Running Header: THE AFFECTIVE HARM ACCOUNT (AHA) OF MORAL JUDGMENT

The Affective Harm Account (AHA) of Moral Judgment: Reconciling Cognition and Affect, Dyadic Morality and Disgust, Harm and Purity

In press, Journal of Personality and Social Psychology

Kurt Gray[1], Jennifer K. MacCormack[2,3], Teague Henry[3], Emmie Banks[4], Chelsea Schein[5], Emma Armstrong-Carter[6], Samantha Abrams[1], Keely A. Muscatell[1,7]

[1] Department of Psychology and Neuroscience
University of North Carolina at Chapel Hill

[2] Department of Psychiatry
University of Pittsburgh

[3] Department of Psychology
University of Virginia

[4] Department of Neuroscience
Emory University

[5] Legal Studies and Business Ethics
The Wharton School

[6] Graduate School of Education
Stanford University

[7] Lineberger Comprehensive Cancer Center
University of North Carolina at Chapel Hill

Contact information:
Kurt Gray
CB 3270
Department of Psychology and Neuroscience
University of North Carolina at Chapel Hill
kurtjgray@gmail.com

Abstract

Moral psychology has long debated whether moral judgment is rooted in harm vs. affect. We reconcile this debate with the Affective Harm Account (AHA) of moral judgment. The AHA understands harm as an intuitive perception (i.e., perceived harm), and divides "affect" into two: embodied visceral arousal (i.e., gut feelings) and stimulus-directed affective appraisals (e.g., ratings of disgustingness). The AHA was tested in a randomized, double-blind pharmacological experiment with healthy young adults judging the immorality, harmfulness, and disgustingness of everyday moral scenarios (e.g., lying) and unusual purity scenarios (e.g., sex with a corpse) after receiving either a placebo or the beta-blocker propranolol (a drug that dampens visceral arousal). Results confirmed the three key hypotheses of the AHA. First, perceived harm and affective appraisals are neither competing nor independent but intertwined. Second, although both perceived harm and affective appraisals predict moral judgment, perceived harm is consistently relevant across all scenarios (in line with the Theory of Dyadic Morality), whereas affective appraisals are especially relevant in unusual purity scenarios (in line with affect-as-information theory). Third, the "gut feelings" of visceral arousal are not as important to morality as often believed. Dampening visceral arousal (via propranolol) did not directly impact moral judgment, but instead changed the relative contribution of affective appraisals to moral judgment—and only in unusual purity scenarios. By embracing a constructionist view of the mind that blurs traditional dichotomies, the AHA reconciles historic harm-centric and current affect-centric theories, parsimoniously explaining judgment differences across various moral scenarios without requiring any "moral foundations."

Abstract word count: 250

Keywords: morality, affect, purity, propranolol, disgust, harm

An Affective Harm Account of Moral Judgment: Reconciling Cognition and Affect, Dyadic Morality and Disgust, Harm and Purity

Everyone agrees that people make moral judgments; not everyone agrees about their psychological basis. Early theories emphasized the role of reasoning and harm in morality, arguing that affect was irrelevant to authentic moral judgment (Kohlberg, 1981). These theories grounded moral judgments in explicit judgments of harm (Turiel, 1983), with more harm leading to more moral condemnation. Later theories dismissed the role of reasoning and instead emphasized the role of affect—particularly visceral "gut feelings" (Haidt & Hersh, 2001)—with more negative feelings leading to more moral condemnation (Haidt, 2012). This "affective" approach is now the dominant paradigm in moral psychology, but doubts remain. First, meta-analyses (Landy & Goodwin, 2015) and inconsistent replication attempts (Johnson et al., 2016; van Dijk et al., 2018) raise questions about the causal role of affect in moral judgment. Second, evidence supports the continued importance of harm in moral judgment (Decety & Cowell, 2017), especially the perceived harm proposed by the Theory of Dyadic Morality (Schein & Gray, 2018). Here we provide a new answer to the question of what matters most in moral judgment— affect or harm? Reconciling past debates, we suggest that the answer is "both."

This paper introduces the Affective Harm Account (AHA; pronounced as "ah-ha!") of moral judgment. It argues that harm and affect do not compete but are instead intertwined. The AHA clarifies the nature of "harm" in moral judgment as *perceived harm*—intuitive appraisals of harmfulness—consistent with past work (Schein & Gray, 2018). The AHA also clarifies the nature of "affect" in moral judgment by distinguishing affective appraisals (i.e., the perceived affective qualities of a scenario) from general visceral arousal (i.e., the "gut feelings" of core physiological arousal). Past work has emphasized the importance of "gut feelings" in morality, but we suggest that visceral arousal may not directly impact moral judgments, instead fueling affective appraisals: stronger visceral "gut feelings" give affective appraisals more moral weight.

The AHA also sheds light on *when* affect is most important in moral judgment: when scenarios are novel or unusual, such as often-used "purity" scenarios (e.g., sex with a dead chicken).

We test these key predictions of the AHA in a randomized, double-blind, placebo-controlled pharmacological experiment, and speak to debates about cognition vs. affect, dyadic morality vs. gut feelings, and harm vs. purity.

A Brief History of Harm vs. Affect in Moral Psychology

The field's understanding of moral judgment—and harm, affect, and purity—has evolved (see **Table 1**). Morality was long thought to be rooted in explicit reasoning about harm (Kohlberg, 1969). In line with many philosophical traditions, harm was understood as a binary and objective feature of the situation—physical/emotional suffering is either present or absent, and people can objectively determine its presence or absence through reasoned deliberation. Turiel's (1983) Domain Theory emphasized that people referenced the harmfulness of a situation when reasoning about violations of the moral domain (e.g., murder) but not for violations of social convention (e.g., wearing pajamas to school). The centrality of reason was soon challenged by multiple trends in psychology, especially the "affective revolution" which revealed that many judgments are based on feelings (Zajonc, 1980). Research suggested that self-reported reasoning can be an unreliable guide to psychological processes (Nisbett & Wilson, 1977). Reasoned harm models of morality were also criticized for studying (and elevating) the moral decisions of white Western men (e.g., Gilligan, 1993), especially in light of work revealing powerful cross-cultural differences across many domains (Markus & Kitayama, 2003), including morality.

The Big Three (Shweder et al., 1997) cemented the importance of culture in morality by using anthropological field work to argue for moral pluralism—that morality contained multitudes—and extended beyond strict concerns of objective harm to three themes of "ethical discourse:" autonomy, community, and divinity. Autonomy focused on rights and harm, community focused on duty, hierarchy, and interdependence, and divinity focused on sanctity

and purity. The Big Three expanded the scope of moral concerns, but still endorsed the value of explicit explanations of morality (i.e., discourse), similar to models of reasoned harm.

The Big Three argued for a role for affect, especially in moral acquisition, and although it is often painted as dismissive of harm (Haidt, 2012), it explicitly argued for the importance of harm, with each of autonomy, community and divinity tied to a different kind of suffering. That is, while advocating for moral pluralism, the Big Three also advocated for "harm pluralism," the idea that harm—ubiquitous across moral discourse—was culturally constructed, with its exact understanding tailored to the value being violated. For example, a violation of community might harm group solidarity, and a violation of divinity might harm immortal souls. This represents a major break from previous psychological conceptions of harm; instead of being *objectively* present or absent, the Big Three recognized that the harmfulness of a situation is culturally *subjective* (Shweder et al., 1997).

The Social Intuitionist Model (SIM; Haidt, 2001) embraced pluralism and affect, but rejected the longstanding importance of harm. This theory argued that moral judgments are driven by intuitive—and embodied—affective feelings of good or bad, and that references to harm are largely post-hoc confabulations used by participants to justify their (affect-based) moral judgment. While the Big Three examined everyday moral violations, the Social Intuitionist Model focused on a subset of unusual purity violations (e.g., sex with a dead chicken, loving consensual incest).

Table 1. Morality Judgment, Harm, Affect, and Purity across Theories of Moral Judgment

Theory and Moral Stance	Harm	Affect	Purity
Moral vs. Conventional Domains Morality and convention are separate domains of judgment (Turiel et al., 1987).	Harm is what separates morality from convention, measured via reasoning. It is objective present or absent.	Affect is not involved in moral judgment.	Purity is unconsidered.
The Big Three Moral judgments extend beyond Western conceptions of rights (Shweder et al., 1997; Rozin et al., 1999).	Harm is especially tied to one of three ethics (i.e., autonomy), but is also understood to vary culturally (as suffering).	Affect helps drive different types of moral judgment and is critical for moral acquisition.	Purity is a key theme of everyday ethical discourse, especially in non-Western cultures.
Social Intuitionist Model (SIM) Moral judgments reflect cultural milieu and are driven by intuitive gut feelings (Haidt, 2001).	Harm rationalizes moral judgments post-hoc via explicit reasoning. It is objectively present or absent (as defined by experimenters). Harmless wrongs exist.	Affect is intuitive gut feelings (e.g., flashes of bad) that drive all moral judgments.	Purity is studied via unusual acts, "proves" that harm does not drive moral judgment.
Moral Foundations Theory (MFT) Moral judgment is grounded in five separate innate, but culturally activated, modular mechanisms (Graham et al., 2009; Haidt, 2012)	Harm is just 1 of 5 moral modules. It can only triggered by the presence of "objective harm" (experimenter-defined). Harmless wrongs exist.	Affect drives moral judgment; specific emotions drive specific judgments.	Purity is studied via unusual acts, is specifically tied to disgust, and is fundamentally distinct from harm.
Theory of Dyadic Morality (TDM) Moral judgment relies on a template of perceived harm (Gray et al., 2012a, 2012b; Schein & Gray, 2018).	Harm is dyadic (tied to a template of agent-patient). It is subjective—varying by person/culture/context— and is intuitively perceived along a continuum. No real harmless wrongs.	Affect is feelings of bad that can—along with norm violations—amplify moral judgment.	Purity is an important value; violations judged as immoral based on their perceived harm.
Affective Harm Account (AHA) Moral judgment relies jointly on perceived harm and affective appraisals, which are intertwined (current paper).	Harm is a subjective, dyadic (template of agent-patient). It is subjective—varying by person/culture/context—and is intuitively perceived along a continuum. No real harmless wrongs. It is also tied to affect.	Affect involves: 1) visceral signals that can contribute to 2) affective appraisals, helping inform moral judgment.	Purity is judged as wrong based on perceived harm and—especially for unusual purity scenarios—by affective appraisals.

Also, while the Big Three acknowledged the validity of participants' judgments of both morality and harm, the Social Intuitionist Model dismissed the validity of harm judgments. In one famous (but unpublished) study, participants read a vignette about a brother (Mark) and sister (Julie) who had consensual incest, apparently without negative consequences (Haidt et al., 2000). Participants generally said that Mark and Julie's actions were immoral and, when asked to explain why, typically invoked harm, saying that incest led to deformed children and ruined relationships. However, experimenters disqualified these statements of harm because they had designed the specific case of Mark and Julie to be objectively harmless. Without being able to reference harm, participants could not the articulate reasons for their moral judgment (they were rendered "dumb"), and so Haidt and colleagues (2000) concluded that they revealed the phenomenon of "moral dumbfounding."

Moral dumbfounding shifted the field's view of moral judgment. Because harm was "objectively" absent from consensual incest and other unusual purity violations (e.g., sex with dead chickens), psychologists reasoned that there must be a non-harm-based driver of moral judgment—negative affect (Haidt, 2001). Spurred on by this logic, researchers argued broadly for the importance of visceral feelings over judgments of harm in morality (Haidt & Hersh, 2001)—especially given that inducing disgust seemed to increase moral condemnation (Schnall et al., 2008).

Moral Foundations Theory (MFT; Graham et al., 2009) was the next evolution of the Social Intuitionist Model. It maintained a focus on intuitions, cultural pluralism, and unusual purity violations, but also incorporated a modular view of the mind from evolutionary psychology (Haidt & Joseph, 2004)—which suggests that adaptive challenges are met via organisms possessing various psychological "switches in the brain" (Haidt, 2012, p. 123). Moral Foundations Theory posits five of these distinct moral switches—harm, fairness, loyalty, authority, and purity.

Moral Foundations Theory makes many claims, but here we focus on two key claims related to affect and harm. First, it claims that all moral judgment is driven by "affective reactions of liking or disliking" (Haidt & Graham, 2007, p. 104)—especially purity-related judgments, which are argued to be grounded in visceral feelings of disgust. Second, Moral Foundations Theory claims that the importance of harm in morality is extremely limited. Harm is only one of five moral concerns and—in line with the Social Intuitionist Model's conception of harm—can only be "triggered" by the presence of objective physical or emotional suffering (as determined by researchers). Harm is argued to be especially absent and irrelevant in moral judgments of usual purity scenarios. The irrelevance of harm to purity judgments and the separation between harm and affect are essential lynchpins in modern formulations of Moral Foundations Theory (Graham et al., 2018).

This progression of theories in moral psychology would seem to argue strongly for the role of affect in moral judgment and against the role of harm, but doubts remain.

Affect and Not Harm?

Studies show that moral judgments are intuitive and vary across cultures, but the evidence that affect *per se* drives moral judgment is less clear. One review concludes: "current evidence is insufficient to support the hypothesis that emotional processes mediate our intuitive moral judgments, or that our moral concepts are emotionally constituted" (Huebner et al., 2009, p. 5)—largely because most studies reveal correlations but not causation between affect and moral judgment. Some studies seem to reveal a causal connection between affect and moral judgment by experimentally inducing disgust to increase moral condemnation (Strohminger & Kumar, 2018; but see Tracy et al., 2019, which we later discuss), yet the robustness of these effects is disputed. A meta-analysis revealed no clear causal link between inducing incidental disgust and the amplification of moral judgment (Landy & Goodwin, 2015; but see Schnall et al., 2015), and a high-powered replication showed similar null effects (Johnson et al., 2016). However, subsequent work has revealed evidence in favor of the causal role of disgust in moral

condemnation, so long as appropriate moderators (e.g., disgust sensitivity, private body consciousness) are measured and appropriate outcomes are assessed (van Dijk et al., 2018).

Not only has the central causal role of affect come under question, but recent research also suggests that harm—at least intuitive perceptions of harm—is more important than is currently appreciated. For example, work on moral grammar finds that the mind quickly processes harm-based elements to form intuitive moral judgments (Mikhail, 2007), and the intuitiveness of harm is underscored by evidence that the brain processes harm in moral violations in under 200 *ms* (Decety & Cacioppo, 2012).

The Theory of Dyadic Morality (TDM) redefines the importance of harm in moral judgment. It argues that harm is not a reasoned calculation but instead an intuitive perception that varies across cultures, just like morality (Schein & Gray, 2018). This theory is called "dyadic morality" because it defines harm as a cognitive template (or schema) involving two minds—a dyad of an intentional agent causing damage to a vulnerable patient (iA $^d\rightarrow$ vP). Dyadic Morality draws from other categorization theories (Murphy, 2004) to suggest that moral judgment (how much does *x* act belong to the category of "immoral?") proceeds by comparing acts to a cognitive template of dyadic harm. The closer the match—i.e., the more an act seems to be dyadic/harmful—the stronger the moral condemnation, explaining why obviously dyadic acts like child abuse seem more immoral (intentional agent, causing damage, vulnerable patient) than accidents (less agency), attempted harms (less damage), or graffiti on concrete (less vulnerable patient).

In contrast to past theories, Dyadic Morality argues that harm and immorality are not binary (i.e., present or absent), but instead continuums (see top panel of **Figure 1**). Harm is not either absent (harmless) or present (harmful), and moral judgments are not "triggered" on or off. Instead, the subjective experiences of both harm and immorality vary in degree, ranging from low to medium to high, and the relative amounts of perceived harm predict the relative amounts of perceived immorality. People clearly view morality as a continuum, judging murder

as more immoral than theft, theft as more immoral than engaging in weird sexual fetishes, and engaging in weird sexual fetishes as more immoral than jogging. Dyadic Morality argues that underneath this continuum of immorality is a continuum of perceived harm (iA $^{d}\rightarrow$ vP): people view murder as more harmful than theft, theft as more harmful than engaging in weird sexual fetishes, and engaging in weird sexual fetishes as more harmful than jogging.

Dyadic Morality also emphasizes that harm is subjective. Whether people perceive harm cannot be independently determined by an external authority (e.g., an experimenter) based on the "objective" presence of physical injury or emotional trauma within a situation (Graham et al., 2013). Instead, harm is something perceived and appraised by each of us within our minds, based on how much intention (Cushman, 2008, 2015; Decety & Cacioppo, 2012; Hesse et al., 2016; Malle & Guglielmo, 2011), causation (Le Guen et al., 2015; Moore et al., 2013; Schlottmann et al., 2002), and suffering (Decety & Cowell, 2018; Fan et al., 2011; Han et al., 2020; Lamm et al., 2011) we infer. People can therefore still perceive harm even when an act is ostensibly—or explicitly argued to be—harmless, as in the case of incest between Mark and Julie (Royzman et al., 2015).

It may seem strange to moral psychologists that harm is a matter of perception because moral psychology has roots in moral philosophy, which emphases the normative and objective features of situations. But the perceived nature of harm is perfectly consistent with social psychology, which reveals how our external social world is filtered through the lens of our own assumptions, cultures, and identities (Henrich et al., 2010; Markus & Kitayama, 2003; Triandis et al., 1988). Regardless of the ontological status of harm as objective in some absolute sense, harm is necessarily subjective when its presence is assessed by social perceivers. Just as two people can see the nature of their relationship very differently (e.g., we are getting married vs. we are casually dating; Schindler et al., 2010), or the mind of a dog differently (e.g., it is a socially intelligent companion vs. a mindless furry robots; Epley & Waytz, 2009; Wegner & Gray, 2016), they can also see different amounts of harm in an act.

Importantly, research finds that people's perceptions of harm are not post-hoc spasms of reasoning (Gray et al., 2014). Instead, they are intuitive appraisals that authentically drive moral judgment (Schein & Gray, 2015). Consider again the case of Mark and Julie. Moral judgments of incest can be driven by harm because people have intuitive perceptions of harm that persist despite explicit arguments that incest is harmless. Just as people can still be afraid of flying even if they "explicitly know" it is safer than driving, people can still think that incest is harmful even if a scenario is designed to be harmless (Gendler, 2008). After all, in almost every real-world case, incest *is* harmful. It makes sense that our intuitions about both morality and harm are calibrated for the real-world, harm-based situations faced by ourselves and our evolutionary ancestors.

The perceptual nature of harm allows it to be pluralistic across values and cultures—to contain multitudes—explaining why people see harm in disloyalty, disrespect, and impurity (Ochoa, 2022; Schein & Gray, 2018). Consider when Brahmin Indians morally condemn the so-called "harmless act" of a son eating a chicken after a father's funeral (Shweder et al., 1997). A full anthropological appreciation of their worldview reveals that these people actually perceive much harm in this act—eating meat is thought to interrupt the cleansing of "death pollution," thereby condemning the father to eternal suffering (Shweder et al., 1997). Thus, a moral stance that seems to Westerners as being about "harmless impurity" is grounded in perceived harm.

Dyadic Morality argues that researchers need to better distinguish their own subjective perceptions of harm(lessness) from those of the people and cultures they study. Past theories like Moral Foundations are paternalistic, viewing Western academics—who are largely liberal and secular—as the ultimate authority for drawing a bright line between acts that are "objectively" harmful and those that are "objectively" harmless. In contrast, Dyadic Morality acknowledges the validity of perceived harm across people and cultures. For example, when evangelical Christians state that they oppose gay marriage because they see it as harming American families (Bryant, 1977), Dyadic Morality does not dismiss these perceptions of harm

as confabulated or invented. Instead, it accepts these intuitions of harm as legitimate, even if they can be hard for liberal academics to understand.

Of course, different perceptions of harm can be more or less widely shared, more or less controversial, and more or less connected to obvious physical pain. Dyadic Morality suggest that moral disagreement across issues and between cultures/politics is grounded in differences in harm perception (Schein & Gray, 2018). One group of people may believe an act causes obvious harm to a vulnerable patient (e.g., Pro-Life advocates thinking about abortion as killing babies), whereas another group may deny that the same act causes harm (e.g., Pro-Choice advocates thinking about abortion as the removal of mindless cells, Gray et al., 2012). Likewise, liberals may see it as good to challenge the police but see obvious harm in purchasing a truck with low-gas mileage (e.g., to the environment). Conversely, conservatives may see obvious harm in challenging the police (e.g., to the social order), but see it as good to buy a powerful American-made truck. The key point that Dyadic Morality makes is that *both* morality *and* harm are subjective judgments that are intuitive, pluralistic, and that vary across people.

Dyadic Morality also acknowledges that different moral values exist and are important. Loyalty is not the same as fairness, which is not the same as purity, and these differences help make sense of cultural differences and moral dilemmas (e.g. whistleblowing; Waytz et al., 2013). However, Dyadic Morality suggests that the dyadic template functions across different values as a common moral currency (Schein & Gray, 2015), and that the relative immorality of a moral violation is robustly predicted by its perceived harm (see top panel of **Figure 1**). Most succinctly, if you want to know how much someone will condemn an act, discover how much that person views that act as harmful. Research supports the idea that dyadic "template matching mediates moral cognition" (p. 13): across diverse acts, perceived harmfulness predicts the severity of moral judgment (Ochoa, 2022; see bottom panel of **Figure 1**).

Critics of Dyadic Morality often point to exceptions to the general predictive power of harm, namely "harmless wrongs" (e.g., sex with chickens; Haidt, 2001) and "permissible harms"

(e.g., honor killing, Fiske & Rai, 2014; or dumping your romantic partner, E. Royzman & Borislow, 2022). Importantly, the moral dyad is an *empirical psychological template,* not a rigid philosophical rule arguing for necessary and sufficient conditions (Sinnott-Armstrong, 2016). A template can still drive general judgments even if exceptions exist. We still say that mammals give live birth despite the platypus, and still say that birds fly despite the penguin. It is exciting to find exceptions to any general rule, but if 99% of morality is predicted by perceived harm, then Dyadic Morality is still a tenable psychological theory. And what of the other 1%? Upon closer inspection, exceptions may not even be exceptions. Because Dyadic Morality argues for a continuum of perceived harm, researchers testing it must allow participants to 1) rate the immorality and *perceived harm* of the apparent exception, and 2) rate various acts that are both more and less harmful/immoral than the apparent exception (to form *a continuum*). Studies critiquing Dyadic Morality seldom do either and never do both. One cannot claim to engage with a theory while ignoring its key claims.

Given the pervasive role of perceived harm in moral judgment as articulated by Dyadic Morality, how can moral psychology reconcile the fundamental importance of perceived harm with work demonstrating affective contributions to moral judgment? This is precisely the theoretical gap that the AHA seeks to bridge. Building on Dyadic Morality, we revisit and resituate the role of affect in moral judgment.
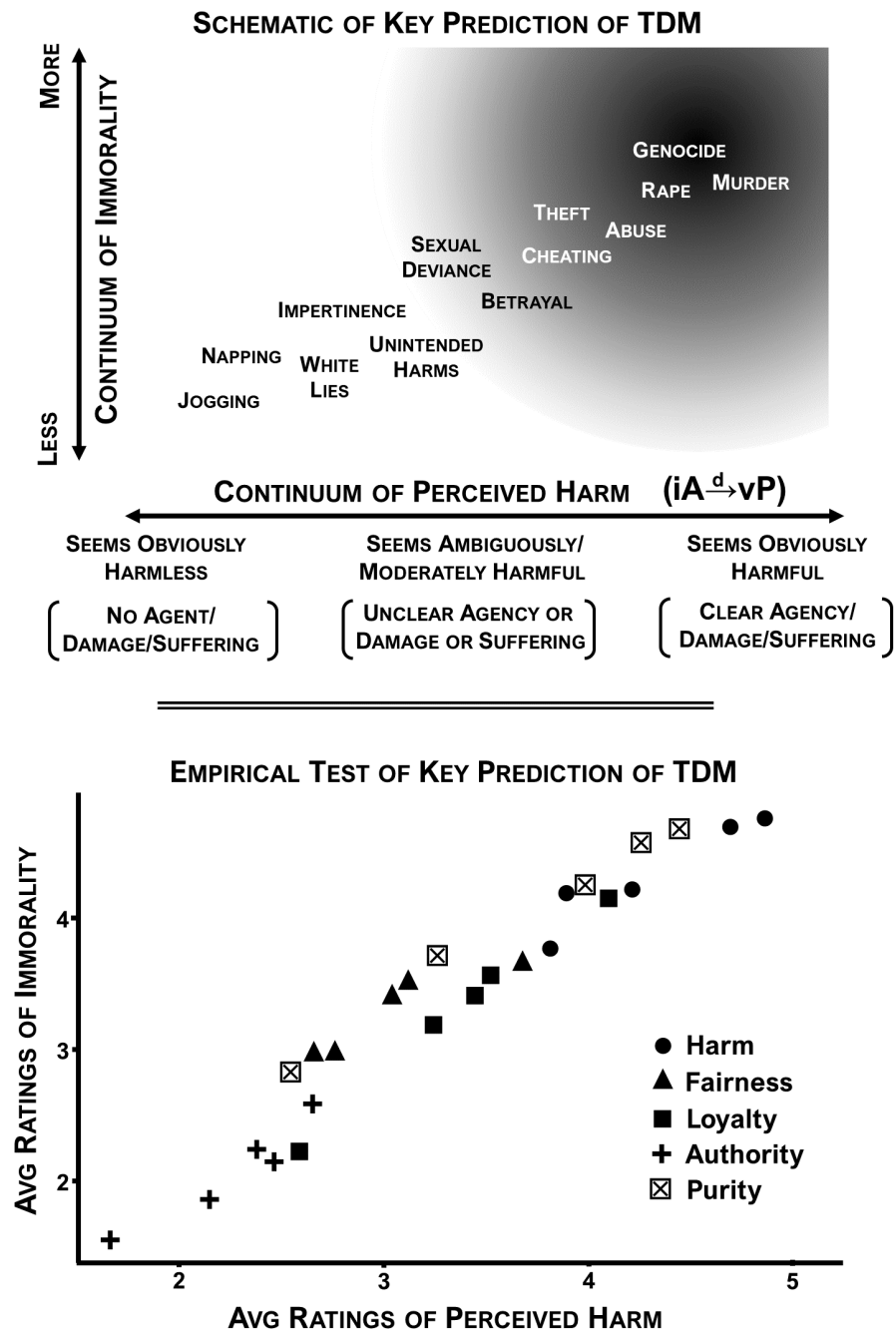
SCHEMATIC OF KEY PREDICTION OF TDM

EMPIRICAL TEST OF KEY PREDICTION OF TDM

Figure 1. Top Panel: Schematic of key prediction of the Theory of Dyadic Morality (Schein & Gray, 2018)—perceived harm should predict immorality across acts. Bottom Panel: Empirical test of key prediction of the Theory of Dyadic Morality—perceived harm does predict immorality across acts (Ochoa, 2022).

Deconstructing Affect in Moral Psychology

"Affect" is a simple word but means different things to different people, like "agency" and "mind." Moral psychology papers speak about affect (Cameron & Payne, 2011), affective reactions (Haidt & Hersh, 2001), gut feelings (Pizarro et al., 2003), and affectively-laden moral intuitions (Wheatley & Haidt, 2005), but do they mean subjective feelings of good or bad (i.e., valence), subjective feelings of arousal (i.e., activation), emotional states such as anger or disgust, more objective changes in physiological activation, ratings of the affective qualities of a stimulus, or affectively-tinged categorizations (Barrett & Bliss-Moreau, 2009; Satpute et al., 2019)? Drawing from diverse theories of emotion (Barrett, 2017; Moors et al., 2013; Storbeck & Clore, 2008), we suggest that "affect" in moral psychology often conflates (at least) two different concepts.[1] The first is "visceral arousal" and the second is "affective appraisals."

We use the term *visceral arousal* to refer to physiological states of activation that typically—but not necessarily—includes feelings of goodness or badness (i.e., valence). Visceral arousal is instantiated via the integration of afferent visceral and autonomic nervous system signals (e.g., sympathetic nervous system, cardiovascular and gastrointestinal systems, etc.) as well as signals from broader neurophysiological signaling pathways such as the dopaminergic and adrenergic/noradrenergic systems (Kleckner et al., 2017; Satpute et al., 2019). When moral psychology research discusses a direct causal link between "gut feelings" and judgment, or the role of "bodily reactions" (Schnall et al., 2008, p. 1097), or the importance of general arousal (Cheng et al., 2013), they are nodding toward visceral arousal. On the other hand, *affective appraisals* are judgments—typically intuitive—about the emotional qualities of a stimulus, and so include semantic or conceptual content. When people label a moral act as "disgusting" or

---

[1] As we note in the Introduction, affective elements and assessments are many and lie along a continuum from most "free floating and embodied" to most "cognitive and context-specific." Interested readers may find it helpful to refer to Clore & Huntsinger's (2007) classic paper discussing differences and relations between affect, affective appraisals, emotions, and moods.

"gross," they are making an affective appraisal (i.e., labeling or evaluating its affective content). When researchers ask participants to explicitly state whether an act is "*really disgusting*" or whether "acting like this is *really gross*" (Rottman & Kelemen, 2012, p. 357, italics in original), they are assessing affective appraisals.

Visceral arousal and affective appraisals are connected but usefully distinguished. Visceral arousal is more embodied and "free-floating"—and is likely tied in part to a person's broader mood state (e.g., feeling more energized vs. drained, tense vs. relaxed). On the other hand, affective appraisals are context-bound and directed toward a stimulus; in moral psychology, this stimulus is typically a scenario. Similarly, visceral arousal is better understood as a "core" embodied reaction (Barrett & Bliss-Moreau, 2009; Haidt et al., 1997) whereas affective appraisals are more "cognitive," because they are explicit responses to specific questions about specific stimuli.

Importantly, past studies that discuss "gut feelings" or "embodied reactions" typically do not measure visceral arousal but instead some form of self-reported emotion ("How much disgust do you feel?"), often within the context of a specific moral scenario. These emotion self-reports likely conflate people's scenario-specific affective appraisals ("How disgusting is this moral act?") and the relevant visceral arousal (e.g., racing heart, nausea) connected to those appraisals. The AHA argues that it is helpful to distinguish between individuals' more ambient visceral-based affect (e.g., arousal) vs. situation-specific affective appraisals (as per Clore & Ortony, 2013; MacCormack & Lindquist, 2017; Schachter & Singer, 1962).

After distinguishing between visceral arousal and affective appraisals, an existing link between affective appraisals and moral judgment now seems obvious, as both are context-bound, affectively-laden judgments about a specific scenario. Affective appraisals are "this is bad," whereas moral judgment is "this is morally bad." Affective appraisals concern the inherent properties of a stimulus (i.e., a scenario) and so they are "integral" (tied to the stimuli at hand) rather than "incidental" (simply layered on top). Work reveals the causal impact of integral

affective appraisals on moral judgment (Wiseneski & Skitka, 2017), but there is less support for the oft-repeated claim that moral judgment is driven by embodied gut feelings (Haidt, 2001).

Is there a *direct causal* connection from visceral arousal to moral judgment? It is difficult to test this claim. Past research on affect and morality typically induces an incidental emotion (e.g., disgust) and then examines whether this emotion amplifies moral judgments (Pizarro et al., 2011). There are two limitations of this approach. First, these emotional manipulations typically involve reading vignettes, watching movies, or recalling memories, which—as explicit scenario-based manipulations—manipulate affective appraisals. Second, even when affect inductions may succeed at altering visceral arousal, any link between gut feelings and the amplification of moral judgment cannot tell us whether affect drives moral judgment (for a concise exploration of this distinction, see Pizarro et al., 2011). You can make a bike go faster by blowing wind behind it, but this doesn't mean that bikes are wind-powered. Likewise, just because making people feel extra bad might make moral violations seem worse doesn't mean that typical moral judgments (i.e., those not made in a lab) are driven by affect.

To reveal whether visceral arousal has a direct causal impact on moral judgment, we cannot simply *increase* arousal (either directly through a physiological manipulation or indirectly through an emotion induction) and then test for an *increase* in moral condemnation. Instead, we need to *reduce* visceral arousal and see if this *decreases* moral condemnation (see Tracy et al., 2019). We do this here using a randomized, double-blind, placebo-controlled trial with the well-known beta-blocker drug propranolol, which dampens visceral arousal by blocking beta-adrenergic receptor signaling (Berridge, 2008; Mueller & Ayres, 1980).

The Affective Harm Account: Beyond Dichotomies

If the distinction between two kinds of affect has been underappreciated in moral psychology, we suggest that the differences between affect and harm have been overstated. Past work in moral psychology often contrasts "cognitive" perceptions of harm against the "affective" experience of negative gut feelings (Greene et al., 2001), but strict dichotomies between

cognition and affect are being challenged (De Neys & Pennycook, 2019; Melnikoff & Bargh,
2018). In morality, affective appraisals are intrinsically "cognitive," because they involve
conceptual categorization at some implicit level as well as explicit self-reports of semantic
content. Perceptions of harm are also intrinsically "affective," because they involve appreciating
the causation of suffering, which is aversive (Blair, 1995; Crockett et al., 2008, 2010). The
overlap between "cognitive" harm and "affective" experiences is supported by bi-directional
links between these neural systems, which are especially prevalent in brain areas that output
higher-order judgments (Oosterwijk et al., 2012).

We formalize the intertwining of affect and harm in morality in the *affective harm
account* (AHA) of moral judgment. Classical models (**Table 1**) emphasize *either* harm *or* affect,
but the AHA emphasizes the joint importance of both harm and affect as mutually-reinforcing
sources of information in moral judgment (**Figure 2**).

First, the AHA builds on Dyadic Morality, which acknowledges the role of affect while
redefining harm as an intuitive, participant-determined perception. In this view, perceived harm
need not be the product of "explicit reasoning"; instead, perceived harm emerges from the
perceptual categorization of a moral act as harmless/harmful along a continuum from low (e.g.,
sleeping) to medium (e.g., eating a dead pet bird) to high harm (e.g., murder). What acts become
perceived and classified as higher or lower in harm depends on one's not-necessarily-conscious
cognitive template of harm, constructed through personal experience and sociocultural schemas.
This conceptualization of harm is consistent with past claims about "informational
assumptions" surrounding harm (Turiel et al., 1987) and with pluralism (Shweder, 2012;
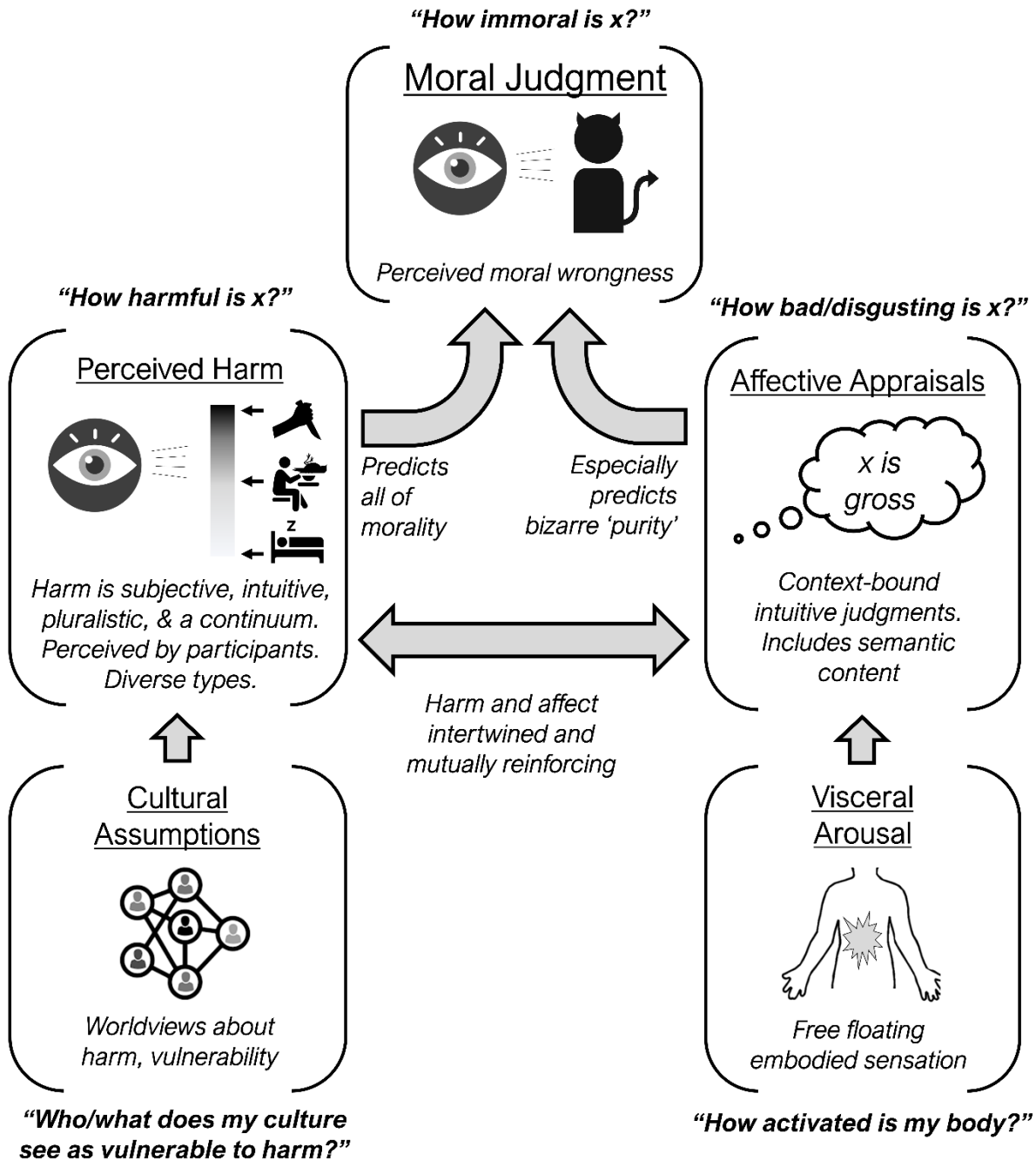Shweder et al., 1997).

Figure 2. Schematic of the Affective Harm Account (AHA) of moral judgment. All questions in quotes should be understood as implicit/intuitive perceptions. (Note: ultimately, all arrows are likely bi-directional. Moral judgments impact both perceived harm and affective appraisals, individual perceptions can shape culture, and affective appraisals can influence visceral arousal. But here we focus on the predictors of moral judgment.)

Second, the AHA divides "affect" into visceral arousal and affective appraisals (in this case, operationalized as disgust about moral scenarios). In doing so, the AHA draws upon classic and contemporary affect-as-information and constructionist models to specify the potential roles of visceral arousal and affective appraisals in moral judgment (Barrett, 2017; Clore & Ortony, 2013; MacCormack & Lindquist, 2017, 2019; Schachter & Singer, 1962; Schwarz & Clore, 2003; Storbeck & Clore, 2008). The AHA argues that visceral arousal may indeed sometimes help fuel moral judgment but does so via its influence on affective appraisals, which help disambiguate and "attach" those visceral feelings to the situated moral context. Thus, visceral arousal acts on moral judgment only *indirectly*, via the more context-bound affective appraisals. Affective appraisals should be especially relevant for moral judgment when they disambiguate the moral meaning of acts, which is especially likely when acts are unfamiliar, unusual, novel, etc. Ultimately, the AHA aims to support descriptive differences in patterns of moral judgment (Young & Saxe, 2011) between the common, everyday violations that people often encounter, versus the more unusual, unfamiliar, or novel violations that people are less likely to encounter (e.g., in our culture, unusual purity violations).

Three Predictions of the Affective Harm Account

The AHA poses three key hypotheses (**Table 2**) which not only clarify the roles of affect and harm in moral judgments, but also predict differences in the contributions of affect and harm across moral contexts (i.e., type of moral scenario) without needing to posit any distinct moral "foundations." Scenarios in moral psychology can be divided into two groups: everyday moral violations (e.g., abuse, cheating, lying) that make up 99% of our moral world (Hofmann et al., 2014), and more unusual violations written by experimenters to capture impurity (e.g., sex with dead animals). In contrast to everyday violations of purity such as pornography and prostitution, the unusual purity violations of moral psychology are seldom important in the real

world but are heavily studied in moral psychology (representing thousands of citations in the field; Gray & Keeney, 2015).

There are two key differences between everyday moral scenarios and unusual purity scenarios. First, purity scenarios are created by researchers to be more disgusting or emotionally evocative (although we note that there is no "special" link between purity and disgust; Cameron et al., 2015). Second, by virtue of their unusualness and infrequency in daily life, purity scenarios are more novel (unlikely to be encountered in daily life due to lower base rates) and thus more likely to be ambiguously immoral. As we detail, both evocativeness and novelty likely make affect more relevant to moral judgments of unusual purity scenarios.

Hypothesis 1. The first hypothesis of the AHA is that perceived harm ("seems harmful") and affective appraisals ("seems bad/disgusting") should be intertwined and mutually reinforcing rather than conflicting or independent mechanisms (H1: Harm and affect are intertwined). However, the specific relationship between perceived harm and affective appraisals likely depends on whether a given moral scenario is common vs. unusual.

Table 2. Hypotheses of the affective harm account (AHA) of moral judgment.

Hypotheses	Predictions
H1. Harm and affect are intertwined	Perceived harm and affective appraisals are connected & mutually reinforcing, not competing nor independent.
---- H1a. Everyday, harm leads affect	In more familiar moral contexts (e.g., everyday moral scenarios) perceived harm leads, predicting affective appraisals more than vice versa
---- H1b. Unusual, affect leads harm	In more unusual, ambiguous, novel, or evocative moral contexts (e.g., unusual purity scenarios), affective appraisals lead, predicting perceived harm more than vice versa
H2. Dual importance of harm and affect	Both perceived harm and affective appraisals predict moral judgment across contexts.
---- H2a. Harm is consistently relevant	Perceived harm should be ubiquitous, consistently predicting moral judgment regardless of the moral context.

---- H2b. Affect is contextually relevant	Affective appraisals should more strongly predict moral judgments of unusual, ambiguous, or novel moral contexts (e.g., unusual purity scenarios).
H3. Visceral arousal empowers affective appraisals	Visceral arousal fuels affective appraisals but only in unusual, ambiguous, novel, or evocative moral contexts (e.g., unusual purity scenarios) where that arousal can serve as information to guide appraisals.
---- H3a. Blunting arousal transforms unusual purity scenarios into everyday scenarios	Without strong visceral arousal, the roles of perceived harm and affective appraisals in unusual purity violations are more similar to their roles in everyday moral contexts.

All theories of moral judgment—even those that disagree (Graham et al., 2013; Schein & Gray, 2018)—agree that most everyday immoral acts (e.g., lying, disloyalty, abuse, unfairness) are dyadic. They involve someone ("agent(s)") who does the immoral act, and someone ("patient(s)") negatively impacted by the immoral act. Accordingly—because harm is dyadic (Schein & Gray, 2018)—the harmfulness of everyday acts should be easy to judge and cognitively salient. Thus, in everyday moral scenarios, perceived harm should "lead," predicting affective appraisals rather than vice-versa (H1a: Everyday, harm leads affect). When people wonder "how bad does it seem?" they rely upon their answer to "how harmful does it seem?" to inform their judgment of badness. In contrast to everyday moral scenarios, the harmfulness of unusual purity scenarios is less obvious. For unusual purity scenarios, affective appraisals should lead, predicting perceived harm more than vice-versa (H1b: Unusual purity, affect leads harm)—when people wonder "how harmful does it seem?" they rely upon their answer to "how bad (or disgusting) does it seem?" This prediction is consistent with "affect-as-information" theories (Schwarz & Clore, 2003; Storbeck & Clore, 2008), which argues that the mind uses ongoing visceral affect to help make sense of ambiguous or novel stimuli.

Hypothesis 2. The second hypothesis of the AHA is that both perceived harm and affective appraisals should matter in predicting moral condemnation across all scenario types (H2: Dual importance of harm and affect). In general, immoral acts are condemned more when they seem more "harmful" and more "bad/disgusting." However, there are likely differences

across scenarios. Harm, as a common currency across moral judgment (Schein & Gray, 2018), should be consistently predictive across all domains of morality, whether that be everyday or more unusual violations (H2a: Harm is consistently relevant). In this way, people should consistently draw upon "harm-as-information," using their template of dyadic harm to identify potential immorality. Conversely, the predictive utility of affect is likely more specific. Not only are unusual purity scenarios more emotionally evocative, but their ambiguous harmfulness likely leads people to draw upon "affect-as-information" to help disambiguate their immorality—as claimed by past work (Schnall et al., 2008). We echo these claims to hypothesize that affect will be particularly relevant in unusual purity scenarios (H2b: Affect is contextually relevant).

Hypotheses H1 and H2 are parallel—they endorse the interconnection and predictive power of both perceived harm and affective appraisals but suggest that affective appraisals exert more influence in unusual scenarios. Importantly, we are parsimoniously explaining differences between everyday moral and unusual purity violations (Young & Saxe, 2011) without needing to posit a separate "foundation" for purity violations. There is little strong evidence that different moral judgments are driven by distinct mechanisms (see Schein & Gray, 2018 and Figure 1).

Hypothesis 3. The third hypothesis is about visceral arousal or "gut feelings." Arguing against past claims that gut feelings drive all moral judgment (Haidt, 2001), we outline a more circumscribed role for visceral arousal. We suggest that gut feelings matter, but only insofar as they help empower affective appraisals, which in turn predict moral judgment within the affect-laden context of unusual moral scenarios (H3: Visceral arousal empowers affective appraisals). Past work suggests that (the more "embodied") visceral arousal impacts judgment by feeding into affective appraisals (Lindquist & Barrett, 2008; MacCormack & Lindquist, 2017, 2019; Niedenthal et al., 2005). This hypothesis argues that, when visceral arousal is blunted, moral judgment *per se* is not impacted; instead, affective appraisals become less predictive of moral judgment. More specifically, blunting visceral arousal should lead unusual purity scenarios to be judged more like everyday moral scenarios—both making affective appraisals less predictive of

moral judgment and also giving perceived harm more explanatory power over affect. (H3a: Blunting arousal transforms unusual purity scenarios into everyday scenarios).

The Two Horses of Moral Judgment. As an analogy for these hypotheses, imagine the cart of moral judgment being pulled by two horses, perceived harm (harm) and affective appraisals (affect). These horses are both always present and pulling (H2) and always—but not rigidly—yoked together (H1). They both encourage each other, such that when one speeds up the other one does too (H1), but sometimes one horse exerts relatively more effort in pulling the cart compared to the other (H1a & H1b). The harm horse is consistent and reliable, always pulling moral judgment a similar amount no matter the location (H2a)—whether on the well-worn roads of everyday morality or when plunging through the eerie forests of unusual purity. In contrast, the affect horse—easily spooked—becomes panicked and pulls much harder through those unfamiliar purity forests (H2b). The harm horse has wide experience on many roads across the land, and so it exerts more leadership upon the affect horse on these "typical" roads (H1a). However, in unusual, eerie forests, the affect horse is so frenzied that the harm horse cannot help but be pulled along by it (H1b). Although visceral panic fuels the affect horse in eerie forests (H3), its panic can be soothed though a special pill (e.g., propranolol). Once soothed, the affect horse pulls less aggressively and allows the harm horse to take the lead again (H3a).

Of course, any analogy is imperfect, but we hope this gives an intuitive sense of the three hypotheses of the AHA. Here, we test the first two hypotheses by assessing moral condemnation, perceived harm, and affective appraisals in response to both everyday moral and unusual purity scenarios. We test the third hypothesis—which lies at the heart of debates about "gut feelings"—via a pharmacological manipulation.

The Current Study: Propranolol and Moral Judgment

The current study examines whether—and how—moral judgments are impacted by propranolol, a well-validated beta-blocker often prescribed to help people manage anxiety and

distress by blocking beta-adrenergic receptors, a key sympathetic nervous system signaling pathway that underpins visceral arousal (Mills & Dimsdale, 1991). Propranolol is effective at reducing the experience of physiological arousal (Grillon et al., 2004) and, to date, at least one other study examined an affective social phenomenon after administration of propranolol: implicit racial bias (Terbeck et al., 2012). These authors suggest that "propranolol may lead to a significant reduction in" ... "somatic experiences dependent on increased sympathetic activation" ... "and thereby reduce the 'embodied' experience of emotional responses to a racial out-group." (Terbeck et al., 2012, p. 423) Given this past work and theorizing connecting moral judgment to an embodied affective response (Schnall et al., 2008), propranolol is an ideal intervention to explore the role of visceral arousal (i.e., "gut feelings") in moral judgment.

To date, only one other study has examined the effect of propranolol on moral judgment, investigating its impact on judgments of moral dilemmas (e.g., trolley problems; Terbeck et al., 2013). Consistent with classic "gut feeling" accounts, the authors predicted that the blunted visceral arousal induced by propranolol would make utilitarian decisions more acceptable in "personal" dilemmas (i.e., those involving inflicting direct harm; Terbeck et al., 2013). However, results demonstrated the exact *opposite* pattern: blunting of visceral arousal via propranolol made inflicting direct harm *less* acceptable. Given the results, the authors concluded that "general emotional physiological arousal is not likely to play an essential role in generating deontological judgments" (Terbeck et al., 2013, p. 326). Of course, the moral domain is richer than trolley problems, which are constructed to be conflicting (Gawronski et al., 2017) by pitting affect *against* harm—not an ideal framework for testing the potential overlap of harm and affect.

Although no studies (to our knowledge) have used propranolol when assessing standard moral judgment scenarios, one study did examine a kind of visceral blunting—operationalized by administering ginger tablets—and its impact on judgments of unusual purity scenarios (Tracy et al., 2019). Ginger inhibits nausea (such as in morning sickness, Thomson et al., 2014), and so the authors hypothesized that it would inhibit the negative gut feelings (measured as "disgust"

ratings) tied to unusual purity scenarios, rendering them less immoral. Although ginger did indeed decrease the moral condemnation of some purity violations, questions remain.

First, the Tracy et al (2019) study did not assess perceived harm, nor use any typical harm-relevant scenarios (beyond a trolley problem in Study 2), which means that those data cannot speak to the relevance of harm and its potential intersection with affect, nor to the generalizability of affect to non-purity scenarios. Second, ginger is more typically classified as a "non-pharmacological" intervention (Thomson et al., 2014) and its full physiological mechanisms of action remain under-characterized, including in the context of nausea (Marx et al., 2017). Indeed, nausea occurs via the coordination of multiple physiological systems—such as sympathetic nervous system increases, parasympathetic nervous system decreases, increased dysrhythmic gastric activity, and vasopressin increases (see Stern, 2002). Yet without nausea manipulation checks (e.g., electrogastrography measures; self-reported gastric symptoms), it is unclear whether ginger reduced disgust ratings via nausea reduction or via other known effects (e.g., reducing inflammation, stimulating blood circulation). Finally, the operationalization of nausea as self-reported "disgust" confounds affective appraisals with visceral arousal.

Given these ambiguities, we administered an arousal-blunting pharmacological agent (i.e., propranolol) known for its well-specified physiological impacts (i.e., blockade of beta-adrenergic receptors). Our goal was to test the causal role of visceral arousal in moral judgment and the predictions of the AHA. In the present study, participants read a series of often-used vignettes that included everyday moral violations and unusual purity violations. They made moral judgments of these scenarios and also rated perceived harm and affective appraisals. To best speak to both classic and modern studies in moral psychology (Schnall et al., 2008; Tracy et al., 2019), affective appraisals were operationalized as in previous studies: via disgust ratings. There are many morally-relevant emotions (e.g., anger, outrage), but disgust is argued to be especially visceral and embodied (Haidt et al., 1997) and especially tied to judgments of unusual purity violations (Horberg et al., 2009, but see Cameron et al., 2015). To situate findings within

this past work, we used disgust ratings as our operationalization of affective appraisals and used unusual purity violations as the moral context for affective appraisals of disgust.

In parallel with affective appraisals of disgust (how disgusting a given moral scenario was), we operationalized harm as appraisals of perceived harm *about* the moral scenario (how harmful a given moral scenario was). Note that it is possible to assess harm more implicitly (e.g., Gray et al., 2014, Study 2), but similar results are obtained when harm is assessed via the typical method of appraisals (How harmful is this act?). Operationalizing perceived harm as an appraisal is consistent with past work (Gray et al., 2014), and although ratings of harm—as with any rating, including disgust, impurity, or loyalty—could involve some post-hoc justification, these harm assessments use the same popular self-report Likert scales used to examine moral intuitions (Graham et al., 2009). Accordingly, post-hoc justifications are unlikely, especially because participants are anonymous, because responses are given alone, and because of past work revealing the intuitive nature of harm perception (Schein & Gray, 2018).

Methodological Innovation. The present study used gold-standard methods in psychopharmacology: a randomized, double-blind, placebo-controlled trial of the beta-blocker drug propranolol to blunt visceral arousal. We examined the effect of this manipulation on judgments of immorality, disgustingness, and harmfulness of both everyday moral violations and unusual purity violations. Pharmacological methods have revealed links between physiological and social psychological processes in other areas (Inagaki, 2018), but these methods remain largely absent from moral psychology (but see Terbeck et al., 2013). Thus, we bring this innovative approach to bear on a key debate in social-personality psychology.

<div align="center">

Method

</div>

Participants

Data presented here were collected as part of a parent study assessing the effects of beta-adrenergic receptor blockade with propranolol on physiological and psychological reactivity to an acute stressor (MacCormack, Armstrong-Carter, Gaudier-Diaz, et al., 2021; MacCormack,

Armstrong-Carter, Humphreys, et al., 2021; MacCormack, Gaudier-Diaz, et al., 2021). None of

the present data are published elsewhere. A total of 90 healthy young adults (*N*=43 propranolol,

N=47 placebo) were recruited in the parent study and all had usable data from the moral

judgment task (35 self-identified women, 47 self-identified men; M_{age}=20.33 years old with

range of 18-25 years; 56% White, 25% Asian/Asian American, 9% Black/African American, 7%

bi- or multi-racial, 2% other). Participants were assigned to placebo vs. propranolol condition by

the primary institution's investigative drug services pharmacy, who also ensured that conditions

were matched on self-ascribed gender and race/ethnicity. Given the use of a pharmacological

drug, participants were extensively screened for eligibility via structured telephone interviews

and an in-person visit. Individuals were excluded based on many criteria, such as prior use of

beta-blockers, BMI over 33 (which can alter dosage effects and impede autonomic nervous

system physiological measurements), or a history of or current diagnosis of physical or mental

illness. See the Supplementary Materials (SMs) for full list of eligibility criteria and for a

CONSORT diagram of enrollment in this randomized, double-blind, placebo-controlled trial.

Power Estimates

A priori power estimates were conducted for primary hypotheses in the parent study, but

not for our secondary, partially exploratory hypotheses about the role of propranolol in moral

judgment. There remains very little work capitalizing on pharmacological blockade paradigms to

examine the effects of physiological and affective processes on moral judgment, likely due to the

difficulty of recruiting and running a randomized, placebo-controlled drug study. Thus, one

secondary goal of this paper was to help establish effect size estimates for the effect of

propranolol on moral judgment. As discussed earlier, to our knowledge, only one prior study has

examined the effect of propranolol on moral judgment (Terbeck et al., 2013). The effect size for

an interaction between a single 40 mg dosage of propranolol x task condition in this study was

moderate-to-large, *d*= .72 (Terbeck et al., 2013, p. 325) with a sample size of *N*=20 on

propranolol vs. *N*=20 on placebo. Given the size of this effect and our sample size of *N*=43 on a

single 40 mg dosage of propranolol vs. *N*=47 placebo, our sample size may be sufficient to clarify main effect and interaction effect sizes for *drug* (i.e., visceral arousal) x *moral scenario type* (everyday and unusual purity) on perceived harm, affective appraisals, and moral judgment. Importantly, we use a within-person design (i.e., moral scenario type: everyday vs. unusual purity), affording us greater power to probe effects. Consistent with this within-person approach, we used multilevel modeling with random effects to improve statistical inferences, resulting in a total of 880 observations across 90 participants.

Procedures

To manipulate visceral arousal, participants were randomly assigned to receive either a one-time, 40 mg oral dose of propranolol or a visually identical placebo pill, self-administered in the presence of a registered nurse. Participants then completed a psychosocial stress task, after which they relaxed for 90 minutes to ensure that the stress task did not influence performance on subsequent tasks. After resting, participants rated and evaluated several moral scenarios (details below) around two hours after drug administration. The half-life of propranolol is five hours (Williams et al., 1986), suggesting that the drug was still active during moral judgment.

Measures

Moral Scenarios. Participants rated 12 scenarios encompassing a diversity of moral concerns. Four scenarios were unusual purity violations (e.g., "Eating a dog after it dies of natural causes") taken directly from previous work (Graham et al., 2009). The other eight scenarios involved more common, everyday violations, also taken from past work (Clifford et al., 2015). There were more of these everyday scenarios to reflect the fact that everyday morality involves many different values violations (Graham et al., 2013) including cheating/lying/stealing (i.e., fairness; "Cheating in class to get a good grade"), betrayal of in-group members (i.e., loyalty; "Deceiving a friend"), and subverting social hierarchies (i.e., authority; "Forging someone's signature on a legal document"). One criticism of past work examining harm is that using severely harmful scenarios (e.g., murder, genocide) leads people to

anchor upon harm more than they might otherwise (Graham, 2015). Accordingly, we used low-level everyday acts that we thought would be seen as *relatively* harmful (compared with unusual purity scenarios) but not so extremely harmful as to provide an overly salient anchor.

In addition to these moral scenarios, participants also rated some non-moral scenarios. The goal here was to provide context for participants (i.e., a spectrum of variation across morality and disgust), consistent with recent methodological recommendations (Cameron et al., 2015). As we had no hypothesis about these non-moral scenarios, they were not analyzed but their data and a full list of scenarios are available on our OSF page (see Results for link).

Scenario Ratings. Participants rated the immorality (our operationalization of moral judgment), harmfulness (perceived harm), and disgustingness (affective appraisal) of each scenario with the following questions—all using 6-point Likert scales from 1 to 6:

Immorality: How immoral is this act? 1 (Not immoral) to 6 (Extremely immoral)

Harmfulness (Harm): How harmful is this act? 1 (Not harmful) to 6 (Extremely harmful)

Harmfulness (Victimhood): Is there a victim in this act? 1 (Definitely no) to 6 (Definitely yes)

Disgustingness: How gross is this act? 1 (Not gross) to 6 (Extremely gross)

Harm was assessed via two questions—victimhood and harmfulness—consistent with past work (Gray et al., 2014), in part because of criticisms that "harmfulness" can be used symbolically (Gutierrez & Giner-Sorolla, 2007). Also consistent with past work, responses to these two questions were combined ($r_{mean} = .3$, $p < .005$; Gray et al., 2014) to yield an overall perceived harmfulness measure. Further, rather than having participants rate the "disgustingness" of scenarios, they rated how "gross" they were (consistent with past research; Widen & Russell, 2002) because "disgusting" is sometimes used as a synonym for immoral (Chapman & Anderson, 2013), and we wanted to tap "core" disgust (Haidt et al., 1994).

Scenario Checks. Although the everyday and unusual purity scenarios used here are standard in the field, research often reveals heterogeneity within these domains. Inspired by reviews, we examined whether these everyday and unusual purity scenarios were homogeneous,

or whether there were scenario outliers. Accordingly, we conducted consensus cluster analyses in R using the *ConsensusClusterPlus* package (Wilkerson & Hayes, 2010). Results confirmed that there were two clusters representing everyday vs. unusual purity moral scenarios, but one vignette within each category did not clearly load onto either cluster. As such, those two scenarios are dropped from subsequent results, leaving a remainder of seven everyday moral and three purity moral scenarios. See **Table 3,** as well as the SMs, for relevant analyses.

Table 3. Moral scenarios organized by type.

Everyday Moral Scenarios	Unusual Purity Moral Scenarios
Leaving spouse after 20 years of marriage	Having sex with a corpse
Deceiving a friend	Rubbing feces on a Bible
Shoplifting a candy bar*	Eating a dog after it dies
Stealing from a neighbor	Signing a paper to sell soul*
Cheating in class	
Forging signature	
Lying in business	
Stealing bank information	

Note: asterisks indicate scenarios that the consensus clustering analysis identified as fitting poorly within any given scenario type.

Physiological Manipulation Check

Consistent with past work (Harris et al., 1967), a physiological manipulation check confirmed that propranolol was effective in reducing visceral arousal, as measured by cardiac pre-ejection period (PEP) at a 5-min resting pre-drug administration baseline vs. at a 5-min resting post-drug baseline acquired 60-minutes after drug administration. PEP was measured using electrocardiography and impedance cardiography to provide an index of cardiac contractility controlled by the sympathetic nervous system (e.g., beta-adrenergic signaling). PEP thus reflects sympathetic nervous system-specific activity and related arousal (Newlin & Levenson, 1979). There was no significant difference between groups in PEP at the pre-drug baseline, but 60-min post-drug administration, the propranolol group showed significantly slower PEP relative to the placebo group, $t(80) = -2.63$, $p= .01$, confirming that propranolol did significantly blunt visceral arousal.

Results

The Supplementary Materials (SMs) include additional results which are signposted where appropriate. Data, syntax, and materials can be found on OSF at the following link: https://osf.io/ak3g9/. In the spirit of transparency and open science, we also acknowledge that these analyses—and hence the framing of this paper—have been sharpened and refined through the review processes.

We present two primary sets of analyses, which we introduce before connecting to specific hypotheses with supporting results. We first calculated zero-order correlations between perceived harm (harmfulness), affective appraisals (disgustingness), and moral judgment (immorality) across and within *drug condition* (between-person: placebo, propranolol) and *scenario type* (within-person: everyday, unusual purity; **Table 4**). These correlations provide an intuitive description of how perceived harm, affective appraisals, and moral judgment interrelate across participants by drug condition and scenario type.

We then used a cross-classified multilevel modeling approach to more rigorously assess how perceived harm, affective appraisals, and moral judgment interrelate and how these may differ between/within person with respect to propranolol administration and scenario type. Three models were fit. In the main model, tabulated in **Table 5**, harmfulness (i.e., perceived harm) and disgustingness (i.e., affective appraisals) predicted immorality ratings (i.e., moral judgment). Fixed effects of harmfulness and disgustingness were allowed to vary by drug condition (0=placebo, 1=propranolol) and scenario type (0=everyday, 1=unusual purity) using multiplicative interaction terms. Two other similar multilevel models were also fit, one with perceived harm predicted by affective appraisals and moral judgment, and another with affective appraisals predicted by perceived harm and moral judgment. Covarying interrelations

between perceived harm, affective appraisals, and moral judgment are visualized in **Figure 3**, as tested by all three MLMs[2].

For all models, random intercepts were specified at the subject level (representing inter-individual variability in their average ratings of a given dimension), and separately at the scenario level (representing dimensional differences due to the specific scenario in question). Additionally, random slopes for both predictors' (drug condition, scenario type) main effects were specified at the subject level (representing inter-individual variability in how each dimension is related to the outcome). All interactions, regardless of significance, were probed at each level of drug condition/scenario type to provide a complete view of how conditions altered interrelations, although only significant moderation effects are interpreted. All models were fit in R (R Core Team, 2021) using the *lme4* package (Bates et al., 2015), with degrees of freedom and *p*-values for fixed effects calculated using *lmertest* (Kuznetsova et al., 2017). Simple slopes were calculated using the *interactions* R package (Long, 2019).

Unlike zero-order correlations, which simply summarize general relations in isolation from each other, the three MLMs each speak to a different angle of the inferential puzzle and should not be considered separate analyses. Instead, these three MLMs should be interpreted together, supporting a clearer understanding of interrelations between moral judgment, perceived harm, and affective appraisals while controlling for inter-individual variability. As such, to evaluate Hypothesis 1, we used estimates from all three MLMs to construct **Figure 3**. This diagram is a visual representation of the estimated fixed effects (and relevant simple slopes) from all three models (it does not reflect mediational analysis), corresponding to results in **Tables 5 and 6,** as well as supplementary tables in the SMs. This approach allowed us to visualize findings across all models holistically, revealing an underlying pattern of relations

---

[2] Because ratings of perceived harm, affective appraisals, and moral judgment are highly collinear, for any given MLM where we assessed one outcome of interest (e.g., moral judgment), we also included the other two measures to adjust for confounds due to shared variance between the three measures. Importantly, given the present nested data structure, this multivariate MLM approach is comparable to a MANOVA but more appropriate for modeling multivariate effects within the context of cross-classified nesting.

between moral judgment, perceived harm, and affective appraisals. However, given that Hypotheses 2 and 3 focus on the contrasting role of perceived harm vs. affective appraisals in relation to moral judgment, we chose to present that specific model in the main text (**Table 5**).

It bears mentioning upfront that, although perceived harm and affective appraisals predicted moral judgment, moral judgment also predicted perceived harm and affective appraisals. Traditional accounts may argue that one (or both) of these links are "post-hoc justification," but a more modern understanding might view these links as simply a product of the dynamic intertwining of affect, cognition, morality, and social cognition (De Neys & Pennycook, 2019; Gray et al., 2017; Melnikoff & Bargh, 2018). Questions of "how harmful" and "how disgusting" not only influence "how immoral," but also vice versa. If someone told you an act was extremely immoral you would infer that it was harmful and perhaps disgusting. Nevertheless, consistent with the focus of the field, here we explored the predictors of moral judgment, but future research should investigate these bidirectional links.

Evaluating Hypothesis 1: The AHA suggests that perceived harm and affective appraisals should be inter-related in the context of moral judgment rather than orthogonal, independent entities (H1). Consistent with this idea, zero-order correlations (**Table 4**) revealed consistent, strong inter-associations between perceived harm, affective appraisals, and moral condemnation, regardless of drug condition or scenario type.

Table 4. Bivariate correlations between ratings of perceived harm, affective appraisals, and moral judgment, overall and split by drug condition and scenario type.

|	Everyday Scenarios		Unusual Purity Scenarios		Across All Scenarios	
	Affective Appraisals	Moral Judgment	Affective Appraisals	Moral Judgment	Affective Appraisals	Moral Judgment
Placebo (*n*=47)						
Perceived Harm	.614***	.625***	.624***	.542***	.622***	.573***
Affective Appraisals	--	.542***	--	.654***	--	.576***
Propranolol (*n*=43)						
Perceived Harm	.425**	.601***	.565***	.511***	.515***	.689***
Affective Appraisals	--	.446**	--	.666***	--	.519***
Overall (*n*=90)						
Perceived Harm	.517***	.616***	.590***	.539***	.572***	.633***
Affective Appraisals	--	.490***	--	.652***	--	.548***

*** $p<.0001$, ** $p<.001$
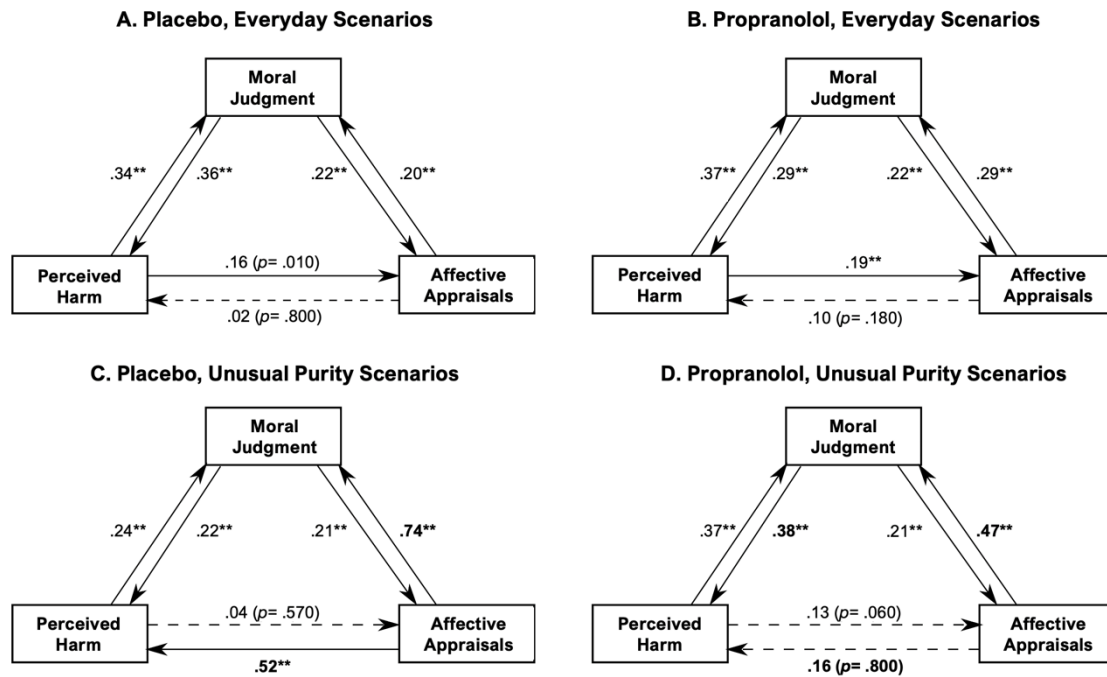
Figure 3. Multilevel model fixed effects and simple slope relations between moral judgment, perceived harm and affective appraisal. See Tables 5, 6, S2-S5 for full models. Double asterisks (**) indicate $p<.001$ with all other p-values provided. Solid arrows denote significant fixed effects/simple slopes. Bolded labels denote that the simple slope has a significant corresponding two- or three-way interaction. Effects provided are standardized

βs ($p − value$). Although this diagram resembles a mediation model, mediation models were not performed for this particular set of analyses nor should this diagram be interpreted as such.

The AHA further proposes that, in everyday moral scenarios, perceived harm should be more predictive of affective appraisals than vice versa (H1a), whereas in the context of unusual purity scenarios, affective appraisals should be more predictive of perceived harm than vice versa (H1b). Consistent with this idea, main effects from the MLMs showed that, for individuals on placebo in the everyday scenarios (**Figure 3, Panel A**), perceived harm predicted affective appraisals, β=.16, p= .01, whereas affective appraisals did not predict perceived harm, β=.02, p>.25. A similar pattern was observed in participants on propranolol, where perceived harm predicted affective appraisals, β=.19, p<.001 (**Panel B**) but not vice versa, β=.01, p>.25. On the other hand, in unusual purity scenarios, for individuals on placebo, affective appraisals predicted perceived harm, β=.52, p< .001 (**Panel C**) while perceived harm did not predict affective appraisals, β=.04, p>.25. Interestingly, propranolol blunted the predictive power of affective appraisals in unusual purity scenarios, with affective appraisals no longer predicting perceived harm, β=.16, p>.25 (**Panel D**). There was also no effect of perceived harm on affective appraisals, β=.13, p=.06, although it was marginal—and we note that these results for the *propranolol x purity scenario* effect were consistent with Hypothesis 3, discussed shortly.

Together, these findings support the first hypothesis of the AHA: perceived harm ("seems harmful") and affective appraisals ("seems bad/disgusting") are closely intertwined rather than orthogonal, independent domains (H1). However, the specific relationship between perceived harm and affective appraisals may vary by context, i.e., scenario type. Specifically, in everyday moral scenarios, perceived harm mattered more for affective appraisals than vice versa (H1a), whereas in the context of unusual moral scenarios (i.e., purity scenarios), affective appraisals mattered more for perceived harm than vice versa (H1b).

Evaluating Hypothesis 2: The second hypothesis of the AHA is that both perceived harm and affective appraisals predict moral condemnation (H2) across moral scenarios. **Table 5** contains the estimates and inferential statistics from the previously described cross-classified multilevel model with moral judgment as the outcome. **Table 6** contains the simple slope estimates for each combination of drug condition and scenario type. As seen in **Table 5**, there were significant effects of perceived harm and affective appraisals on moral judgment, βs=.34, .20, ps< .001 respectively, consistent with Hypothesis 2. However, the AHA also posits that the relative "weights" of perceived harm vs. affective appraisals should differ by moral context (whether everyday vs. more unusual, i.e., purity). Perceived harm should consistently predict across contexts given its ubiquity in moral judgment (H2a) whereas affective appraisals should be especially important in contexts that are unusual, evocative, ambiguous, etc. (e.g., purity scenarios), as they help provide disambiguating information (H2b). Evidence for both H2a and H2b can be seen in **Figure 3**. Across scenario types, perceived harm showed a relatively stable effect size in its relation to moral judgment, even when considering drug condition, βs=.34, .24, .37, .37; ps< .001. Correspondingly, there was no significant interaction of *scenario type x perceived harm* on moral judgment, β= -.10; p> .25, in line with the interpretation that effects of perceived harm on moral judgment remain consistent across all moral scenarios (H2a). This consistency was not the case for affective appraisals.

Table 5. Fixed effect estimates for moral judgment as predicted by drug condition, scenario type, perceived harm, and affective appraisals.

Predictors	Moral Judgment as Outcome			
	Est.	SE	95% CI	*p*
(Intercept)	0.01	0.14	[-0.29, 0.31]	.95
Drug Condition (0=Placebo, 1=Propranolol)	0.00	0.09	[-0.18, 0.17]	.97
Scenario Type (0=Everyday, 1=Purity)	-0.57	0.27	[-1.15, 0.01]	.052
Perceived Harm	**0.34**	**0.07**	**[0.21, 0.47]**	**<.001**
Affective Appraisals	**0.20**	**0.07**	**[0.07, 0.33]**	**<.001**
Drug x Scenario	0.30	0.19	[-0.08, 0.68]	.12
Drug x Perceived Harm	0.03	0.09	[-0.15, 0.21]	.74
Drug x Affective Appraisals	0.09	0.09	[-0.10, 0.27]	.34
Scenario x Perceived Harm	-0.10	0.09	[-0.28, 0.07]	.25
Scenario x Affective Appraisals	**0.53**	**0.14**	**[0.25, 0.82]**	**<.001**
Drug x Scenario x Perceived Harm	0.10	0.12	[-0.13, 0.34]	.38
Drug x Scenario x Affective Appraisals	**-0.36**	**0.18**	**[-0.72, 0.00]**	**.048**

Note: Significant effects are bolded. The outcome and continuous predictors were standardized before model fit. As such, the estimates can be interpreted as standardized βs.

Table 6. Simple slopes for perceived harm and affective appraisals by drug condition and scenario type.

Condition	Scenario	Effect	Est.	SE	t-value	p-value
Placebo	Everyday	Perceived Harm	0.34	0.07	5.06	<.001
Placebo	Everyday	Affective Appraisals	0.20	0.07	3.10	<.001
Placebo	Purity	Perceived Harm	0.24	0.08	3.13	<.001
Placebo	**Purity**	**Affective Appraisals**	**0.74**	**0.14**	**5.35**	**<.001**
Propranolol	Everyday	Perceived Harm	0.37	0.07	5.47	<.001
Propranolol	Everyday	Affective Appraisals	0.29	0.07	4.33	<.001
Propranolol	Purity	Perceived Harm	0.37	0.08	4.80	<.001
Propranolol	**Purity**	**Affective Appraisals**	**0.47**	**0.11**	**4.29**	**<.001**

Note: Bolded entries denote a significant interaction related to the simple slope.

Although affective appraisals showed a significant main effect on moral judgment on average, as discussed above, the main effect was qualified by an interaction with moral scenario type. Specifically, as shown in **Table 6** and **Figure 3 (Panels A & C)**, when individuals were on placebo and judged everyday moral scenarios, the simple slope effect size of affective appraisals was modest, β = .20, *p* < .001, but in the context of unusual purity moral scenarios, the simple slope effect size of affective appraisals was large, β = .74, *p* < .001. Furthermore, there was a significant interaction between *scenario type x affective appraisals*, β = .53, *p* < .001 (**Table**

5), such that affective appraisals were more predictive of moral judgment for unusual purity scenarios relative to everyday moral scenarios (H2b). Beyond these main effects, scenario type had a main effect on moral judgments: everyday scenarios were rated on average as slightly more immoral than unusual purity scenarios, but we are cautious not to draw strong inferences given that this was a marginal effect, β= -.57, p= .052.

Evaluating Hypothesis 3: The final prediction of the AHA is that visceral arousal should matter only insofar as it empowers the role of affective appraisals in moral judgment in an affectively relevant context, e.g., purity scenarios (H3). Furthermore, blunting visceral arousal should lead purity scenarios to be judged more like everyday scenarios—both making affect less predictive of moral judgment and giving perceived harm more explanatory power (H3a). Again, **Figure 3** alongside **Tables 5 and 6** provide evidence for these hypotheses, when considering how propranolol administration alters these effects. As described above in Hypothesis 2, perceived harm showed a relatively stable effect size regardless of drug condition, suggesting that blunting visceral arousal did not alter the role of perceived harm in relation to moral judgment. Accordingly, there was no two-way interaction of *drug x perceived harm,* β= .03, p>.25 nor a three-way interaction of *drug x scenario x perceived harm,* β= .10, p>.25.

The impact of propranolol administration was specific: it blunted affective appraisals but only for unusual purity scenarios (and not everyday scenarios), evidenced by a nonsignificant two-way interaction of *drug x affective appraisals,* β= .09, p> .25, yet a significant three-way interaction between *drug x scenario x affective appraisals* on moral judgment, β= -.36, p= .048. This three-way interaction is further illustrated in **Figure 3**. Interestingly, consistent with H3a, there was no main effect of drug condition on moral judgment, suggesting that blunting of visceral arousal via propranolol did not directly alter moral judgment, β= .00, p> .25. Instead, we see that propranolol (relative to placebo) reduced the effect size of affective appraisals for purity scenarios (compare in **Figure 3, Panels D vs. C**: β= .20 for propranolol relative to β=

.74 for placebo, *ps*<.001, **Table 6**), in turn altering the relative degree to which perceived harm

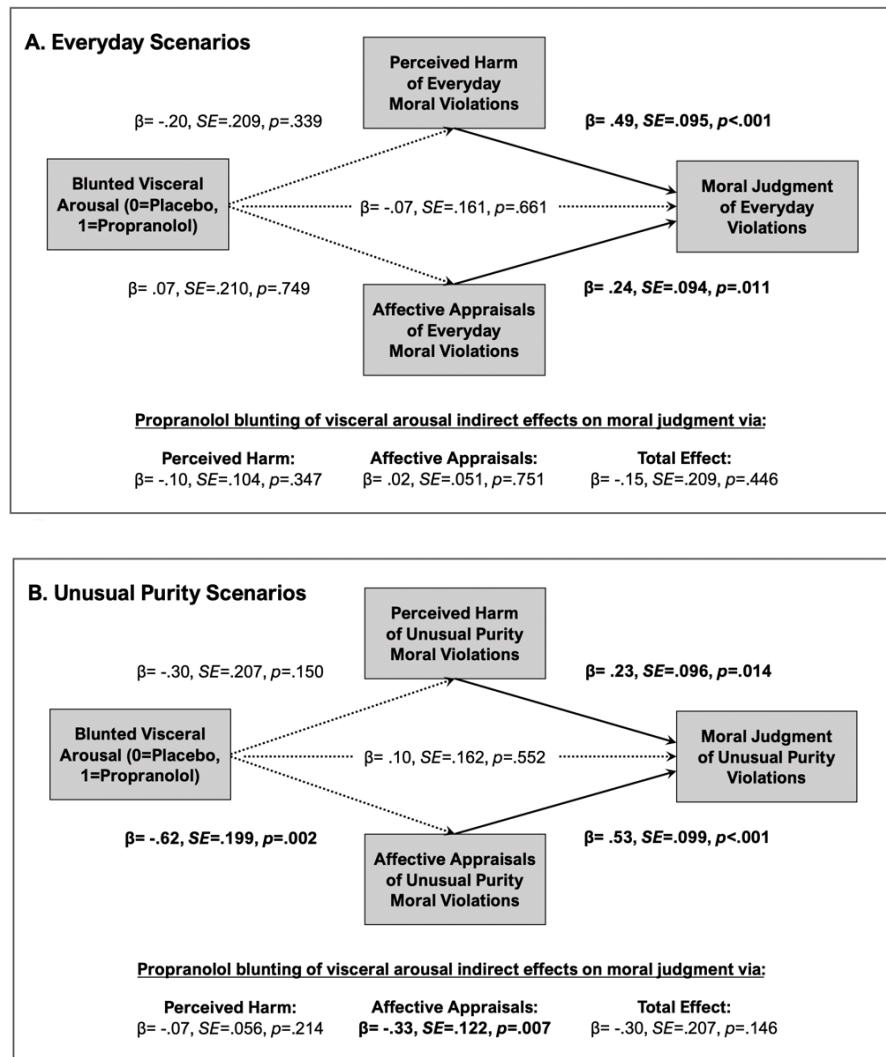vs. affective appraisals matter for moral judgment.

Figure 4. Manifest variable structural equation models testing direct vs. indirect paths for effects of propranolol on moral judgment in everyday moral scenarios (Panel A) vs. unusual purity scenarios (Panel B). Significant paths shown with solid lines and nonsignificant effects with dotted lines. Coefficients are standardized betas to provide effect size comparisons. Covariance between perceived harm and disgust appraisals was included in both models, but not depicted here, with a covariance of β= .52, *SE*=.117, *p*<.0001 in the everyday moral scenarios model and a covariance of β= .54, *SE*=.113, *p*<.0001 in the unusual purity scenarios model.

Finally, as a secondary demonstration of H3, we conducted supplementary mediation

models, as reported in **Figure 4**. Using manifest structural equation modeling (SEM; see SMs

for full modeling details), we sought to further investigate the above three-way interaction

between *drug x scenario x affective appraisals* by modeling the potential direct and indirect

paths by which perceived harm vs. affective appraisals might explain variance in propranolol's

effect on moral judgment, while also controlling for shared covariances between perceived harm

and affective appraisals. **Figure 4** shows that there was no direct effect of propranolol

administration (i.e., blunting of visceral arousal) on moral judgment, but there were significant

direct effects of both perceived harm and affective appraisals on moral judgment in the unusual

purity scenarios (consistent with the MLMs).

Importantly, there was a significant indirect effect of propranolol on moral judgment via

affective appraisals (β=-.33, *SE*=.122, *p*=.007)—but only for the unusual purity scenarios. There

was also no indirect effect of propranolol on moral judgment via perceived harm in the unusual

purity scenarios. As can be seen, propranolol administration was irrelevant to everyday moral

judgment, regardless of perceived harm or affective appraisals (**Figure 4**). These findings are

consistent with the third AHA hypothesis and the above MLM findings. The effects of visceral

arousal (i.e., gut feelings) in moral judgment were restricted to changing the predictive power of

affective appraisals and only for unusual purity scenarios.

Discussion

Moral psychology has long debated whether moral judgment is grounded in affect *or*

harm. Seeking to reconcile these apparently competing perspectives, we have proposed an

Affective Harm Account (AHA) of moral judgment. This account is conciliatory because it

highlights the importance of both perceived harm and affect, not as competing considerations

but as joint partners—two different horses yoked together pulling the cart of moral judgment.

The AHA also adds clarity to the previously murky nature of "affect" in moral

psychology, differentiating it both in nature and measurement as (at least) two phenomena—

embodied, free-floating, visceral arousal (i.e., "gut feelings") and self-reported, context-bound,

affective appraisals (i.e., "this situation is gross"). The importance of affect in moral judgment—

especially the "gut feelings" of visceral arousal—was tested via administration of propranolol,

which dampens visceral arousal via beta-adrenergic receptor blockade. Importantly, propranolol allows us to manipulate more general visceral arousal (rather than targeting a specific organ, like the gut, or a specific state, like nausea). This increases the potential generalizability of these findings to other moral scenarios (beyond disgust) where visceral arousal might be relevant. We measured the effect of propranolol (vs. placebo) on ratings of moral condemnation, perceived harm, and affective appraisals (i.e., operationalized as ratings of disgust, as in much past work). These ratings were obtained for both everyday moral scenarios (Hofmann et al., 2018)—which are dyadic in structure and thus obviously linked to harm—and for unusual purity scenarios, which are frequently linked to affective appraisals of disgust (Horberg et al., 2009). This study offers support for the three hypotheses of the AHA (**Table 2, Figure 2**).

First, the significant correlations and predictive links between perceived harm and affective appraisals reveal support for the first hypothesis of the AHA (H1: Harm and affect are intertwined). However, these links vary by scenario type. For everyday moral scenarios, it is harm that leads, predicting affective appraisals (H1a: Everyday, harm leads affect), consistent with the idea that people draw upon a dyadic harm-based template ("harm-as-information") to guide the vast majority of moral judgments (Schein & Gray, 2018). In contrast, for novel and ambiguous moral scenarios (e.g., unusual purity scenarios), it is affective appraisals that lead and predict harm (H1b: Unusual purity, affect leads harm), consistent with the idea that people may draw more upon "affect-as-information" to guide moral judgment when the features of the situation are particularly ambiguous or novel.

Nevertheless, across all scenario types, *both* affective appraisals and perceived harm mattered in moral judgment (H2: Dual importance of harm and affect). Again, however, scenario differences emerged. Harm, as a common currency across moral judgment (Schein & Gray, 2018), was consistently and robustly predictive across all domains of morality, whether everyday immorality or unusual impurity (H2a: Harm is consistently relevant). Conversely,

affect was most predictive for unusual purity scenarios (H2b: Affect is contextually relevant), which are less obviously harmful, consistent with affect-as-information (Schnall et al., 2008).

Finally, because propranolol impacted affective appraisals—but not moral judgment directly—we suggest that visceral arousal (i.e., "gut feelings") likely does not causally "drive" moral judgment in the strong sense advanced by past work (Haidt & Hersh, 2001). Visceral arousal is involved in moral judgment only indirectly, via its influence on affective appraisals in the context of unusual moral scenarios (H3: Visceral arousal empowers affective appraisals). Again, we interpret this as evidence in line with affect-as-information and related accounts (e.g., MacCormack & Lindquist, 2019; Storbeck & Clore, 2008; Schachter & Singer, 1962), where visceral arousal helps to disambiguate and guide mental inferences in the context of attributional ambiguity. Here, it appears that visceral arousal shapes the power of affective appraisals only when the "output" of harm-based template matching is ambiguous (i.e., the presence of harm is ambiguous in unusual purity scenarios). Consistent with the idea that visceral arousal empowers affective appraisals in unusual moral scenarios, blunting visceral arousal via propranolol caused purity scenarios to be judged more like everyday scenarios (H3a: Blunting arousal transforms unusual purity into everyday).

Our findings that propranolol reduced the predictive ability of affective appraisals for moral judgment within unusual purity scenarios is consistent with evidence from Tracy et al. (2019). They observed that ginger administration reduced ratings of disgust and immorality in unusual purity scenarios (but not other moral scenarios). However, it is important to note that these ginger administration studies only measured self-reported experiences of disgust without including nausea manipulation checks (i.e., assessments of nausea-related physiological changes; self-reports of nausea or gastric changes). As such, it is unclear whether ginger administration reduced the severity of these moral judgments directly via effects on nausea (or some other physiological pathway) or indirectly via affective appraisals of disgust. In the present study, we were able to confirm that propranolol blunted both visceral arousal (i.e., cardiac pre-

ejection period as an index of sympathetic nervous system activity) and affective appraisals of disgust, thus disentangling the interrelated effects of "gut feelings" as physiological changes vs. affective appraisals.

More generally, one way of understanding why visceral arousal only mattered for moral judgment indirectly—via affective appraisals (e.g., three-way MLM interaction, indirect mediation effects)—is that visceral arousal may be conceptually "further" from moral judgment compared to affective appraisals (and perceived harm). One could imagine a continuum of moral relevance, with moral judgments being maximally relevant (they are exactly about the immorality of an action), harm appraisals being highly relevant (given our dyadic harm-based moral template), affective appraisals also being highly relevant (they are about the affective qualities of an action), and visceral arousal being of low relevance (it is a free-floating feeling which may become linked or attached to a given stimulus, depending on the current situational features and related appraisals). The conceptual distance of visceral arousal from moral judgment makes it less relevant to moral judgment directly until it is "made relevant" via the more proximal, congruent psychological pathway of affective appraisals.

Implications

Support for the AHA offers several implications for moral psychology. First, it suggests that we should move past false dichotomies and ongoing debates about whether morality is driven by "harm" or by "affect." Instead, perceptions of harm are imbued with affect, and affect—at least in morality—is tied to harm. Crucially, both harm and affect appear to travel together in their relationship with moral judgment.

Second, the AHA redefines how we should think of the concepts of "harm" and "affect." In contrast to the arguments of popular accounts like Moral Foundations Theory, harm is not reasoned, nor binary, nor objective, nor monistic. Instead, harm is an intuitive continuum that is perceived by the participant (not decided by the experimenter) and varies across cultures and contexts. This redefinition draws from the Theory of Dyadic Morality (Schein & Gray, 2018) and

helps to vindicate early harm-based theories in moral psychology that argued for the importance of "informational assumptions" about harm (Turiel et al., 1987).

The pervasive role of perceived harm across moral judgment—including in unusual purity scenarios—provides strong evidence against classic formulations of Moral Foundations Theory. Harm is not just restricted to an artificially- and experimenter-defined "harm/care foundation" but is instead a pervasive concern across all moral judgment. Harm is not a set of specific moral violations (e.g., hitting a kid), but instead an overarching understanding of morality, a common currency, and a domain-general cognitive template for moral judgment. This is why perceptions of harm robustly predict moral judgment across all "foundations" (**Figure 1**).

If separate moral "foundations" do not exist (see also Schein & Gray, 2018), what are we to make of the many studies revealing correlations between interesting phenomena and measurements of fairness, loyalty, authority, and purity (Dungan & Young, 2019; Graham et al., 2009, 2016; Graham & Haidt, 2010)? The answer is to pivot our understanding of "foundations" away from the modules of evolutionary psychology and back to the "themes of ethical discourse" of anthropology (Shweder et al., 1997). Culturally important themes can help explain and predict cultural differences without needing a specific number of moral mechanisms. After all, the most successful modern application of Moral Foundations Theory is via natural language processing of moral language (Sagi & Dehghani, 2014), which measures the kinds of words that people use when discussing moral issues. These language analyses are merely *descriptive*, assessing the prevalence of 5 themes of "ethical discourse"—harm, fairness, loyalty, authority, and purity—without providing evidence for cognitive mechanisms. Counting the number of different animals encountered in different forests tells you little about their fundamental biological processes. Moreover, the scholars developing these language-based tools explicitly admit that they are measuring Moral Foundations "rhetoric" (Sagi & Dehghani, 2014)—not intuitive gut reactions but explicit (and perhaps even reasoned) moral arguments. People use

different words to argue for different values, but all these values—at least to the extent that they

are moralized—are grounded in a template of perceived harm (Gray & Keeney, 2015; Schein &

Gray, 2018).

The AHA also redefines "affect" in morality (and more broadly), by distinguishing

between embodied-free-floating visceral arousal and more explicit, contextualized affective

appraisals. In order to truly understand the role of "gut feelings" and affect in shaping moral

judgments, researchers should carefully distinguish actual physiological states/sensations from

affective appraisals—which appear to connect physiology to moral judgment. Furthermore,

whereas Moral Foundations Theory posits a pervasive role of affect and specific emotions across

all moral domains (e.g., anger-harm, disgust-purity; (Haidt & Joseph, 2004, 2007)), the AHA

argues for a more circumscribed causal role for affect in moral judgment[3]. Instead of "driving"

all moral judgments, affect helps to guide moral judgments when there is uncertainty,

ambiguity, novelty, etc. about the moral meaning of a given act or scenario. Thus, affect can still

matter for moral judgment without needing to posit moral "foundations," "modules," or special

links between specific emotions and moral content. Rather, we can explain the importance of

affect—and affect as a source of moral information—in the unusual purity scenarios by the fact

that: (1) the content in these scenarios is both more ambiguously harmful and more affectively

evocative than everyday moral scenarios, and (2) this ambiguity and evocativeness is likely

---

[3] Differences in Moral Foundations Theory vs. AHA characterizations of "affect" reflect differences in causal appraisal vs. constitutive appraisal models of emotion. For example, causal appraisal models tend to assume that appraisal dimensions of uncertainty, power, threat, etc. refer to literal cognitive mechanisms, paralleling how Moral Foundations Theory argues for discrete moral foundations of harm, purity, loyalty, etc. that serve as distinct cognitive mechanisms in moral judgment. On the other hand, constructionist-leaning theories, which encompass constitutive appraisals models such as affect-as-information theory and Schachter and Singer's two-factor theory, do not argue for discrete/dissociable cognitive appraisal types. These instead argue that emotional experiences and judgments emerge when interoceptive/visceral signals (e.g., "visceral arousal," "core affect") are made situationally relevant (via "appraisals"). Causal appraisal models also tend to see physiological or visceral changes as an *outcome* of appraisals, whereas constitutive appraisal models tend to see such bodily changes as a *source* of affective information that fuels and infuses appraisals. For further articulations of causal vs. constitutive appraisal theories, see Clore & Ortony, 2013; Gendron & Barrett, 2009; MacCormack & Lindquist, 2017; Moors, 2014).

because they are created by researchers to be ambiguously harmful and evocative, and from the

participant's perspective—strange, weird, and perhaps even bizarre (Gray & Keeney, 2015).

Building on this last insight, the AHA transforms how we understand patterns of

judgment surrounding unusual moral scenarios such as the often-used-but-unusual purity

scenarios of moral psychology (e.g., sex with a dead chicken). In the present data, these

scenarios were the only moral scenarios where affective appraisals were especially predictive of

moral judgment and where visceral arousal played a role—albeit indirectly. In everyday

scenarios, it was perceptions of harm that consistently predicted *both* moral judgment and

affective appraisals. We suggest that the field's long reliance on these unusual purity scenarios

has led to an inaccurate overemphasis on the role of affect in moral judgment. Rather than

focusing on contrasting specific moral domains of purity, loyalty, etc., we recommend that

future studies can instead take a domain-general approach, systematically manipulating

features such as novelty, ambiguity, extremity, or self-relevance across moral scenarios. This

approach could help clarify the boundary conditions surrounding *when* and *how much* affect

will matter for moral judgment.

Finally, the AHA argues for psychological *degeneracy,* a concept borrowed from biology,

which argues that it is adaptive for a complex system to have structurally different processes

that can perform the same functions, thereby promoting survival (Edelman & Gally, 2001). If

one process or structural pathway goes "offline" (e.g., is impaired or inactive), then other

processes or pathways can maintain system functioning. Because humans, as a social species,

rely on moral judgments to make social inferences that could make the difference between life or

death, wellbeing or harm, it is logical that psychological degeneracy may characterize moral

judgments. Our minds typically rely upon perceptions of harm—the ubiquitous common

currency—to inform judgments of morality, but when these perceptions become difficult (as

when confronted with unusual or novel moral situations), our minds turn to affect to guide such

judgments. However, if affect somehow becomes inhibited (here, via propranolol), the mind

may then resort back to its typical source of information for forming moral judgments: harm. In this way, questions of moral mechanisms are less about "which" source of information (harm or affect) drives moral judgment, but rather, which gets prioritized first and when.

Limitations

Social psychology has been transformed by the open science initiative (Open Science Collaboration, 2017), and one of its key tenets is the importance of well-powered studies. Pharmacological studies require intensive efforts (e.g., extensive screening to minimize participant risk, a physician to prescribe medication, a research pharmacy to dispense drugs, a nurse on hand to monitor for side-effects). We aimed for a larger-than-typical sample size (*N*=90) relative to most past studies (*Ns*~20-50) that have examined the psychological effects of propranolol (for reviews, see MacCormack, Armstrong-Carter, Gaudier-Diaz, et al., 2021; MacCormack, Armstrong-Carter, Humphreys, & Muscatell, 2021). We also used a within-person design (i.e., repeated measures of the two moral scenario types), which affords greater power to probe effects. Given the paucity of work on beta-blockers in moral judgment, we hope that this study provides valuable effect size estimates for future work.

One potential limitation is that the moral scenarios were rated two hours after drug administration, raising questions about the strength of propranolol effects in dampening visceral arousal at the time the moral judgment task was given. Perhaps the time between drug administration and task administration helps to explain the null direct effect of visceral arousal on moral judgment? This explanation is unlikely. The half-life of propranolol is five hours, providing ample opportunity for it to continue blunting visceral arousal and moral judgments at the time of task administration (Williams et al., 1986). Further, propranolol exerted a moderate-to-large effect on affective appraisals—showing that propranolol was impactful on some judgments—just not moral judgment directly. Furthermore, pre-ejection period—a reliable measure of sympathetic nervous system-driven cardiac arousal—was lower in the propranolol condition (vs. placebo) even one hour after drug administration. These results suggest that

propranolol was effective in reducing physiological arousal and that effects persisted. However, future work that includes ratings of moral judgments, harm, and affective appraisals closer to the time of propranolol administration will provide clarity on the strength of effects.

Another question is whether more evocative moral stimuli—such as videos of transgressions—would lead visceral arousal to directly impact judgment. Future studies could test which features (e.g., ambiguity, novelty) lead to greater inputs from visceral arousal and affective appraisals. In line with Dyadic Morality and work on moral template-matching (Ochoa, 2022), this proposal also implicitly suggests that people's moral templates—acquired from personal experience and cultural understandings of harm—will help determine what moral scenarios are "everyday" vs. "unusual." Past work argues that the content of moral violations varies by culture, while also suggesting at least some universality among this variation (Shweder, 2012). Future work could examine whether the role of affect is consistent or variable across cultures, in collaboration with local communities, anthropologists, and cultural psychologists.

We also note some theoretical limitations. Although we manipulated both the kind of moral scenarios (everyday vs. unusual purity) and visceral affect, we measured ratings of perceived harm, affective appraisals, and moral judgments cross-sectionally using self-reports, which prevents strong claims about the causal effects of perceived harm vs. affective appraisals in determining moral judgments. Nevertheless, past work has found individual strands of causal evidence for each of the AHA hypotheses (Ochoa, 2022; Schein & Gray, 2015), and we believe there is value in synthesizing these hypotheses together in this first formulation of the AHA.

Some may also take issue with how we divided up the realm of affect in moral judgment. For example, one could separate out a third category of affect beyond affective appraisals and visceral arousal, such as self-reports of emotional experience (e.g., "I feel gross"). Although these emotion self-reports could act differently than visceral arousal and the scenario-tied affective appraisals, we suggest that emotion self-reports may confound visceral arousal and appraisals.

Constructionist and constitutive appraisal models (e.g., affect-as-information) suggest that emotion self-reports reflect a mixture of "cognitive" affective appraisals and "embodied" arousal, whether integral to a stimulus (e.g., a scenario makes people feel arousal) or incidental (e.g., someone feels generally hungry, tired, or stressed). Of course, even the terms "embodied" and "cognitive" should be taken with a grain of salt, given our argument that affect and cognition are often intertwined. No matter how exactly these phenomena are categorized, our overarching goal is to have researchers reconsider their assumptions about sharp divisions *between* constructs (e.g., "affect" vs "harm") and homogeneity *within* constructs (e.g., affect).

In this paper, we focused on affective appraisals of disgust and did not consider how much an act is angering, distressing, saddening, etc. As a first foray into exploring the AHA, we focused on disgust given that it is the most often-discussed affective appraisal in moral psychology. However, future work that similarly blunts vs. exacerbates visceral affect should examine a greater diversity of affective appraisals and emotions, especially given the obvious importance of anger and outrage in moral judgment (Kahneman et al., 1998).

Finally, there are several important moderators that we did not consider here but that may prove fruitful in future extensions of the AHA. For example, the role of affect in moral judgment may be moderated by factors such as attentional control (Van Dillen et al., 2012), emotional differentiation (Cameron et al., 2013), and disgust sensitivity (Ong et al., 2014). Similarly, if visceral arousal helps potentiate affective appraisals in the context of moral judgment, then trait and state differences in interoceptive awareness and sensitivity may particularly warrant closer investigation. This possibility is borne out by findings in both Schnall et al. (2008) and Tracy et al. (2019) where private body consciousness (i.e., self-reported sensitivity to bodily sensations) significantly moderated effects.

Conclusion

Altogether, results suggest that modern harm-centric and affect-centric perspectives may both help explain moral judgment—and can be reconciled with each other. The Affective Harm

Account highlights the intertwining of harm and affect, clarifies how we should think about both harm and affect, and explains how differences in judgment across scenarios can co-exist within a parsimonious model of moral cognition. When it comes to harm and affect, rather than asking "which one," we sought to ask "when" and "how." Future research should do the same.

It may seem like we are returning to the past of moral psychology. It is true that we are advocating for a greater role of harm and a reduced role of affect in explaining moral judgment. We also suggest that we should understand different moral concerns (e.g., purity) more like Shweder's themes of ethical discourse (Shweder et al., 1997) and less like Haidt's (2012) modular "foundations." However, what may seem like "circling back" when viewed from directly above, can be seen as "spiraling upwards" when viewed from a greater distance. The AHA seeks to move beyond old debates of reasoned harm vs. gut feelings by relying less on traditional dichotomies—including "cognition vs. affect," "dyadic morality vs. disgust," and "harm vs. purity"—and instead provides a more parsimonious predictive framework of moral judgment in line with newer theories of morality and harm (Schein & Gray, 2018), and affect and visceral embodiment (Barrett et al., 2017; MacCormack & Lindquist, 2019; Storbeck & Clore, 2008).

Science progresses not only through running new experiments with innovative methods and developing and testing new theories, but also by reexamining old evidence, redefining old concepts, and reconsidering old assumptions. The Affective Harm Account integrates wisdom from the past, provides a new perspective on theories of the present, and serves as a parsimonious and predictive framework for the future of moral psychology.

Author's Note

References

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology*, *41*, 167–218.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Berridge, C. W. (2008). Noradrenergic modulation of arousal. *Brain Research Reviews*, *58*(1), 1–17. https://doi.org/10.1016/j.brainresrev.2007.10.013

Blair, R. J. (1995). A cognitive developmental approach to mortality: Investigating the psychopath. *Cognition*, *57*(1), 1–29. https://doi.org/10.1016/0010-0277(95)00676-p

Bryant, A. (1977). *The Anita Bryant story: The survival of our nation's families and the threat of militant homosexuality*. Revell.

Cameron, C. D., Lindquist, K. A., & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review*, 371–394. https://doi.org/10.1177/1088868314566683

Cameron, C. D., & Payne, B. K. (2011). Escaping affect: How motivated emotion regulation creates insensitivity to mass suffering. *Journal of Personality and Social Psychology*, *100*(1), 1–15. https://doi.org/10.1037/a0021643

Cameron, C. D., Payne, B. K., & Doris, J. M. (2013). Morality in high definition: Emotion differentiation calibrates the influence of incidental disgust on moral judgments. *Journal of Experimental Social Psychology*, *49*(4), 719–725. https://doi.org/10.1016/j.jesp.2013.02.014

Chapman, H. A., & Anderson, A. K. (2013). Things rank and gross in nature: A review and

synthesis of moral disgust. *Psychological Bulletin*, *139*(2), 300–327.

https://doi.org/10.1037/a0030964

Cheng, J. S., Ottati, V. C., & Price, E. D. (2013). The arousal model of moral condemnation.

Journal of Experimental Social Psychology, *49*(6), 1012–1018.

https://doi.org/10.1016/j.jesp.2013.06.006

Clifford, S., Iyengar, V., Cabeza, R., & Sinnott-Armstrong, W. (2015). Moral foundations

vignettes: A standardized stimulus database of scenarios based on moral foundations

theory. *Behavior Research Methods*, *47*(4), 1–21. https://doi.org/10.3758/s13428-014-

0551-2

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought.

Trends in Cognitive Sciences, *11*(9), 393–399.

https://doi.org/10.1016/j.tics.2007.08.005

Clore, G. L., & Ortony, A. (2013). Psychological construction in the OCC model of emotion.

Emotion Review, *5*(4), 335–343. https://doi.org/10.1177/1754073913489751

Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively

influences moral judgment and behavior through effects on harm aversion. *Proceedings

of the National Academy of Sciences*, *107*(40), 17433–17438.

https://doi.org/10.1073/pnas.1009396107

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D., & Robbins, T. W. (2008). Serotonin

modulates behavioral reactions to unfairness. *Science*, *320*(5884), 1739–1739.

https://doi.org/10.1126/science.1155577

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional

analyses in moral judgment. *Cognition*, *108*(2), 353–380.

https://doi.org/10.1016/j.cognition.2008.03.006

Cushman, F. (2015). Deconstructing intent to reconstruct morality. *Current Opinion in*

    Psychology, *6*, 97–103. https://doi.org/10.1016/j.copsyc.2015.06.003

De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing.

    Current Directions in Psychological Science, *28*(5), 503–509.

    https://doi.org/10.1177/0963721419855658

Decety, J., & Cacioppo, S. (2012). The speed of morality: A high-density electrical neuroimaging

    study. *Journal of Neurophysiology*, *108*(11), 3068–3072.

    https://doi.org/10.1152/jn.00473.2012

Decety, J., & Cowell, J. M. (2017). Interpersonal harm aversion as a necessary foundation for

    morality: A developmental neuroscience perspective. *Development and*

    Psychopathology, 1–12.

Decety, J., & Cowell, J. M. (2018). Interpersonal harm aversion as a necessary foundation for

    morality: A developmental neuroscience perspective. *Development and*

    Psychopathology, *30*(1), 153–164. https://doi.org/10.1017/S0954579417000530

Dungan, J. A., & Young, L. (2019). Asking 'why?' enhances theory of mind when evaluating

    harm but not purity violations. *Social Cognitive and Affective Neuroscience*, *14*(7), 699–

    708. https://doi.org/10.1093/scan/nsz048

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems.

    Proceedings of the National Academy of Sciences, *98*(24), 13763–13768.

    https://doi.org/10.1073/pnas.231499798

Epley, N., & Waytz, A. (2009). Mind perception. In S. T. Fiske, D. T. Gilbert, & G. Lindzey

    (Eds.), *The Handbook of Social Psychology* (5th ed., pp. 498–541). Wiley.

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in

    empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral*

    Reviews, *35*(3), 903–911. https://doi.org/10.1016/j.neubiorev.2010.10.009

Fiske, A. P., & Rai, T. S. (2014). *Virtuous violence: Hurting and killing to create, sustain, end, and honor social relationships*. Cambridge University Press.

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. *Journal of Personality and Social Psychology*, *113*(3), 343–376. https://doi.org/10.1037/pspa0000086

Gendler, T. S. (2008). Alief in action (and reaction). *Mind & Language*, *23*(5), 552–585. https://doi.org/10.1111/j.1468-0017.2008.00352.x

Gendron, M., & Barrett, L. F. (2009). Reconstructing the past: A century of ideas about emotion in psychology. *Emotion Review*, *1*(4), 316–339. https://doi.org/10.1177/1754073909338877

Gilligan, C. (1993). *In a different voice: Psychological theory and women's development* (29th ed.). Harvard University Press.

Graham, J. (2015). Explaining away differences in moral judgment: Comment on Gray and Keeney (2015). *Social Psychological and Personality Science*, *6*(8), 869–873. https://doi.org/10.1177/1948550615592242

Graham, J., & Haidt, J. (2010). Beyond beliefs: Religions bind individuals into moral communities. *Personality and Social Psychology Review*, *14*(1), 140–150. https://doi.org/10.1177/1088868309353415

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., Motyl, M., Meindl, P., Iskiwitch, C., & Mooijman, M. (2018). Moral Foundations Theory: On the advantages of moral pluralism over moral monism. In *Atlas of Moral Psychology* (pp. 211–222). Guilford Press.

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, *8*, 125–130. https://doi.org/10.1016/j.copsyc.2015.09.007

Gray, K., & Keeney, J. E. (2015). Impure or just weird? Scenario sampling bias raises questions about the foundation of morality. *Social Psychological and Personality Science*, *6*(8), 859–868. https://doi.org/10.1177/1948550615592241

Gray, K., Schein, C., & Cameron, C. D. (2017). How to think about emotions and morality: Circles, not arrows. *Current Opinion in Psychology*, *17*, 41–46.

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*(4), 1600–1615. https://doi.org/10.1037/a0036149

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*, 101–124. https://doi.org/10.1080/1047840x.2012.651387

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Grillon, C., Cordova, J., Morgan, C. A., Charney, D. S., & Davis, M. (2004). Effects of the beta-blocker propranolol on cued and contextual fear conditioning in humans. *Psychopharmacology*, *175*(3), 342–352. https://doi.org/10.1007/s00213-004-1819-5

Gutierrez, R., & Giner-Sorolla, R. (2007). Anger, disgust, and presumption of harm as reactions to taboo-breaking behaviors. *Emotion*, *7*(4), 853–868. https://doi.org/10.1037/1528-3542.7.4.853

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral

judgment. *Psychological Review*, *108*(4), 814.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*.

Pantheon Books.

Haidt, J., Bjorklund, D. F., & Murphy, S. (2000). *Moral dumbfounding: When intuition finds

no reason* [Unpublished].

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral

intuitions that liberals may not recognize. *Social Justice Research*, *20*(1), 98–116.

https://doi.org/10.1007/s11211-007-0034-z

Haidt, J., & Hersh, M. A. (2001). Sexual morality: The cultures and emotions of conservatives

and liberals. *Journal of Applied Social Psychology*, *31*(1), 191–221.

https://doi.org/10.1111/j.1559-1816.2001.tb02489.x

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate

culturally variable virtues. *Daedalus*, *133*(4), 55–66.

https://doi.org/10.1162/0011526042365555

Haidt, J., & Joseph, C. (2007). The moral mind: How five sets of innate intuitions guide the

development of many culture-specific virtues, and perhaps even modules. In P.

Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind* (Vol. 3, p. 31).

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A

scale sampling seven domains of disgust elicitors. *Personality and Individual

Differences*, *16*(5), 701–713. https://doi.org/10.1016/0191-8869(94)90212-7

Haidt, J., Rozin, P., Mccauley, C., & Imada, S. (1997). Body, psyche, and culture: The

relationship between disgust and morality. *Psychology & Developing Societies*, *9*(1),

107–131. https://doi.org/10.1177/097133369700900105

Han, X., Zhou, S., Fahoum, N., Wu, T., Gao, T., Shamay-Tsoory, S., Gelfand, M. J., Wu, X., &
     Han, S. (2020). *Cognitive and neural bases of collateral damage during intergroup
     conflict* [Preprint].

Harris, W. S., Schoenfeld, C. D., & Weissler, A. M. (1967). Effects of adrenergic receptor
     activation and blockade on the systolic preejection period, heart rate, and arterial
     pressure in man. *Journal of Clinical Investigation*, *46*(11), 1704–1714.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral
     and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hesse, E., Mikulan, E., Decety, J., Sigman, M., Garcia, M. del C., Silva, W., Ciraolo, C.,
     Vaucheret, E., Baglivo, F., Huepe, D., Lopez, V., Manes, F., Bekinschtein, T. A., & Ibanez,
     A. (2016). Early detection of intentional harm in the human amygdala. *Brain: A Journal
     of Neurology*, *139*(Pt 1), 54–61. https://doi.org/10.1093/brain/awv336

Hofmann, W., Brandt, M. J., Wisneski, D. C., Rockenbach, B., & Skitka, L. J. (2018). Moral
     punishment in everyday life. *Personality and Social Psychology Bulletin*, 1–15.

Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life.
     Science, *345*(6202), 1340–1343.

Horberg, E. J., Oveis, C., Keltner, D., & Cohen, A. B. (2009). Disgust and the moralization of
     purity. *Journal of Personality and Social Psychology*, *97*(6), 963–976.
     https://doi.org/10.1037/a0017423

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends
     in Cognitive Sciences*, *13*(1), 1–6. https://doi.org/10.1016/j.tics.2008.09.006

Inagaki, T. K. (2018). Opioids and social connection. *Current Directions in Psychological
     Science*, *27*(2), 85–90. https://doi.org/10.1177/0963721417735531

Johnson, D. J., Wortman, J., Cheung, F., Hein, M., Lucas, R. E., Donnellan, M. B., Ebersole, C.
     R., & Narr, R. K. (2016). The effects of disgust on moral judgments: Testing moderators.

Social Psychological and Personality Science, *7*(7), 640–647.

https://doi.org/10.1177/1948550616654211

Kahneman, D., Schkade, D., & Sunstein, C. (1998). Shared outrage and erratic awards: The

psychology of punitive damages. *Journal of Risk and Uncertainty*, *16*(1), 49–86.

https://doi.org/10.1023/A:1007710408413

Kleckner, I. R., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W. K., Quigley, K. S.,

Dickerson, B. C., & Barrett, L. F. (2017). Evidence for a Large-Scale Brain System

Supporting Allostasis and Interoception in Humans. *Nature Human Behaviour*, *1*.

https://doi.org/10.1038/s41562-017-0069

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to

socialization. In T. Mischel (Ed.), *Cognitive development and epistemology* (pp. 151–

235). Academic Press.

Kohlberg, L. (1981). *The philosophy of moral development: Moral stages and the idea of

justice*. HarperCollins.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in

linear mixed effects models. *Journal of Statistical Software*, *82*(13).

https://doi.org/10.18637/jss.v082.i13

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct

neural networks associated with directly experienced pain and empathy for pain.

NeuroImage, *54*(3), 2492–2502. https://doi.org/10.1016/j.neuroimage.2010.10.014

Landy, J. F., & Goodwin, G. P. (2015). Does incidental disgust amplify moral judgment? A meta-

analytic review of experimental evidence. *Perspectives on Psychological Science*, *10*(4),

518–536. https://doi.org/10.1177/1745691615583128

Le Guen, O., Samland, J., Friedrich, T., Hanus, D., & Brown, P. (2015). Making sense of

(exceptional) causal relations. A cross-cultural and cross-linguistic study. *Frontiers in

Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01645

Lindquist, K. A., & Barrett, L. F. (2008). The embodiment of emotion. In G. R. Semin & E. R. Smith (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches*. Cambridge University Press.

Long, J. A. (2019). *Interactions: Comprehensive, user-friendly toolkit for probing interactions*. https://cran.r-project.org/package=interactions

MacCormack, J. K., Armstrong-Carter, E., Humphreys, K. L., & Muscatell, K. A. (2021). Neurophysiological contributors to advantageous risk-taking: An experimental psychopharmacological investigation. *Social Cognitive and Affective Neuroscience*, *16*(9), 926–936. https://doi.org/10.1093/scan/nsab047

MacCormack, J. K., Armstrong-Carter, E. L., Gaudier-Diaz, M. M., Meltzer-Brody, S., Sloan, E. K., Lindquist, K. A., & Muscatell, K. A. (2021). Beta-adrenergic contributions to emotion and physiology during an acute stressor. *Psychosomatic Medicine*, *83*, 959–968. https://doi.org/10.1097/PSY.0000000000001009

MacCormack, J. K., Gaudier-Diaz, M. M., Armstrong-Carter, E. L., Arevalo, J. M. G., Meltzer-Brody, S., Sloan, E. K., Cole, S. W., & Muscatell, K. A. (2021). Beta-adrenergic blockade blunts inflammatory and antiviral/antibody gene expression responses to acute psychosocial stress. *Neuropsychopharmacology*, *46*(4), 756–762. https://doi.org/10.1038/s41386-020-00897-0

MacCormack, J. K., & Lindquist, K. A. (2017). Bodily contributions to emotion: Schachter's legacy for a psychological constructionist view on emotion. *Emotion Review*, *9*(1), 36–45. https://doi.org/10.1177/1754073916639664

MacCormack, J. K., & Lindquist, K. A. (2019). Feeling hangry? When hunger is conceptualized as emotion. *Emotion*, *19*(2), 301–319. https://doi.org/10.1037/emo0000422

Malle, B. F., & Guglielmo, S. (2011). Are intentionality judgments fundamentally moral? In C. Mackenzie & R. Langdon (Eds.), *Emotion, imagination, and moral reasoning (Macquarie monographs in cognitive science)* (pp. 275–293). Psychology Press.

Markus, H. R., & Kitayama, S. (2003). Culture, self, and the reality of the social. *Psychological Inquiry*, *14*(3/4), 277–283. JSTOR.

Marx, W., Ried, K., McCarthy, A. L., Vitetta, L., Sali, A., McKavanagh, D., & Isenring, L. (2017). Ginger—Mechanism of action in chemotherapy-induced nausea and vomiting: A review. *Critical Reviews in Food Science and Nutrition*, *57*(1), 141–146. https://doi.org/10.1080/10408398.2013.865590

Melnikoff, D. E., & Bargh, J. A. (2018). The mythical number two. *Trends in Cognitive Sciences*, *22*(4), 280–293. https://doi.org/10.1016/j.tics.2018.02.001

Mikhail, J. (2007). Universal moral grammar: Theory, evidence and the future. *Trends in Cognitive Sciences*, *11*(4), 143–152. https://doi.org/10.1016/j.tics.2006.12.007

Mills, P. J., & Dimsdale, J. E. (1991). Cardiovascular reactivity to psychosocial stressors: A review of the effects of beta-blockade. *Psychosomatics*, *32*(2), 209–220. https://doi.org/10.1016/S0033-3182(91)72094-X

Moore, J. W., Teufel, C., Subramaniam, N., Davis, G., & Fletcher, P. C. (2013). Attribution of intentional causation influences the perception of observed movements: Behavioral evidence and neural correlates. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00023

Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review*, *6*(4), 303–307. https://doi.org/10.1177/1754073914534477

Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal Theories of Emotion: State of the Art and Future Development. *Emotion Review*, *5*(2), 119–124. https://doi.org/10.1177/1754073912468165

Mueller, H. S., & Ayres, S. M. (1980). Propranolol decreases sympathetic nervous activity reflected by plasma catecholamines during evolution of myocardial infarction in man. *The Journal of Clinical Investigation*, *65*(2), 338–346. https://doi.org/10.1172/JCI109677

Newlin, D. B., & Levenson, R. W. (1979). Pre-ejection Period: Measuring Beta-adrenergic Influences Upon the Heart. *Psychophysiology*, *16*(6), 546–552. https://doi.org/10.1111/j.1469-8986.1979.tb01519.x

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review*, *9*(3), 184–211. https://doi.org/10.1207/s15327957pspr0903_1

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.

Ochoa, N. R. (2022). Template matching and moral judgment: A new method and empirical test. *Poetics*, 101643. https://doi.org/10.1016/j.poetic.2021.101643

Ong, H. H., Mullette-Gillman, O., Kwok, K., & Lim, J. (2014). Moral judgment modulation by disgust is bi-directionally moderated by individual sensitivity. *Frontiers in Psychology*, *5*. https://www.frontiersin.org/article/10.3389/fpsyg.2014.00194

Oosterwijk, S., Lindquist, K. A., Anderson, E., Dautoff, R., Moriguchi, Y., & Barrett, L. F. (2012). States of mind: Emotions, body feelings, and thoughts share distributed neural networks. *NeuroImage*. http://www.sciencedirect.com/science/article/pii/S1053811912005678

Open Science Collaboration. (2017). Maximizing the reproducibility of your research. In *Psychological Science Under Scrutiny* (pp. 1–21). Wiley-Blackwell. https://doi.org/10.1002/9781119095910.ch1

Pizarro, D., Inbar, Y., & Helion, C. (2011). On disgust and moral judgment. *Emotion Review*, *3*(3), 267–268. https://doi.org/10.1177/1754073911402394

Pizarro, D., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology*, *39*(6), 653–660. https://doi.org/10.1016/S0022-1031(03)00041-6

R Core Team. (2021). *R: A language and environment for statistical computing.*

> https://www.R-project.org/

Rottman, J., & Kelemen, D. (2012). Aliens behaving badly: Children's acquisition of novel

> purity-based morals. *Cognition*, *124*(3), 356–360.

> https://doi.org/10.1016/j.cognition.2012.06.001

Royzman, E. B., Kim, K., & Leeman, R. F. (2015). The curious tale of Julie and Mark: Unraveling

> the moral dumbfounding effect. *Judgment and Decision Making*, *10*(4), 18.

Royzman, E., & Borislow, S. H. (2022). The puzzle of wrongless harms: Some potential concerns

> for dyadic morality and related accounts. *Cognition*, *220*, 104980.

> https://doi.org/10.1016/j.cognition.2021.104980

Sagi, E., & Dehghani, M. (2014). Measuring moral rhetoric in text. *Social Science Computer*

> *Review*, *32*(2), 132–144. https://doi.org/10.1177/0894439313506837

Satpute, A. B., Kragel, P. A., Barrett, L. F., Wager, T. D., & Bianciardi, M. (2019). Deconstructing

> arousal into wakeful, autonomic and affective varieties. *Neuroscience Letters*, *693*, 19–

> 28. https://doi.org/10.1016/j.neulet.2018.01.042

Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of

> emotional state. *Psychological Review*, *69*(5), 379–399.

Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the

> same harm-based moral template. *Personality and Social Psychology Bulletin*, *41*(8),

> 1147–1163. https://doi.org/10.1177/0146167215591501

Schein, C., & Gray, K. (2018). The theory of dyadic morality: Reinventing moral judgment by

> redefining harm. *Personality and Social Psychology Review*, *22*(1), 32–70.

> https://doi.org/10.1177/1088868317698288

Schindler, I., Fagundes, C. P., & Murdock, K. W. (2010). Predictors of romantic relationship

> formation: Attachment style, prior relationships, and dating goals. *Personal*

> *Relationships*, *17*(1), 97–105. https://doi.org/10.1111/j.1475-6811.2010.01255.x

Schlottmann, A., Allen, D., Linderoth, C., & Hesketh, S. (2002). Perceptual causality in children. *Child Development*, *73*(6), 1656–1677. https://doi.org/10.1111/1467-8624.00497

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2008). Disgust as embodied moral judgment. *Personality and Social Psychology Bulletin*, *34*(8), 1096–1109. https://doi.org/10.1177/0146167208317771

Schnall, S., Haidt, J., Clore, G. L., & Jordan, A. H. (2015). Landy and Goodwin (2015) confirmed most of our findings then drew the wrong conclusions. *Perspectives on Psychological Science*, *10*(4), 537–538. https://doi.org/10.1177/1745691615589078

Schwarz, N., & Clore, G. L. (2003). Mood as information: 20 years later. *Psychological Inquiry*, *14*(3–4), 296–303. https://doi.org/10.1080/1047840X.2003.9682896

Shweder, R. A. (2012). Relativism and Universalism. In D. Fassin (Ed.), *A Companion to Moral Anthropology* (pp. 85–102). John Wiley & Sons, Ltd.

Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The "big three" of morality (autonomy, community, and divinity), and the "big three" explanations of suffering. In *Morality and Health* (pp. 119–169). Routledge.

Sinnott-Armstrong, W. (2016). The Disunity of Morality. In S. M. Liao (Ed.), *Moral Brains: The Neuroscience of Morality*. Oxford University Press.

Stern, R. M. (2002). The psychophysiology of nausea. *Acta Biologica Hungarica*, *53*(4), 589–599.

Storbeck, J., & Clore, G. L. (2008). Affective arousal as information: How affective arousal influences judgments, learning, and memory. *Social and Personality Psychology Compass*, *2*(5), 1824–1843. https://doi.org/10.1111/j.1751-9004.2008.00138.x

Strohminger, N., & Kumar, V. (Eds.). (2018). *The Moral Psychology of Disgust*. Rowman & Littlefield Publishers.

Terbeck, S., Kahane, G., McTavish, S., Julian, S., Levy, N., Hewstone, M., & Cowen, P. J. (2013). Beta adrenergic blockade reduces utilitarian judgement. *Biological Psychology*, *92*(2), 323–328. https://doi.org/10.1016/j.biopsycho.2012.09.005

Terbeck, S., Kahane, G., McTavish, S., Savulescu, J., Cowen, P. J., & Hewstone, M. (2012). Propranolol reduces implicit negative racial bias. *Psychopharmacology*, *222*(3), 419–424. https://doi.org/10.1007/s00213-012-2657-5

Thomson, M., Corbin, R., & Leung, L. (2014). Effects of ginger for nausea and vomiting in early pregnancy: A meta-analysis. *The Journal of the American Board of Family Medicine*, *27*(1), 115–122. https://doi.org/10.3122/jabfm.2014.01.130167

Tracy, J. L., Steckler, C. M., & Heltzel, G. (2019). The physiological basis of psychological disgust and moral judgments. *Journal of Personality and Social Psychology*, *116*(1), 15–32. https://doi.org/10.1037/pspa0000141

Triandis, H., Bontempo, R., Villareal, M., Asai, M., & Lucca, N. (1988). *Individualism and collectivism: Cross-cultural perspectives on self-ingroup relationships*. https://doi.org/10.1037/0022-3514.54.2.323

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge University Press.

van Dijk, E., van Beest, I., van Kleef, G. A., & Lelieveld, G.-J. (2018). Communication of anger versus disappointment in bargaining and the moderating role of power. *Journal of Behavioral Decision Making*, *31*(5), 632–643. https://doi.org/10.1002/bdm.2079

Van Dillen, L. F., van der Wal, R. C., & van den Bos, K. (2012). On the role of attention and emotion in morality: Attentional control modulates unrelated disgust in moral judgments. *Personality and Social Psychology Bulletin*, *38*(9), 1222–1231. https://doi.org/10.1177/0146167212448485

Waytz, A., Dungan, J., & Young, L. (2013). The whistleblower's dilemma and the fairness–
loyalty tradeoff. *Journal of Experimental Social Psychology*, *49*(6), 1027–1033.
https://doi.org/10.1016/j.jesp.2013.07.002

Wegner, D. M., & Gray, K. (2016). *The Mind Club: Who Thinks, What Feels, and Why It
Matters*. Viking.

Wheatley, T., & Haidt, J. (2005). Hypnotic disgust makes moral judgments more severe.
Psychological Science, *16*(10), 780–784.

Widen, S. C., & Russell, J. A. (2002). Gender and preschoolers' perception of emotion. *Merrill-
Palmer Quarterly*, *48*(3), 248–262. https://doi.org/10.1353/mpq.2002.0013

Wilkerson, M. D., & Hayes, D. N. (2010). ConsensusClusterPlus: A class discovery tool with
confidence assessments and item tracking. *Bioinformatics*, *26*(12), 1572–1573.
https://doi.org/10.1093/bioinformatics/btq170

Williams, F. M., Leeser, J. E., & Rawlins, M. D. (1986). Pharmacodynamics and
pharmacokinetics of single doses of ketanserin and propranolol alone and in
combination in healthy volunteers. *British Journal of Clinical Pharmacology*, *22*(3),
301–308. https://doi.org/10.1111/j.1365-2125.1986.tb02891.x

Wiseneski, D., & Skitka, L. (2017). Moralization through moral shock: Exploring emotional
antecedents to moral conviction. *Sage Journals*, *43*(2).
https://doi.org/10.1177/0146167216676479

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across
moral domains. *Cognition*, *120*(2), 202–214.
https://doi.org/10.1016/j.cognition.2011.04.005

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American
Psychologist*, *35*(2), 151.