was **sinful**		1			
Total count out of 30 participants	**12**	**17**	**7**	**11**	

Appendix 5. Aggregated Results from Cameron et al. (2017)

Cameron, C. D., Payne, B. K., Sinnott-Armstrong, W., Scheffer, J. A., & Inzlicht, M. (2017). Implicit moral evaluations: A multinomial modeling approach. *Cognition, 158*, 224–241. https://doi.org/10.1016/j.cognition.2016.10.013

Method. Participants were presented with words selected to represent actions that are usually considered either morally wrong or morally neutral. For each word, participants had to judge whether the described action was morally wrong (M key) or not (Z key). They were given a response deadline of 400 - 800 ms (depending on study), but all their responses, even those past the deadline, were accepted in the original article's calculations. Two types of errors are possible in this design: Trials in which a neutral word is judged as morally wrong represent false alarms; trials in which a morally wrong word is judged as neutral represent misses. I copied these error rates for Studies 1-4 (Tables 1, 3, 5, and 7) into summary table 5.1 below, using the neutral-prime data only. (The authors' use of primes is not of relevance here.) The original article did not report reaction times (how long it took to make moral wrongness judgments), but Daryl Cameron generously provided these data, and I included relevant summary statistics in table 5.1: mean RTs; how many people showed RTs of below 500 ms (averaged across 120 trials); and the estimated interval of 95% of participants' RTs (M +/- 1.96*SD). I grouped the studies that used 400-500 ms response deadlines together and aggregated them via an unweighted average, and contrasted them with one condition in Study 1 in which people had a 800 ms response deadline.

Results. When the response deadline was between 400 and 500 ms, almost 80% of people made wrongness judgments in under 500 ms on average. At this speed, however, they showed average false-alarm rates of 32% and misses of 30%. When the response deadline was extended to 800 ms, people were still able to make wrongness judgments in 555 ms on average, now with reduced errors of 10% false alarms and 12% misses.

Table 5.1 Summary of error rates and reaction times for labeling action descriptions as "morally wrong" across four studies

	Response deadline (in ms)	Reaction Times				Error Rates	
		Mean RT (in ms)	Percent participants with RT < 500 ms	Estimated 95% interval of RTs (ms)		False alarms	Misses
Detailed breakdown							
Study 1.1	400	439	83%	195	682	34%	34%
Study 1.2	500	450	65%	173	727	32%	29%
Study 2	500	430	82%	163	698	33%	27%
Study 3	500	404	83%	142	665	39%	23%
Study 4	450	428	83%	178	678	24%	39%
Unweighted average		**430**	**79%**	**179**	**679**	**32%**	**30%**
Study 1.3	800	555	23%	394	717	10%	12%

Note. RT = reaction time for judging words denoting actions as "morally wrong" or not, averaged for each person across multiple trials and blocks
Study 1.1-1.3 correspond to subsamples in Study 1 that varied in response deadlines of 400, 500, and 800 ms, respectively

Appendix 6. Comparison of present-tense and past-tense uses of classes of moral judgments

Method: I examined three classes of moral judgments: badness, norm, and wrongness judgments. Because of the diversity of norm judgments I selected two prescription terms (required, mandatory), two prohibition terms (prohibited, forbidden), and two permission terms (permissible, acceptable). To capture the temporal focus of these moral judgments (i.e., declaring that something is vs. was bad, is vs. was permissible, is vs. was wrong) I searched the 1-billion Corpus of Contemporary American English (COCA) (as of 5/11/20) for instances of each moral judgment term that followed forms of the verb *be* — e.g., this *was* bad, it*'s* forbidden, this must *be* wrong, it might've *been* acceptable. I tabulated the raw frequencies of all relevant forms (see Table 5.2), broken down by present vs. past. I also separately tabulated singular (rather than plural) forms to ensure that any patterns hold across grammatical number. I then computed the percent of present focus out of total uses and the percent of past focus out of total use, but because these percentages differ as a function of other dominant uses of the terms I computed a present over past ratio to make comparisons possible.

Results: Four main results emerge: (1) All examined moral judgment terms are two to five times more frequently used in present tense than in past tense. (2) Norm judgments have the highest present/past ratio (4.2 on average), but as one exception, *forbidden* has a surprisingly large number of past-tense uses. (3) Wrongness judgments have a lower ratio (2.8) and badness judgments are in the middle (3.5). (All of these differences are significant.) (4) Instances of the term *wrong* include a high number of cases of "what's wrong" and "you are wrong," which barely ever occur with other moral judgment terms. If one considers these unique phrases that distort the comparison, an adjusted present/past ratio for wrongness can be calculated (Table 5.3), which is substantially lower, namely 1.6. Similar adjustments for *bad* are miniscule, and norm judgments are hardly ever paired with these terms.

Table 6.1 Summary of present-focused and past-focused uses of different classes of moral judgment

	Evaluation	Norm judgments						Wrongness
Tense	*bad*	*required*	*mandatory*	*prohibited*	*forbidden*	*permissible*	*acceptable*	*wrong*
Present	23,296	29,203	1,444	2,393	1,852	799	5,326	60,034
Past	6,616	6,425	262	626	1,139	175	991	21,189
Total	**286,811**	**90,302**	**10,809**	**7,161**	**7,823**	**2,263**	**20,649**	**233,420**
Present % of total	8.1%	32.3%	13.4%	33.4%	23.7%	35.3%	25.8%	25.7%
Past % of total	2.3%	7.1%	2.4%	8.7%	14.6%	7.7%	4.8%	9.1%
Present/past ratio	***3.52***	***4.55***	***5.51***	***3.82***	***1.63***	***4.57***	***5.37***	***2.83****
Present-singular only	15,370	12,284	864	995	988	527	3,242	43,558
Past-singular only	4,550	3,182	172	269	564	135	681	16,182
Singular p/p ratio	*3.38*	*3.86*	*5.02*	*3.70*	*1.75*	*3.90*	*4.76*	*2.69*

* See Table 5.3 for a calculation of this ratio without the phrases "what's wrong" and "you're wrong," highly frequent but uniquely associated with *wrong*

Table 6.2 COCA counts of present tense and past tense forms of *be* + moral judgment target adjective, as declarations and negations

Declarations

present	is bad	7433	is required	10098	is mandatory	598	is prohibited	891	is forbidden	802	is permissible	400	is acceptable	1860	is wrong	17061
	's bad	5524	's required	675	's mandatory	103	's prohibited	34	's forbidden	141	's permissible	58	's acceptable	391	's wrong	25967
	be bad	3021	be required	7491	be mandatory	299	be prohibited	578	be forbidden	217	be permissible	132	be acceptable	1192	be wrong	6905
	are bad	3510	are required	7285	are mandatory	205	are prohibited	705	are forbidden	588	are permissible	122	are acceptable	579	are wrong	4285
	're bad	736	're required	297	're mandatory	5	're prohibited	26	're forbidden	42			're acceptable	11	're wrong	4893
past	*was bad*	*3797*	*was required*	*2848*	*was mandatory*	*148*	*was prohibited*	*252*	*was forbidden*	*552*	*was permissible*	*121*	*was acceptable*	*525*	*was wrong*	*15925*
	were bad	*1033*	*were required*	*2453*	*were mandatory*	*60*	*were prohibited*	*243*	*were forbidden*	*411*	*were permissible*	*28*	*were acceptable*	*165*	*were wrong*	*3328*
	been bad	*747*	*been required*	*566*	*been mandatory*	*25*	*been prohibited*	*100*	*been forbidden*	*162*	*been permissible*	*11*	*been acceptable*	*105*	*been wrong*	*1569*
present total		20224		25846		1210		2234		1790		712		4033		59111
past total		*5577*		*5867*		*233*		*595*		*1125*		*160*		*795*		*20822*
present-singular only		12957		10773		701		925		943		458		2251		43028
past-singular only		*3797*		*2848*		*148*		*252*		*552*		*121*		*525*		*15925*
Present/past ratio		3.63		4.41		5.19		3.75		1.59		4.45		5.07		2.84
Singular p/p ratio		3.41		3.78		4.74		3.67		1.71		3.79		4.29		2.70

Negations

present	's not bad	1447	is not required	1272	is not mandatory	112	is not prohibited	63	is not forbidden	26	is not permissible	61	is not acceptable	679	is not wrong	195
	is not bad	389	's not required	95	's not mandatory	31	's not prohibited	2	's not forbidden	10	's not permissible	4	's not acceptable	229	's not wrong	221
	is n't bad	577	is n't required	144	is n't mandatory	20	is n't prohibited	5	are not forbidden	9	is n't permissible	4	is n't acceptable	83	is n't wrong	114
	're not bad	188	are not required	1087	are not mandatory	41	are not prohibited	30	is n't forbidden	6	are not permissible	7	are not acceptable	148	are not wrong	79
	are not bad	160	are n't required	198	are n't mandatory	10	are n't prohibited	5	are n't	1			are n't acceptable	25	are n't wrong	46
	are n't bad	238	're not required	92	're not mandatory	2	're not prohibited	2					're not acceptable	6	're not wrong	169
	not be bad	71	not be required	469	not be	17	not be	52	not be	10	not be	10	not be	122	not be	99
	be not bad	2			be not	1						1	be not	1		
past	*was not bad*	*115*	*was not required*	*243*	*was not mandatory*	*15*	*was not prohibited*	*14*	*was not forbidden*	*11*	*was not permissible*	*12*	*was not acceptable*	*121*	*was not wrong*	*90*
	was n't bad	*638*	*was n't required*	*91*	*was n't mandatory*	*9*	*was n't prohibited*	*3*	*was n't forbidden*	*1*	*were not*	*2*	*was n't acceptable*	*35*	*was n't wrong*	*167*
	were not bad	*51*	*were not*	*202*	*were not*	*5*	*were not*	*12*	*were n't forbidden*	*2*	*not have been*	*1*	*were not*	*39*	*were not wrong*	*29*
	were n't bad	*231*	*not have been*	*9*			*not have been*	*2*					*have not been*	*1*	*were n't wrong*	*65*
	not have been	*2*	*have not been*	*13*											*not have been*	*10*
	have not been	*2*													*have not been*	*6*
present total		3072		3357		234		159		62		87		1293		923
past total		*1039*		*558*		*29*		*31*		*14*		*15*		*196*		*367*
present-singular only		2413		1511		163		70		45		69		991		530
past-singular only		*753*		*334*		*24*		*17*		*12*		*14*		*156*		*257*
Present/past ratio		2.96		6.02		8.07		5.13		4.43		5.80		6.60		2.51
Singular p/p ratio		3.20		4.52		6.79		4.12		3.75		4.93		6.35		2.06

Table 6.3 Reanalysis of *wrong* without the phrases "what's wrong" and "you're wrong"

Full set		Subset		Remainder	Adjusted present/past comparison		By comparison	
is wrong	17061	what is wrong	3923	13138			what is bad	137
's wrong	25967	what's wrong	20159	5808	Adjusted present	30,058	what's bad	152
be wrong	6905			6905	Adjusted past	18,285		
are wrong	4285	you are wrong	1249	3036			you are bad	101
're wrong	4893	you're wrong	3722	1171	Adjusted present % of total	12.9%	you're bad	286
was wrong	*15925*	what was wrong	1765	14160	Adjusted past % of total	7.8%	what was bad	27
were wrong	*3328*	you were wrong	772	2556	Adjusted present/past ratio	***1.64***	you were bad	47
been wrong	*1569*			1569				

Appendix 7. Recommendations for measuring the moral judgments we intend to measure

1 To test hypotheses about the differences among classes of moral judgments, a between-subjects approach (e.g., one group answering badness questions, another group answering blame questions) will often be best. To compare these judgment conditions, however, variation across stimuli must be created (e.g., violations varying in severity and in intentionality/unintentionality) so that the different sensitivities of the moral judgments can be compared (rather than merely mean differences, which are rarely of interest).

2 When a within-subject approach is adopted and different questions are posed to tap into different moral judgments (e.g., badness and blame), we must prevent participants from collapsing the different questions into one judgment. We might (a) intermingle the target questions with a few other, nonmoral questions; (b) explain to participants why we ask the different questions and meaningfully mark their difference; (c) ask people before the task to reflect on what the judgments mean to them (e.g., "When you [judge something as bad]/[blame someone for what they did], how do you do that?" or "…what does this mean to you?").

3 Another approach to ensure differentiation among judgments is to train participants on the different judgments with examples, introduce distinct cue words, and present the cues (hence judgment questions) multiple times, in randomized order. With a sufficient sample of stimuli, each probe can be presented multiple times while it remains unpredictable which judgment succeeds any given stimulus (Malle & Holbrook 2012; Smith & Miller 1983).

4 According to the presented framework of moral judgments, some moral judgments have preferred objects of judgment. For example, questions about permissibility are most meaningful when asked about intentional actions before they happen, while questions about moral wrongness are most meaningful when asked about intentional actions after they happened. When participants are asked about the permissibility of unintentional violations (e.g., accidents), they might be confused, and the data may reflect different judgments from the ones that researchers asked for. Participants will rarely volunteer their confusion, so we need to find out how people conceptualize the scenarios we present them with and what questions they most naturally ask themselves. The next few recommendations are all targeted at this issue and may also have some ancillary benefits.

5 Pretest what possible interpretations or questions people might have about a scenario, by asking them open-ended questions such as "what's going on here?" "What's not right here?" "What are you wondering about?" Good dependent variables capture the observations, inferences, and questions people themselves have about a scenario. If there is a mismatch, and the dependent variable doesn't match people's own conceptualization, then the data a very difficult to interpret. For example, in Guglielmo and Malle (2010) we have proposed, and provided evidence for the proposal, that the well-known "side-effect effect" (Knobe, 2003) forces people to make an intentionality judgment (and they are willing to respond to the question) but that this is not the primary way people conceptualize the event (see also Laurent et al., 2015, 2019). When participants are offered a wider range of judgments, it appears that intentionality is not what they are most concerned with.

6 Allowing participants to opt out of a question may be helpful in diagnosing possible mismatches between the researcher's intended question and people's own conceptualization. Opting out must be costly for participants so that they go that route only if they really have a problem with the question. For example, the opt-out could be "This question does not make sense to me because: _____," requiring a free-response answer).

7 We have learned a lot from asking people to explain their judgments after they made them. When they don't have access to the basis of their judgments, their answers will show that (e.g., short, uninformative, low consensus). If they do have access, however, their answers will show *that* (e.g., elaborate responses, systematic themes, high consensus).

8 Sometimes researchers use ambiguous terms in their measures and assume people interpret them in similar ways. That should be demonstrated. For example, what do people think about when we ask how much "punishment" a person deserves? Do they think of social or legal punishment? Punishment as inflicting pain or as teaching a lesson? It may be useful to ask people to define their understanding of key concepts such as punishment at the end of the study (e.g., "When you made your judgments, what kind of 'punishment' were you imagining the person deserved?"). Alternatively, one might provide definitions of concepts at the beginning of the study, to shape people's interpretations into a more consensual pattern (e.g., "By 'punishment' we mean fines or incarceration" or "…we mean social acts that hurt the other person").

References

Laurent, S. M., Clark, B. A. M., & Schweitzer, K. A. (2015). Why side-effect outcomes do not affect intuitions about intentional actions: Properly shifting the focus from intentional outcomes back to intentional actions. *Journal of Personality and Social Psychology, 108* (1), 18–36. https://doi.org/10.1037/pspa0000011

Laurent, S. M., Reich, B. J., & Skorinko, J. L. M. (2019). Reconstructing the side-effect effect: A new way of understanding how moral considerations drive intentionality asymmetries. *Journal of Experimental Psychology: General, 148* (10), 1747–1766. https://doi.org/10.1037/xge0000554

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* (12), 1635–1647. https://doi.org/10.1177/0146167210386733

Malle, B. F., & Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology, 102* (4), 661–684. https://doi.org/10.1037/a0026790

Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology* , *44* , 492–505.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis, 63* (3), 190–194. https://doi.org/10.1093/analys/63.3.190

Appendix 8. Aggregated Results from McHugh et al (2017)

McHugh, C., McGann, M., Igou, E. R., & Kinsella, E. L. (2017). Searching for moral dumbfounding: Identifying measurable indicators of moral dumbfounding. *Collabra: Psychology, 3(1, Art. 23)* , 1–24. https://doi.org/10.1525/collabra.79

Table 8.1 Average percentage of dumbfounded responses and reasons responses across four dilemmas

		Heinz	Cannibal	Incest	Trolley	**Total**
		Average percent	Average percent	Average percent	Average percent	**Average percent**
Percent out of total responses	Dumbfounded	14.4	26.5	33.9	19.5	**23.6**
	Reasons	65.6	62.1	40.0	57.5	**56.3**
Out of dumbfounded and reasons only	Dumbfounded	18%	34%	48%	27%	**32%**
	Reasons	82%	66%	52%	73%	**68%**

Table 8.2 Counts and percentages of all responses (dumbfounded, reasons, and others) across studies*

		Heinz		Cannibal		Incest		Trolley	
Study	Category	N	percent	N	percent	N	percent	N	percent
Study 1	Nothing wrong	6	19.4	8	25.8	11	35.5	8	25.8
	Dumbfounded	0	0.0	11	35.5	18	58.1	3	9.7
	(admissions)	0	0.0	8	25.8	10	32.3	3	9.7
	(declarations)	0	0.0	3	9.7	8	25.8	0	0.0
	Reasons	25	80.7	12	38.7	2	6.5	20	64.5
Study 3a (critical slide)	Nothing wrong	14	19.4	4	5.6	12	16.7	15	20.8
	Dumbfounded	13	18.1	14	19.4	18	25.0	14	19.4
	Reasons	45	62.5	54	75.0	42	58.3	43	59.7
Study 3a (coded)	Nothing wrong	14	19.4	4	5.6	12	16.7	15	20.8
	Dumbfounded	19	26.4	21	29.2	31	43.1	22	30.6
	Reasons	39	54.2	47	65.3	29	40.3	35	48.6
Study 3b (critical slide)	Nothing wrong	21	20.8	10	9.9	31	30.7	24	23.8
	Dumbfounded	12	11.9	19	18.8	16	15.8	16	15.8
	Reasons	68	67.3	72	71.3	54	53.5	61	60.4
Study 3b (coded)	Nothing wrong	21	20.8	10	9.9	31	30.7	24	23.8
	Dumbfounded	16	15.8	30	29.7	28	27.7	22	21.8
	Reasons	64	63.4	61	60.4	42	41.6	55	54.5

* Study 2 is omitted because, by authors' own admission, it was flawed. Results are displayed in Table 7.3 belov

Table 8.3 Results of omitted Study 2

	Heinz		Cannibal		Incest		Trolley	
Category	N	percent	N	percent	N	percent	N	percent
Nothing wrong	8	11.1%	4	5.6%	2	2.8%	10	13.9%
Dumbfounded	45	62.5%	46	63.9%	54	75.0%	45	62.5%
Reasons	19	26.4%	22	30.6%	16	22.2%	17	23.6%