

The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence

Decety, Jean ^{1 2}

¹ Department of Psychology and the Department of Psychiatry and Behavioral Neuroscience, The University of Chicago
² Paris Institute for Advanced Study

DOI [10.5281/zenodo.11001101](https://doi.org/10.5281/zenodo.11001101)

TO CITE

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). <https://doi.org/10.5281/zenodo.11001101>

PUBLICATION DATE

02/04/2024

ABSTRACT

Our moral sense is a biological adaptation specific to our species. It has evolved to facilitate group living, regulate social interactions, and promote cooperation beyond kinship. Beliefs intensely associated with moral values can motivate positive societal progress. However, moral convictions can also lead to dogmatism, antagonistic attitudes, a polarized political climate, and violent collective actions. This article draws on theories and empirical evidence from anthropology, neuroscience, cognitive science, social psychology, economics, and political science to explain why and how moralization increases the strength of beliefs and attitudes—in certainty and importance. Moral convictions are perceived as absolute, universal, and definite beliefs or principles. They engage the brain circuit associated with reward processing and valuation, which motivates and reinforces behavior. Moral commitments act as signals to conform to the rules of one's coalition, which causes predictable cognitive biases that benefit the self and the group. Moral convictions are held with great confidence. This leads to confirmation bias, which means that people are unwilling to seek or accept corrective information no matter how accurate the content of their beliefs. Once an issue is moralized, it becomes more challenging to engage in a cost-benefit analysis. A deeper understanding of attitudes, values, and moralization dynamics may allow social scientists and policymakers to identify and support pathways away from extremism and to strengthen and sustain engagement in benevolent forms of collective action.

Introduction

Do we respond similarly to news of petty crimes, police brutality, insurrection on the US Capitol, discriminatory social policies, or terrorist attacks? Some of these events betray basic values and principles, such as people's inherent worth and equality, respect for human rights, the importance of civil society, and the rule of law. This is the case

because humans are a judgmental bunch. Cross-culturally, since the dawn of humanity,

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). <https://doi.org/10.5281/zenodo.11001101>

2024/1 - paris-ias-ideas - Article No.9. Freely available at <https://paris.pias.science/article/the-power-of-moral-conviction-how-it-catalyzes-dogmatism-intolerance-and-violence> - ISSN 2826-2832/© 2024 Decety J.

This is an open access article published under the [Creative Commons Attribution-NonCommercial 4.0 International Public License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

people judge some behaviors are right or wrong and deserving rewards or punishments. Perceptions of morality are increasingly generalized as people inhabit larger and more unfamiliar social networks and interact with anonymous persons (Jackson et al., 2024). Moreover, and contrary to what many people believe, morality (kindness, civility, honesty, and basic human decency) is not declining. It is just the opposite, as shown by a study spanning 60 nations worldwide (Mastroianni & Gilbert, 2023). On average, we treat each other far better than our forebears.

Converging lines of evidence—from game theory, biology, anthropology, psychology, and economics—suggest that morality is a collection of evolved adaptations for promoting group living and cooperation (Curry et al., 2019; Enke, 2019). As an ultrasocial species, humans highly value teamwork and cooperation, and our collective orientation benefits us individually by helping gain coalition partners, mates, and friends. The constellation of thoughts and feelings that constitute a sense of morality has evolved to enable individuals to uphold cooperative social relations that maximize their biological benefits (Krebs, 2008). However, while the capacity for morality is universal, there is tremendous variability in which behaviors people think are immoral. Despite substantial individual and cultural variation, nearly all manifestations of morality involve, are based on, influence, and govern our relations with other people (Rai & Fiske, 2011). Furthermore, morality has evolved as a natural extension of our coalitional psychology (Tooby, 2017). As a result, it is strategically tuned to specific aspects of group dynamics. This cultural-evolutionary account explains why morality can lead to dogmatism, intolerance, and violence.

What people believe to be of moral value has a huge impact on society, especially by motivating participation in collective action (cooperative effort towards group status improvement) and inspiring the courage to oppose injustices despite personal cost. Many positive societal changes, such as the struggle to obtain the right to vote for women in the United States in the 1920s, the civil rights movement in the 1950s, and the legalization of abortion in the 1970s around many countries, were the result of people holding strong moral convictions and their tireless engagement in a collective effort.

At the same time, moral convictions can facilitate hostile, inflexible, and at times violent actions, ranging from the cancellation of a book launch ¹ to inter-ethnic riots

that leave scores of people dead. For better or for worse, moral convictions trigger more ardent reactions than other attitudes and beliefs. The extent to which a person considers a given topic (e.g., mining, the use of glyphosate, illegal immigration or the commercial use of body parts to take a few examples) to be within the scope of moral beliefs that are perceived as transcending the boundaries of persons, cultures, and contexts has wide-ranging and a variety of predicted negative interpersonal outcomes. These include dogmatism, intolerance of those who do not share the same views, unwillingness to compromise or accept procedural solutions to conflicts with little space for economic gain, and a host of other potentially harmful consequences.

This dark side of morality is the focus of this essay. Given that strong moral convictions are likely to be associated with acceptance of any means to achieve preferred goals (Skitka, 2010), a better understanding of why attitudes rooted in morality can promote disruptive forms of social engagement constitutes an important endeavor for interdisciplinary research with consequences to public policy.

In this article, I adopt a naturalistic perspective ², integrating empirical evidence and theories from multiple disciplines. Social phenomena like morality can be more comprehensively and accurately understood if their explanations maintain consistency and coherence across different levels of analysis. A vertical integration of knowledge across the biological sciences and the social sciences requires an effort to make them coherent and consistent (Cosmides, Tooby & Barkow, 1992). A biological perspective can inform social systems and provide a comprehensive understanding of social phenomena. Observations, knowledge, and concepts of higher and lower levels of analysis can mutually inform and calibrate those of other levels. Moreover, because of the emergent properties of complex organizations, it is unlikely that an understanding of higher-level psychological processes can be satisfactorily derived from neurophysiology alone (Berntson et al., 2012). Nor, for that matter, by a restricted analysis of any single level of organization or function, whether it be sociological, psychological, or neurobiological. Thus, these differing perspectives need to be integrated.

This article develops an integrative analysis of the empirical and theoretical literature based on a selection of articles from the social sciences and the biological sciences (Figure 1), to explain *why* and *how* the moralization process increases the strength of

beliefs and attitudes - in certainty and importance - which, in turn, motivates social commitment and induces dogmatism and attitudinal extremism at the individual level, regardless of ideological or political affiliation.

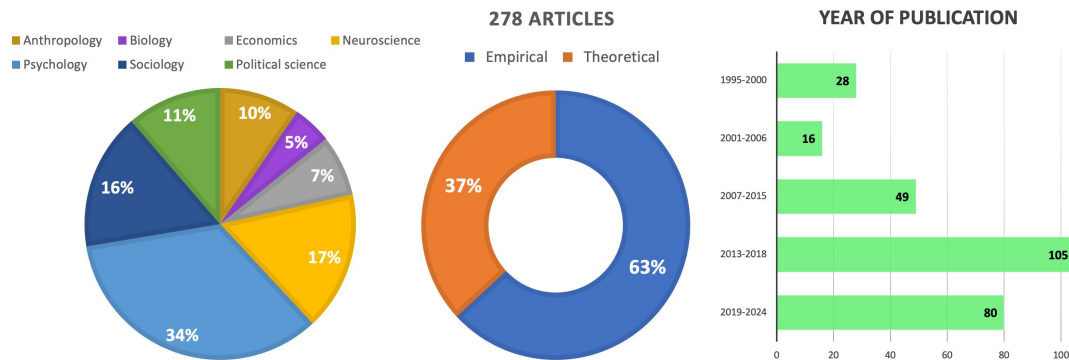


Figure 1: This integrative analysis includes 278 articles (67% published in the last decade) in anthropology, biology, neuroscience, psychology (developmental, cognitive & social), sociology, economics, and political science to explain *why* and *how* the moralization process increases the strength of beliefs and attitudes - in certainty and importance - which, in turn, motivates social commitment and induces dogmatism and attitudinal extremism at the individual level, regardless of ideological or political affiliation.

The article is organized into eleven sections. Following this general introduction, the second section explains why studying moral conviction matters. The third section presents a functionalist perspective on morality as it is currently described in evolutionary anthropology and psychology (developmental, cognitive, and social), followed by a section on the social domain theory of morality. The fifth section focuses on virtuous violence. The interwoven nature of morality and group dynamics, which stems from the selective pressures inherent in the evolution of our species is the subject of the sixth section. The cognitive architecture of moral convictions is outlined in the seventh section, followed by a section describing the moralization process and its neural mechanisms. Mental rigidity, a sense of superiority, and high subjective confidence are frequently associated with moral convictions. This class of phenomena, grouped under the umbrella term of metacognition, is the subject of the ninth. A final section before a general conclusion briefly addresses strategies to counter moralization. A glossary of the

terms and key concepts used in the paper (e.g. coalition, reward system, social influence, valuation) is provided before the list of references.

Overall, the understanding that emerges from integrating social sciences and biological science knowledge, demonstrates that moralization strongly increases the strength of attitudes –their certainty and importance– which in turn motivate political commitment, dogmatism, and ideological extremism, independent of the political sophistication of the actors and across the political spectrum. Moral conviction is the catalyst that turns beliefs into action, for better and for worse.

Why studying moral conviction matters

Moral convictions can inspire change and positive collective action but can also prompt dogmatism, intolerance, and societal divisions. Every day, we are bombarded with news reports of protests related to conflicts associated with social values. Some of them are extremely violent and result in the loss of human lives. As a dramatic example, in November 2002 in the city of Kaduna, Nigeria, around 250 people were killed in a series of religiously-motivated riots. The apparent trigger for the violence, which became known as the "Miss World riots," was an article published in a Lagos-based newspaper that was perceived as blasphemous by some Muslims. Within days, expressions of displeasure or offense at the article were seized upon by some militant groups, and the protests turned violent. Muslims attacked Christians and Christians retaliated against Muslims. Both groups went on a rampage, killing, burning, and looting. Not lethally violent, yet still concerning, decades ago, following the publication of Thornhill and Palmer's academic book on "A natural history of rape –biological bases of sexual coercion," which offered a different view on rape contrary to social constructivism that considers it an expression of male domination without sexual arousal, Joan Roughgarden vehemently responded in the journal *Ethology* (2004), arguing that *"critics of evolutionary and human-sexuality psychology should realize that they're dealing with a political fight more than an academic dispute. We must organize as activists to oppose this junk and get out of our safe comfortable armchairs, for much is at stake. Thornhill and Palmer are guilty of all allegations, and they deserve to hang."*³

Moralization can therefore inspire benevolent forms of collective action like the American civil right movement, but it can also incite dogmatism, intolerance, division, authoritarianism and harmful consequences (i.e., aggressive attitudes, justification of prejudice, vigilantism, and political violence) against people or groups who share different values or practices (Yoder & Decety, 2022).⁴ All are serious threats to personal autonomy, civil liberties and political rights. In the United States as well as a growing number of countries, liberals and conservatives are often incapable of reaching policy agreements on polarized social issues, from gun control to economic policy to health care and climate change. This partly stems from divides in fundamental values (Marie, Altay & Strickland, 2023). In August 2023, the number of Americans who believe the use of force is justified to restore Trump to the White House increased by roughly 6 million in the last few months to an estimated 18 million people, according to a survey conducted by the University of Chicago Project on Security and Threats.

Moralization frequently occurs in the public domain (e.g., about smoking, sexually transmitted diseases, vaccinations). It can be a positive force by signaling that an issue is morally important, and lead to good outcomes, but it can backfire and lead to psychological reactance, a motivational state of resistance, and bring about the opposite effect of the intended behavioral change (Kraaijeveld & Jamrozik, 2022). For instance, moralization of public health issues can legitimize stigmatization, ostracism and political division with counterproductive outcomes. In one study, female subjects were exposed to a stigmatizing versus non-stigmatizing health message with forceful versus non-forceful wording (Schnepper, Blechert & Stok, 2022). Then, the effects on a virtual food choice task (healthy versus unhealthy), diet intentions and concerns to be stigmatized were measured. The results indicate that in the non-stigmatizing and non-forceful condition, participants made the highest number of healthy food choices. Conversely, in the two stigma conditions, higher body mass index correlated with higher concern to be stigmatized, highlighting the adverse effect a health message can have. Opiates were not moralized in the past century, and morphine was a respectable medical treatment for alcohol addiction (Siegel, 1986). Today, the use of opiates is moralized and criminalized, and yet according to the National Center for Health Statistics over 106,699 drug-involved deaths (mostly opioids) were reported in 2021 in the US.

It is, therefore, important to understand the motivations, psychological mechanisms, and social variable that underlie and predict why and when moralizing can be dangerous. This knowledge is valuable in finding ways to facilitate tolerance between individuals, to encourage openness to alternative points of view, to reduce dogmatism, and to avoid legislative gridlocks. For example, the propensity in the United States to moralize attitudes on a variety of issues such as social justice, illegal immigration, vaccination, health care coverage, income inequality, or gender equality predicts greater social distance as well as prejudice, anger, incivility, and antagonism toward supporters who hold opposing beliefs (Garrett & Bankert, 2020). This polarization, which is largely affective, is not the direct cause of political violence. But it does help create an environment that allows opinion leaders to increase violence against politicians, election officials, women, and many types of minorities. It fosters an anti-establishment sentiment with a distrust of institutions and science. An editorial published in *Science*, one of the world's top academic journals, by a group of psychologists and sociologists sounded the alarm about political bigotry in the United States, calling it a poisonous cocktail of alienation, aversion, and moralizing that poses a serious threat to democracy (Finkel et al., 2020).

While morality binds us to others, allowing for more harmonious cooperation and coordination—undoubtedly essential foundations of our civilization, it can also blind us to distrust and demonize those who do not think, behave, or agree with us. At times, morality can motivate and justify intolerance and violence. Some people commit acts of violence because they sincerely believe that it is the morally right thing to do. In the minds of many perpetrators, violence can be a morally necessary and appropriate way to regulate social relationships according to precepts and cultural prototypes. These moral motivations apply equally to violent heroes of the Iliad as they do to contemporary environmental activists who destroy agricultural crops in opposition to genetically modified organisms, death threats against two Australian philosophers who suggested in an academic paper that newborns and fetuses are morally equivalent "potential persons" whose family's interests override theirs, rioters of religious and ethnic rights through India and Africa, or parents who use corporal punishment to discipline their children.

Dogmatism, intolerance, and resulting physical, verbal, or symbolic violence motivated by moral conviction are thus serious concerns in today's world. Moralized identification

with one group/coalition leads to polarized attitudes, identity-based ideology (whether it's gender, sexual orientation/identification, religion, race, ethnicity, or socioeconomic status), and sectarianism. The more individuals are certain of a belief—regardless of its objective correctness—the more durable it will be, and the more willing they are to express their opinion and shun others who do not accept their views (Tormala & Rucker, 2015), and polarize the political debate. Social media, currently used by 5 billion people, amplifies polarizing and divisive messages, especially for moralized content (Rathje et al., 2021). Such a combination is potentially toxic and undermines our democracies, as well as our capacity to harmoniously live together and find pragmatic solutions to current challenges. It seems rational to strongly defend a view about facts that one evaluates as beyond doubt, like the earth is round or viruses evolved faster than any other living organisms. Distinguishing the epistemology of factual and normative beliefs is important to consider.

A functionalist perspective on morality

Moral thinking pervades our lives, and this is specific to our species. The human predisposition for social preferences⁵ and cooperation has emerged through natural selection in part due to the benefits it conferred on our ancestors living in large groups. This requires complex systems of social evaluation to distinguish individuals who can be trusted, who are likely to cooperate from those who are not. Our species has the propensity to produce cultural organizations that create social norms (regulation of acceptable and standard ways of behaving and achieving goals) and enforce them through institutions designed to assess the acceptability of individuals' behaviors and assign appropriate punishments to those who violate social norms (Tomasello, 2016).

A large body of literature in anthropology, psychology and economics indicates that our species has evolved interlocking sets of values, virtues, norms, practices, identities, institutions that work together to suppress or regulate self-interest and make cooperative societies possible (Haidt, 2011). These adaptations are thought to be the product of gene-culture coevolution (Bowles, 2016; Boyd, 2018; Gintis, 2011; Henrich, 2016). All human societies develop social norms and systems of social regulation, which owe their

effectiveness to the ability of individuals to integrate, share, and respect them as much as have them respected by others. This enables stability in living together and promotes large-scale cooperation (Boyd, 2018). Morality regulates interpersonal exchange, facilitates survival, coexistence, and cooperation, minimizes aggression, and more generally provides a balance when individual interests conflict with collective interests (Curry et al., 2009). While it is true that many non-human animals collaborate with their own species, humans are exceptional in their social preferences. They help one another based not only on kinship and reciprocity (direct and indirect), but also on apparently unselfish motives driven by altruism, empathy, and benevolence. Humans seek means for all to benefit through enacting norms to promote and maintain fairness, equity, and justice within the community. Social preferences are associated with happiness and well-being (Iwasaki, 2023). Many studies have reported a correlation, and some have found a causal reciprocal relationship between behaving prosocially and happiness (Aknin & Whillans, 2021; Meier & Stutzer, 2008).

We are motivated by morality because it is advantageous at the individual level –a non-zero-sum game (Delton & Krasnow, 2015)⁶. These moral concerns are not located in an abstract world characterized by ivory tower speculation. We are inherently and deeply social animals. Nearly all manifestations of morality involve, build upon, influence, and often govern our relationships with others. The ability to think and act in accordance with moral norms is a hallmark of our species. Of course, we are not conscious of the ultimate explanation of behaviors (the *why* question), those concerned with fitness consequences.

A strong argument in favor of the evolutionary origins of morality comes from developmental research demonstrating that the foundational aspects of social evaluation and precursor of morality are present early in ontogeny, before input from socialization (Decety, Steinbeis & Cowell, 2021; Tomasello, 2014). Preverbal infants possess an innate system of social evaluation that is similar to what evolutionary biologists, anthropologists, and psychologists have posited to be required for the emergence of our species' sociability (Hamlin, 2013). They distinguish between positive and negative social interactions, positively evaluate those who cooperate and negatively evaluate those who do not, expect other people to prefer those who have helped versus hindered them, and consider mental states when evaluating others (Decety, 2019; Cowell & Decety,

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). <https://doi.org/10.5281/zenodo.11001101>
 2024/1 - paris-ias-ideas - Article No.9. Freely available at <https://paris.pias.science/article/the-power-of-moral-conviction-how-it-catalyzes-dogmatism-intolerance-and-violence> - ISSN 2826-2832/© 2024 Decety J.
 This is an open access article published under the [Creative Commons Attribution-NonCommercial 4.0 International Public License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

2015; Hamlin, 2015; Ting et al., 2020). Before their second year, toddlers direct their own antisocial behaviors appropriately, selectively taking resources from someone who has previously hindered a third party (Hamlin, Wynn, Bloom, & Mahajan, 2011), and showing concern for issues of justice and fairness (Ting et al., 2020).

Children as young as 15 months perceive the difference between equal and unequal distribution of food, and their awareness of equal rations is linked to their willingness to share a toy (Schmidt & Sommerville, 2011). Subsequent studies with infants of the same age demonstrate that they use fairness concerns to guide their social selections (Burns & Sommerville). However, infants also take into consideration the race of individuals, and the consequences of the behavior of these individuals for their own- versus other-race individuals. When the distributor's race (Caucasian vs. Asian) is pitted against prior fair behavior, infants no longer systematically select the fair distributor, suggesting that infants incorporate information about the individuals' races when making social selection, and weigh race and fairness as competing dimensions in their social decision.

Ingroup favoritism – the tendency to favor members of one's own group over those of other groups emerges early. For instance, research indicates that 17-month-old infants possess an abstract expectation of ingroup support that already guides their reasoning about how individuals will act toward others (Jim & Baillargeon, 2017). When an individual needed help and another individual from the same minimal group was present, infants expected this second individual to support her ingroup member; no such expectation arose when the two individuals were not members of the same group. Infants around 14 months old begin to show cooperative behaviors, such as spontaneously and indiscriminately helping others (Warneken & Tomasello, 2007). This disposition is then influenced by judgments of likely reciprocity and their concern about how they will be evaluated by others (reputation) (Engelmann & Rapp, 2018). In costly sharing and prosocial tasks, when sharing incurs no cost, 5-year-olds treat kin and friends more favorably than strangers, while 6-year-olds favor kin over friends when sharing resources incurs a cost, showing kin favoritism (Lu & Chang, 2006).

In their second year, infants possess context sensitive-expectations relevant to fairness. They expect an experimenter to give a reward to each of the two individuals when both have worked to complete an assigned chore, but not when one has done all the work

while the other was playing (Sloane et al., 2002). Across cultures, young children show a desire to appear fair in resource distribution and are more willing to distribute resources equitably with age, suggesting a universal increased preference for equity over equality throughout development (Huppert et al., 2019). A rudimentary understanding of the normative dimension of human actions can be observed in 18-month-old toddlers (Schmidt, Rackoczy & Tomasello, 2019). Normative behavior develops around 3 to 5 years of age, without explicit teaching. Young children of this age can distinguish prescriptive from descriptive social rules (Nucci, 2001; Smetana, 2013; Sripada & Stich, 2006). Around the same time, children express obligatory judgments based on moral concerns with others' welfare, rights, and fairness through spontaneous reactions and reasoning about perceived violations, and begin to enforce social norms on others (Killen & Smetana, 2015).

Morality incorporates multiple dimensions, including sensitivity to social norms, values, reputation, and the set of capacities involved in reinforcement learning (including reward and punishment-based decision-making). It involves nonconscious and implicit processes such as an aversion to harming the vulnerable, concern for others, social emotions (e.g., guilt, shame, outrage, gratitude, and pride), as well as more conscious processes such as theory of mind and deliberative reasoning (Decety & Cowell, 2018; Van Bavel et al., 2015). Although moral values are personal in the sense that they are internalized, they are acquired in a social context and are shared with others from our social group (Schwartz, 2014). The mechanisms by which beliefs are conferred and become shared include the propagation of social norms about common goals and the establishment of moral principles or standards of conduct that should be followed.

Social groups play a key role in the establishment of prescriptive social norms. Importantly, decisions and behaviors that uphold moral values are perceived as highly rewarding (Bartels et al., 2015; Yoder & Decety, 2018). Moreover, judgments about moral norms are highly affected by social conformity. In keeping with the large body of work showing how decisions and behaviors are influenced by majority opinions, some studies have demonstrated that moral judgments are subject of conformity pressure. In one study, participants were asked to make decisions about sacrificial dilemmas either alone or in a group of confederates (Kundu & Cummins, 2013). Under social conformity, permissible actions were deemed less permissible and impermissible actions

were judged more permissible. Another study found that while subjects do tend to conform to the majority opinion on moral matters, they do so in a selective manner by conforming more when the majority opinion is of a deontological nature than when it is of a consequentialist nature (Bostyn et al., 2017). This asymmetric conformism effect is theorized by the authors to be driven by strategic concerns related to mutualistic partner choice preferences.

Across studies in social psychology and cognitive neuroscience, it is well documented that the opinion of others modulates the reward and value systems. In particular, the magnitude of activity in the ventral striatum reflects the value of reward-predicting stimuli (Campbell-Meiklejohn et al., 2010). Social influence mediates very basic value signals in known reinforcement learning circuitry and explains the swift spread of values throughout a group of individuals. The reward circuit generates signals related to a broad spectrum of social functioning, including making decisions influenced by social factors, learning about other members of our communities, and following social norms (Bhanji & Delgado, 2013).

Morality is generally thought to be based on injunctive and descriptive norms about how people should behave toward each other (Killen & Smetana, 2015). Descriptive norms are simply regularities of behavior, what most people do in a given situation, while injunctive norms are behavioral expectations supported by social or material sanctions (Simpson & Willer, 2015). A substantial body of evidence shows that people conflate what is descriptively normal (common or frequent) with what is prescriptively normal (allowed or required), reflecting the common-is-moral association (Heyes, 2024). When 35,000 participants from 30 European countries were asked to evaluate the morality of questionable behaviors (e.g., cheating on tax declaration or having casual sex), there was a positive correlation between their ratings of the frequency and justifiability of the behavior (Eriksson et al., 2021). Injunctive norms have an effect. For instance, studies conducted among Christians-American have found that participants donated more money to charity (Malhotra, 2010) and watched less porn on Sundays (Edelman, 2009). However, they compensated on both accounts during the rest of the week. Likewise, a study conducted in Morocco found that whenever the Islamic call to prayer was publicly audible, local shopkeepers contributed more money to charity (Duhaime, 2015).

However, these effects were short-lived. Donations increased only within a few minutes of each call and then dropped again.

Numerous other studies have yielded similar results. People become more generous in a donation task (Xygalatas et al., 2016) and cooperative in a bargaining game (Xygalatas, 2013) when they found themselves in a place of worship. Importantly, this effect is not due to intrinsic religiosity (the disposition) but an artifact of contextual cues (the situation).

In addition, one's own survival and success depend on trustworthy, competent, and motivated collaborative partners (Noë & Hammerstein, 1994). Thus, within a cultural group, individuals are concerned with how others evaluate them and invest energy to maintain the appearance of being fair and keep a good reputation (Caviola & Faulmüller, 2014). Conformity to the prevailing practices and social norms is vital to advertise group identity. It is plausible that morality is the product of cultural evolution over a potentially shorter period, by selecting cognitive and motivational mechanisms specialized for processing social norms (Heyes, 2024). Her cultural-evolutionary model proposes that norm psychology depends on implicit domain-general processes (perceptual, attentional, learning, motivation) that are genetically inherited as well as explicit domain-specific processes that are culturally inherited. Such capacities and practices have evolved over hundreds of thousands of years since humans first banded together for fitness and survival.

Moreover, some evolutionary anthropologists have argued that moral judgment is routinely strategic (DeScioli et al., 2014). Human societies have a variety of rules, and some benefit some people more than others. This panoply of moral rules with differential effects is part of the enduring social ecology of *Homo sapiens*, and natural selection has favored cognitive adaptations for advocating rules that enhance the individual's fitness (Mercier, 2011). Group living has persistently constituted a fundamental aspect of our evolutionary history, and thus selection has favored psychological adaptations to enable and regulate behavior both within and between groups (Forgas et al., 2007).

Overall, morality has evolved as a regulatory system that includes the creation and sensitivity to social norms, notions of justice, fairness, and rights, and a meta-perception

of some of our decisions, judgments, and attitudes that can vary in intensity. Overall social and moral norms facilitate a harmonious and cooperative co-existence, especially when individuals have a high degree of conformity.

Social domain theory of morality

Over decades of research in psychology, it has become clear that what people subjectively experience as moral is psychologically different from what they experience as social preferences or conventions. Across cultures and very early in development, people classify behaviors, values, beliefs, and practices into distinct categories (Smetana, 2013; Nucci, 1996; Turiel, 2006). The moral domain includes actions considered inherently wrong because of their detriments on the well-being of others, such as physical and psychological harm, violations of property, or damage to equity. The moral domain thus brings together prescriptive judgments of justice, rights, and welfare considerations relating to how people should treat each other. Moral judgments are impersonal, generalizable, and motivate the possibility of sanction to enforce respect for people and their rights. The conventional domain includes actions that are not inherently wrong but whose wrongfulness depends on the existence of a rule or norm (such as how to dress for a wedding or a funeral), and these rules and judgments of the conventional domain vary according to time and place. They are, therefore, not universal. Judgments that fall into this category are justified by reference to the maintenance of social norms and respect for authority. Finally, the personal domain encompasses preferences and actions whose consequences do not affect other people or society (the private sphere). Though people disagree about what falls into each of these three domains, once a belief, value, or practice has been categorized into one of them, the categorization consistently and powerfully predicts people's responses to it, especially when it differs from their own (Wright, 2023). Overall, the wrongness of moral transgressions is seen as stemming from their intrinsic features, such as their intrusion on others' rights and welfare (Smetana, 2013). Morality is seen as normatively binding, and thus, moral rules are hypothesized to be unalterable. However, depending on the culture, moral principles may carry different weight, and the boundaries between domains are fuzzy. Depending on historical epochs, social ecology (population density, mobility, pathogen avoidance,

weather), religious beliefs, and institutional rules (e.g., kinship structure and economic markets), each society develops a moral system that emphasizes specific orientations (Bentahila et al., 2021). Furthermore, there are cross-cultural differences in the tightness of social norms that impact the enforcement of moral rules and punishments (Gelfand et al., 2017).

Another theoretical perspective proposes that moral beliefs and attitudes are a matter of degree rather than just a matter of kind. The moral significance that people attach to different issues or problems varies according to historical time, culture, and geography (Skitka et al., 2021). Attitude toward smoking, for example, has evolved from a matter of preference to increasing moralization over the past 60 years. Similarly, there was a time when abortion was unrestricted by law in the United States, and abortion services were openly marketed. Abortion restrictions in the United States were not initially grounded on moral concerns, but were rather rooted in concerns about medical licensure and the desire of increasingly professionalized healthcare providers to stem competition from midwives and homeopaths (Reagan, 2022). Attitudes toward abortion vary widely across cultures (Ryan, 2014). Today, however, abortion has become a highly moralized and politically polarizing issue in the United States, mostly for white evangelical Christians. They consider abortion to be morally wrong, and argue that it should be illegal. However, relatively few Americans actually view the morality of abortion in stark terms. According to a 2022 Pew Research Center report, 7% of all U.S. adults say abortion is morally acceptable in all cases, and 13% say it is morally wrong in all cases. A third of them say that abortion is morally wrong in most cases, while about a quarter (24%) say it is morally acceptable most of the time. About an additional one in five do not even consider abortion a moral issue.

In this theoretical context, morality is not an essential characteristic of certain domains of decisions, choices, judgments, or attitudes. Rather, it is a meta-perception that people have of some of their decisions, beliefs, judgments, and attitudes, which can vary in strength. Morality is a matter of degree rather than a matter of kind.

The fact remains that our capacity for morality is universal. It must be understood by distinguishing a moral capacity and moral values: while our moral capacity is universal, the moral rules and values we develop are quite diverse. They vary according to times,

places, cultures, history, and our choices. For example, bullfighting is considered a barbaric and immoral practice of animal torture in many countries where it is therefore prohibited, while it is perceived as a cultural heritage in Portugal, Spain, Venezuela, and southern France where it receives public subsidies. In 2020, same-sex relationships were punishable by law in 69 out of 193 countries and carry the death penalty in 11 countries. Currently, gay marriage is legal in 34 countries.

How do we reconcile the universality of our moral capacity and the plurality of our morals? Contrary to what is too often believed in social sciences, a naturalist perspective on morality is not necessarily reductive. It does not eliminate levels of organization. Rather it seeks conceptual unity and relations that cut across levels. Complex phenomena require accounts at multiple levels of organization. Different levels must be mutually compatible, but no one level reduces to another (Barkow, 1991). In fact, this perspective can help us understand our moral capacity's natural basis without limiting its diversity. Here, the usual dichotomy between nature and nurture is misleading and unnecessary. It is no longer defended by most evolutionary/cultural anthropologists. All humans possess an evolved ability to create, learn, and uphold moral rules and social norms and adapt them to new situations. Therefore, moral cognition is itself universal, whereas moral rules differ from group to group and within the same group over time.

Morality corresponds to a set of principles and norms that guide the conduct of individuals within their group. While some moral principles seem to transcend time and cultures, (e.g., fairness or family obligations), moral values are generally not set in stone. Almost all manifestations of morality involve, rely on, influence, and govern our relationships with others.

Virtuous violence

Violence is generally and wrongly considered as the antithesis of sociality. It is usually seen as an expression of our animalistic nature, which is expressed when learned cultural norms break down or when our ability to control and inhibit ourselves is deficient. Violence is also perceived as the essence of evil, the prototype of immorality.

It manifests in many different ways such as a reflex behavior (reactive aggression), coercion, threat of physical force, hostility, and terror, and thus needs to be determined by assessing intentionality. Studies in behavioral ecology and anthropology indicate that humans have a high potential for proactive aggression (involving some degree of premeditation) toward their fellow humans, including lethal violence, an attitude we share with other primates, and chimpanzees in particular (Gómez et al., 2016). Furthermore, this potential for violent aggression is much higher in men than in women, likely due to an unusually high benefit/cost ratio for intraspecific aggression in the former (Georgiev et al., 2014). Overall, human aggression is characterized as a combination of low propensities for reactive aggression and coercive behavior and high propensities for proactive aggression (especially coalitionary proactive aggression). These tendencies are associated with the evolution of groupishness, self-domestication, and social norms (Sarkar & Wrangham, 2023).

Most psychological theories of morality hold that at their core, moral judgments always include a prohibition against unprovoked killing, harming, stealing, and lying (Turiel, 2012; Gray et al., 2012). Support for violence can only be construed as a moral violation, an error in moral performance, or a necessary evil to achieve a greater good. Yet, historically, harming others, even family members, has been considered morally laudable, consistent with coalitional psychology. A striking illustration can be found in the Old Testament (or Torah), Exodus 32:28, where Moses commands, “Let each one of you lay down his sword to one side; cross and go through the camp from gate to gate, and let each kill his brother, his relative”. About three thousand people perished that day for committing the sin of worshiping a golden calf. Punishing one's own children for disobedience, sometimes very severely, is seen not only as a necessary evil but as a virtuous action in many parts of the world. This story in Exodus is consistent with the common observation that people respond aggressively, with anger, to those who act in ways that threaten their group's identity, particularly when they affiliate strongly with their group. Experiments in social psychology reliably show that behaviors that violate moral norms when displayed by ingroup as compared to outgroup members are associated with ingroup-directed hostility through collective shame, psychological distress, and experience of threat to shared values (Fousiani et al., 2019; Marques et al., 1988). When a threat to one's group identity comes from the inside, that is, from an ingroup member, people react with hostility and respond by treating the ingroup

transgressor very negatively (Marques & Yzerbyt, 1988). This happens because a norm violation by ingroup members jeopardizes the reputation of the ingroup and increases ingroup identity threat (Van der Toorn et al., 2015).

Likewise, many people feel morally obligated to physically and violently punish certain transgressions committed by individuals from an outgroup. Two polls of American public opinion found that support for disproportionately killing civilians and enemy combatants, with nuclear or conventional weapons, was deeply divided along partisan political lines (Slovic et al., 2020). Those who approve of these excessively lethal attacks generally adhere to conservative values. They feel socially distant from the enemy and therefore believe that the victims are responsible for their fate. These same people also tend to support national policies that protect gun ownership, limit abortion, and harshly punish immigrants and criminals.

Across cultures and history, violence is used with the intention of maintaining order and can be expressed through war, torture, genocide, homicide, or violent conflict between ethnic groups (Fiske & Rai, 2014). Anthropological, evolutionary, and sociological analyses indicate that many people believe that violence can be just and that victims of violence were the cause of what happened to them—the husband who avenges the murder of his wife; the vigilante who punishes criminals; the soldier who kills his enemy; the public who votes for capital punishment; family members who torture and kill a young girl because of whom she dates, how she dresses or her refusal to submit to forced marriage; and even the suicidal terrorist who detonates a bomb—all evoke violence as morally justified, obligatory and praiseworthy (Atran, 2010; Boyd et al., 2003; Black, 1983). People may commit acts of extreme violence because they sincerely believe it is an imperative, the right thing to do. This is the case of vigilantism, an organized effort by a group of "ordinary citizens" to uphold norms and maintain public order on behalf of their coalition (Asif & Weenink, 2022). In the minds of these actors, such violence is a necessary and legitimate means of regulating social relations in order to restore the integrity of moral imperatives according to principles, precedents, and cultural prototypes specific to their group.

Anthropologists Rai and Fiske (2011) propose that moral judgments are not independent of the socio-relational contexts in which they occur. Our moral intuitions are defined by

the relations in which they appear. Intentionally harming others will be perceived as more or less acceptable, even morally laudable, depending on the relational and cultural context. These prejudices range from verbal aggression to large-scale ethnic conflict. A given action is deemed good, just, equitable, honorable, pure, virtuous, or morally correct when it occurs in specific socio-relational contexts. It is also be deemed wrong or unjust when it occurs in other socio-relational contexts. Guided by this theoretical framework, Fiske and Rai (2014) proposed that any action, even violent, can be morally just insofar as it regulates the socio-relational context defined by culture, whether it is within a family, ethnic group, military unit, religion or nation. Actions that violate the relational model that actors use are considered immoral. For relationships to work, people need competing motivations that lead them to regulate and sustain social relationships by controlling their own behavior and sanctioning others. Therefore, the core of our moral psychology consists of grounds for evaluating and orienting our judgments and behaviors, and those of others (including speech, emotions, attitudes, and intentions) with reference to prescriptive models. Behaviors that do not conform to relational prescriptions are considered moral transgressions and arouse emotions such as guilt, shame, disgust, envy, or indignation. These emotions motivate sanctions, including apologies, repairs and rectifications, self-punishment, and modulation or termination of the relationship.

The existence of groups and the social processes associated with them lead to solidarity from the group and disapproval from without under conditions where differences between groups become salient. Also, strong moral beliefs and values are often linked to greater defense when threatened. This can be explained by both social identity theory (Tajfel and Turner, 1986) and coalition psychology (Boyer, 2020).

A stark example is the religious fundamentalism of which the French became sorely aware of after the Charlie Hebdo massacre in Paris on January 7, 2015, where 17 people who worked in the offices of the satirical magazine and a Jewish supermarket lost their lives, shot by Islamic militants. In a study of more than 52,000 people in 59 countries, Muslims reported a greater number of fundamentalist beliefs than followers of other religions, including Protestants, Catholics, Hindus, and Jews (Wright, 2016). The link between fundamentalism, dogmatism and hostility is well-established. The higher the level of fundamentalism, the greater the hostility towards an outgroup (Henderson-

King et al., 2004; Johnson et al., 2012; Rothschild et al., 2009). However, it is unclear whether fundamentalism is a direct cause of or exacerbates reactions to social identity threats. Among most of the Muslim publics polled by the Pew Research Center in July 2011, an important variable emerged: Muslims tend to identify with their religion, rather than their nationality. This is particularly true in Pakistan, where 94% think of themselves primarily as Muslim instead of Pakistani. This is also the case in Israel, Jordan and Turkey. What makes religion unique apart from generic discussions of social identity is a belief in the supernatural, the meaning of this belief to the individual and the group, and the internalization and integration of religious identity to the individual. However, all aspects of religion that are associated with group hostility, aggression, and violence have direct secular counterparts and thus are not unique to religion (Wright & Khoo (2019). For instance, a series of studies across political and religious contexts examined the association between religion and popular support for suicide attacks (Ginges et al., 2010). The findings show that the relationship between religion and support for suicide attacks is real, but orthogonal to devotion to religious belief in general. The association between religion and suicide attacks seems to be a function of collective religious activities that facilitate popular support for suicide attacks and parochial altruism more generally.

Secular groups also display dogmatic beliefs and fundamentalist values and react violently to threats to them. In the 80's, loosely organized groups fought against a host of issues, including logging, drift-net fishing, nuclear energy, whaling, and road construction (Eagan, 1996). Their actions were motivated by environmental necessity, leading to uncompromising attitudes and violent actions. Most of their destructive acts targeted property: the spiking of trees to present hazards to loggers who would cut the trees, the dismantling of an electrical transmission tower, and the sinking of ships involved in whaling and drift net fishing.

In North American public and private universities, an alarming number of students approve of the use of violence and find it acceptable to use physical force to silence someone with whom they disagree. Several incidents have taken place on campus, illustrating how moralized and polarized opinions can lead to verbal abuse or physical bullying. For example, a political science professor at Middlebury College in Vermont was injured in the neck by a group of students who opposed a lecturer she had invited

to campus. At Reed College in Oregon, students interrupted lectures and bullied a humanities professor who they perceived to be too “Eurocentric” in her class on ancient Greece.

More recently, in several European countries, demonstrations against genetically modified organisms (GMOs) with the destruction of many experimental plots have been motivated by beliefs and values of safeguarding biodiversity. This anti-GMO attitude is not really grounded in the objective scientific knowledge of biology, as indicated by a study conducted in the United States, France and Germany with representative samples (Fernbach et al., 2019). Those with the most extreme anti-GMO attitudes turned out to be the least scientifically knowledgeable. This overconfidence and certainty in judgments and beliefs, combined with cognitive simplicity (i.e., seeing things black or white) makes it easier for people who are extreme in their political attitudes, regardless of political affiliation, to perceive their beliefs as moral absolutes (actions are intrinsically right or wrong, regardless of context or consequence) that reflect an objective, universal truth (van Prooijen & Krouwel, 2019).

Such strong beliefs about what is right or wrong are usually shared within a group or a community, consolidating them and providing them with a unity of motivation that can induce intolerance and at times, extreme actions, including giving one's life for one's group. This is the case of suicide bombers (Decety, Pape & Workman, 2018), ordinary citizens risking their lives for a cause, like an activist in France going on hunger strike to prevent the cutting of trees needed to widen a road (Box 1 for a dramatic example), or less dramatically, refusing any discussion and public debate with other people, including experts, who have a divergent opinion, as we see more and more often with specific communities of activists on the transgender movement (Habib, 2019; Grossman, 2023). The reason often invoked in that context is that debate is hate speech. Restrictions on free expression are demanded in the name of preventing offense. In 2021, transgender activists threatened to beat, rape, assassinate and bomb Harry Potter author J. K. Rowling on Twitter for comments she made about transgender women using women's restrooms. In the US, several children's hospitals have become targets of far-right attacks and harassment over gender-affirming medical care for transgender youth.

Box 1: Sacrificing her life to protest the release of a comedy film.

On October 18, 1973, 35-year-old Danielle Cravenne, armed with a handgun, successfully hijacked an Air France flight from Paris to Nice to protest the release of "*The Mad Adventures of Rabbi Jacob*," one of the first films to showcase the orthodox Jewish community. The passionately pro-Palestinian woman who thought that the film was too pro-Israeli launched this desperate act to cancel its release. On board the Paris-Nice flight, Cravenne threatened to destroy the plane if the film was not banned and requested that the plane be directed to Cairo, Egypt. She died on the plane that same day, shot by a tactical police operator disguised as a maintenance worker.

In 1968, Danielle Bâtisse converted to Judaism to marry Georges Cravenne, a French producer and advertising executive. In 1973, he promoted Gérard Oury Tannenbaum's film *Les Aventures de Rabbi Jacob*. The film was released on October 18, 1973. Danielle vehemently disagreed with the film's release at a time when a conflict in the Middle East between Arabs and Israelis, known as the Yom Kippur war, took place. The film deals with relationships between Arabs and Jews and has a happy ending about tolerance towards other religions and cultures. Danielle thought that the film was indecently pro-Israel. According to her friends, "the Arab-Israeli conflict was the trigger for everything in her. She could no longer live in a certain way and couldn't bear the Palestinian minority being persecuted. She was a bit of an idealist."

Danielle Cravenne's moral valuation of the rights of the Palestinian people trumped her Jewish social identity and relationship with her husband. This is a strong illustration of the power of moral conviction.

Hatred has been characterized as the negative evaluation of a target that is linked with a moral judgment, and this emotion is rooted in seeing the hated person as morally deficient or as a violation of moral norms (Royzman et al., 2006; Staub, 2004). People are more inclined to engage in attack-oriented actions when they experience hate as

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). <https://doi.org/10.5281/zenodo.11001101>

2024/1 - paris-ias-ideas - Article No.9. Freely available at <https://paris.pias.science/article/the-power-of-moral-conviction-how-it-catalyzes-dogmatism-intolerance-and-violence> - ISSN 2826-2832/© 2024 Decety J.

This is an open access article published under the [Creative Commons Attribution-NonCommercial 4.0 International Public License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

compared to anger, contempt, and dislike (Martinez et al., 2022). One study analyzed the linguistic content of 150 online websites of hate groups in North America, like Stormfront.org, which is one of the world's oldest hate groups—providing an Internet forum for over 300,000 registered users who are affiliated with neo-Nazi and White supremacist groups (Pretus et al., 2023). The results show that hate groups use more moral language in expressing their beliefs as compared with users who complain on online forums, used as controls.

Counterintuitively, violence is not antithetical to morality. It can be a means to justify a virtuous end like in cultures of honor or when restoring social order. Similarly, self-sacrificial behavior, including giving one's life for a cause, is generally motivated by strong moral conviction.

Coalitional dynamics and morality

Humans are an ultra-social species. This sociality is not a happy coincidence but a biological adaptation: a survival response to evolutionary pressures (Cacioppo et al., 2010; von Hippel, 2018). The degree of sociability between individuals is directly related to the survival and reproductive strategies that dictate the success of a species. When the benefits of living together outweigh the costs of living alone, animals tend to form groups (Alexander, 1974). A phylogenetic comparative analysis of ~1000 mammalian species on three states of social organization (solitary, pair-living, and group-living) and longevity shows that group-living species generally live longer than solitary species, and that the transition rate from a short-lived state to a long-lived state is higher in group-living than non-group-living species (Zhu et al., 2023). Living in groups allows for better protection against predators, facilitates access to resources (information, food, sexual partners), and in humans promotes communal parental care, cooperation, and social learning. Evolutionary and ecological pressures toward individualism and competition are met by the countervailing pressures placed on us by our role in cooperative and interdependent groups. We were successful on the savannah (and everywhere else) because we evolved to develop the capacity to work incredibly well as a team and cooperate in sharing information. For this reason, teamwork and

cooperation are highly valued in potential friends and mates, and our collective orientation benefits us individually by helping us gain coalition partners (von Hippel, 2018). Throughout evolution, humans have relied not only on the support of their relatives, but also that of genetically unrelated individuals. Humans have evolved a range of psychological mechanisms that promote an attraction to and capacity for living in groups. The consequence of this obligatory interdependence is that the building blocks of human psychology –emotion, motivation, cognition– have been shaped by the demands of social interdependence (Baumeister & Leary, 1995). Humans evolved in the context of intense intergroup competition, and groups comprised of loyal members more often succeeded than nonloyal ones (Richerson et al., 2010). Mathematical modeling of social evolution combined with empirical evidence from anthropology and behavioral ecology, reveal that the motivation to be invested in their own members' welfare co-evolved with intergroup competition over resources. An optimal condition for genetically encoded hyper prosociality to propagate is when groups are in conflict. In line with cultural group selection theory (Richerson, Boyd & Henrich, 2010), it has been proposed that during the late Pleistocene, groups with higher numbers of prosocial individuals cooperated more effectively and thus outcompeted others (Marean, 2015). As a non-zero-sum game, our morality is grounded in coalitional psychology with competition-cooperation dynamics. Our cooperative dispositions were superimposed onto human psychology—without eliminating those older social adaptations that favor friends and kin—and this renders us susceptible to an inherent conflict between other-regarding and public-minded concerns, as well as gives us more selfish and nepotistic dispositions that other species are spared with (Richerson & Boyd, 2005).

Selection pressures have thus sculpted human minds to be tribal (i.e., a social group sharing a common interest) and fostered adaptations to regulate within-coalition cooperation and between-coalition conflict in what, under ancestral conditions, was a fitness-promoting mechanism (Clark et al., 2019; Kurzban, Tooby & Cosmides, 2001). Specific abilities and motivations have been selected by evolution to enable individuals to act effectively within groups, and at times compete against other groups. Group loyalty and associated cognitive biases exist in all human societies, and explain how they can distort certain beliefs in favor of one's coalition. This is clearly the case in politics, and can be found among both conservatives and liberals in the USA (Ditto et al., 2019). Intragroup bias is not limited to a preference for members of one's own

group but encompasses multiple domains of cognition (Boyer, 2020; Petersen, 2015). For example, people do not retain information about their own group and others in the same way. They are far more attentive to disagreements between members of their own group than with members of other groups.

Group dynamics is reflected at a primary sensory level. Perceiving ingroup members suffering causes greater activation of brain regions associated with empathy compared to perceiving outgroup members suffering (Contreras-Huerta et al., 2013; Decety, Echols & Corell, 2010; Decety, 2021). Interaction with members of other groups also has a host of physiological impacts, including cardiovascular, hormonal and stress-related consequences (Boyer, Firat & van Leeuwen, 2015). Identity fusion, a visceral sense of oneness with a group (or group mind), occurs when personal and group identities collapse into a single identity to generate a collective sense and unique destiny (Swann & Buhrmester, 2015). This motivates costly pro-group behavior. Experiments conducted with undergraduate students indicate that strongly fused individuals endorsed self-sacrifice to save others (Swann et al., 2014). When they recognized that their group members were in danger, strongly fused persons became highly emotional and immediately felt as if they themselves were in danger and it did not matter whether one or several ingroup members were threatened.

Social groups—whether small bands of friends, entire nations, gangs, tribes, or unions—exist only because individuals have reasons to participate in them over the long term (Boyer, 2018). Because this reciprocity heuristic is always activated when it comes to real groups, people spontaneously apply it to all group situations. In behavioral economics games, when participants learn that they will receive nothing from others in return for the gifts they have given them, the bias in favor of the group disappears (Yamagishi & Mifune, 2009). When people behave in a way that seems to favor their group members, it's because they are implicitly using a social exchange heuristic, a set of assumptions about the task before them (evaluating different individuals or allocating resources to them) as a form of reciprocal cooperation (Boyer, 2018).²

Coalitions can be characterized as sets of individuals interpreted by their members and others as sharing a common social identity, including propensities to act as a unit, defend common interests, and have shared mental states. People are motivated to help

the status of their coalition rise, and will feel increased entitlement to better treatment as it does (Petersen, Osmundsen & Tooby, 2022). They will be motivated to fight any threat to the status of their coalition but will feel lower entitlement if its status falls. This provides a functional account of the intuition that group attitudes provide intangible feelings of self-esteem. An inflated belief in self-superiority is prone to encountering threats and causing violence (Baumeister et al., 1996). This coalitional psychology expresses itself in group psychology, politics, and morality (Tooby & Cosmides, 2010). Thus, it should come as a surprise that the surface contents of moral values function as signals of coalition coordination within groups rather than doctrines selected for their intrinsic appeal (DeScioli & Kurzban, 2018). Moral content is often selected to communicate the emergence of a new coalition or to morally legitimize attacks on rivals based on pretexts arising from the superficial properties of those rivals' moralities. Indeed, people support moral projects not because they have intrinsic appeal, but because of their downstream effects on rivals—for example, by lowering their status or weakening their social power.

From that perspective, morality is not a domain of content—moral values are heterogeneous and contradictory—and group affiliations and coalitions shape the perception and evaluation of the events that concern them and motivate moral judgments accordingly. In line with this theory, a large body of empirical literature shows that individuals are more likely to cooperate and reciprocate with their group partners than strangers (Böhm, et al., 2020). At the proximal level, cooperation is highly rewarding and is associated with increased neural activity in brain areas involved in reward processing, including the nucleus accumbens, caudate nucleus, and ventromedial prefrontal cortex (Rilling et al., 2012). Activation of this neural network reinforces reciprocal altruism by motivating participants to resist the temptation to selfishly accept, but not reciprocate, favors. Similar results were found in an fMRI study when participants played a game in cooperation with another person in comparison to playing in competition against someone else (Decety et al., 2004).

Religion is a source of social identity in many parts of the world. Religion provides particularly powerful narratives and conveys a sense of security, stability, and shared values that unify the social fabric. It delivers powerful social signals for cooperation or conflict by making group boundaries explicit and salient. Religious identity marker

signals influence perception in an implicit and rapid way. For example, a study conducted in China included participants of Han ethnicity, who are similar in terms of facial features (Huang & Han, 2014). Some of these participants were Christians, and some were atheists. Visual evoked potentials were recorded by electroencephalography (EEG) while the participants perceived faces of this ethnicity either with a neutral expression or with an expression of pain. These faces were marked (with a symbol attached to their collars) as Christians or Atheists. Participants explicitly reported greater discomfort and less empathy for the faces of people who hold different religious beliefs. Religious/non-religious identification significantly modulated the amplitude of evoked potentials (ERPs) as early as 200 ms after the onset of the appearance of the face expressing pain compared to neutral expressions. The amplitude of this signal was significantly stronger when participants perceived the faces of people with whom they shared the same religious affiliation.

A single-word label indicating a person's religious affiliation (Hindu, Christian, Jewish, Muslim, Scientologist, or Atheist), without any prior interpersonal interaction or additional information, is sufficient to modulate neural activity in the observer and its direction (positive or negative), and can be predicted from their religious affiliation (Vaughn et al., 2018). In this fMRI study, the neural response in the network involved in emotional empathy was significantly greater when participants perceived a painful event occurring on a hand labeled with their own religion rather than a hand labeled with a different religion. Furthermore, the magnitude of this bias is proportional to the participants' level of dispositional empathy. The more the participants have an empathetic personality, the more they will exhibit a negative bias towards people in the outgroup.

When making moral decisions after considering a hypothetical crime, people are less willing to report a member of their group than an outsider, despite knowing that the person is guilty (Lee & Holyoak, 2020). Even when the situation is unambiguous and the guilt of their sibling or ally is clear, participants still often refuse to report him to the police. Experiments examining moral obligations showed that people who are impartially prosocial are viewed as less moral and less trustworthy precisely because they are seen as not fulfilling special obligations to their family or group (McManus et al., 2020).

The great religious traditions emphasize the responsibility of believers towards their fellow human beings, by encouraging benevolence and preaching caring for others. It is often said that religious beliefs arouse compassion toward the disadvantaged, which logically can be extended to the attitude toward migrants. However, religion's propensity to elicit compassion is not immune to the phenomenon of social identity and group dynamics. In general, people are more empathetic and prosocial towards members of the group with which they identify, whether it is ethnic, sporting (Fourie et al., 2017) religious, or political (Saroglou et al., 2004). In this latter case, empathy towards members of one's group strongly promotes polarization and intolerance towards people who have other affiliations (Simas et al., 2020). Specifically, religious beliefs are a powerful motivator for compassion, pious people typically direct their kindness toward loved ones and co-religionists (Norenzayan 2014).

Using two priming experiments conducted among American Catholics, Turkish Muslims, and Israeli Jews, Bloom and her colleagues (2015) were able to distinguish the role of religious social identity and religious belief in attitudes toward immigrants. The authors found that religious social identity plays an important role in the integration of immigrants. In the study sample, religious social identity increases opposition to immigrants who do not resemble group members in terms of religion or ethnicity, while religious belief engenders welcoming attitudes toward immigrants from the same religion and ethnicity, especially among less conservative devotees.

The essential function of morality is to regulate social life. It is adapted to the type of group it regulates and is sensitive to relational violations. Thus, it is not paradoxical that any action— even a violent one—can be perceived morally just insofar as it regulates the social-relational context defined by culture (Fiske & Rai, 2014). Furthermore, and consistent with coalitional psychology theory and the social heuristic exchange theory,⁸ a large body of research demonstrates people's tendency to derogate deviant ingroup members as a means to protect the group from the threat that they pose to their social identity, coining it the black sheep effect. People display a stronger preference for (specifically) exclusionary punishments against deviant ingroup members through the experience of increased ingroup threat (Fousiani et al., 2019). Third-party experimental economics punishment games demonstrate that non-cooperative in-group members are more harshly judged and severely punished than out-group members, demonstrating the

black sheep effect (Shinada et al., 2004). The black-sheep effect is most likely to occur when group-based motivational concerns are activated—for example, when one strongly identifies with the group, perceives a threat to its reputation, or believes that its members are seen as similar to each other and when a group-specific norm has been violated (Jordan et al., 2014). Immoral behaviors by ingroup rather than outgroup members jeopardize the group's reputation and therefore activate utilitarian (i.e., exclusion-oriented) motives for punishment. Conversely, in-group members who prove their loyalty when a threat is present receive the most positive evaluations (Branscombe et al., 1993).

Political orientation modulates whether people feel inspired and uplifted after watching a video of large-scale protests demanding racial equity in policing. Participants' political orientation influenced whether they experienced moral elevation while watching a video of large-scale protests for racial equity or a counter-protest video (Holbrook et al., 2023). Conservatives experienced elevation in response to the Back the Blue (BtB) video, (protesters from a counter-movement-BLM, who support the police), while liberals experienced elevation in response to the Black Lives Matter video. Furthermore, the state of elevation experienced by participants in response to the videos was associated with costly behaviors, such as expression in their preferences regarding funding for the police. Elevation evoked by the BLM video was related to a preference for defunding the police, while elevation evoked by the BtB video was related to a preference for increasing police funding. These findings indicate that people's political attitudes and coalitional affiliation influence their emotional responses and preferences regarding allocating funds to policing and social services. This shows that the same actions or policies can be experienced as moral, and as emotionally moving, in opposite directions depending on our group biases.

Morality is in the eye of the beholder. Prosocial emotions may drive feelings toward enemies we intuitively perceive as driven by hate. Understanding their moral sentiments and motivations may be helpful when negotiating with—or even strategizing against—members of opposing groups.

In one study, college students were first surveyed about their political affiliation and moral convictions about various sociopolitical issues (e.g., Black Lives Matter, abortion

rights, foreign aid). Then, while in the scanner they were requested to rate the appropriateness of photos depicting violent protests in the United States that were ostensibly in line or not with their moral beliefs (Workman, Yoder & Decety, 2020). The results indicate that the appropriate violence-judgments were associated with a parametric increase in hemodynamic signal within the vmPFC and the ventral striatum (Figure 2). Furthermore, the more the violence is perceived as legitimate, the less the amygdala that encodes emotionally salient information and the prefrontal regions engaged in deliberative reasoning react to these stimuli. The neural circuitry that mediates reaction to congruent (“my-side”) and incongruent (“other-side”) engages circuits involved in the calculation of value and cost-benefit analysis that underpins decision-making.

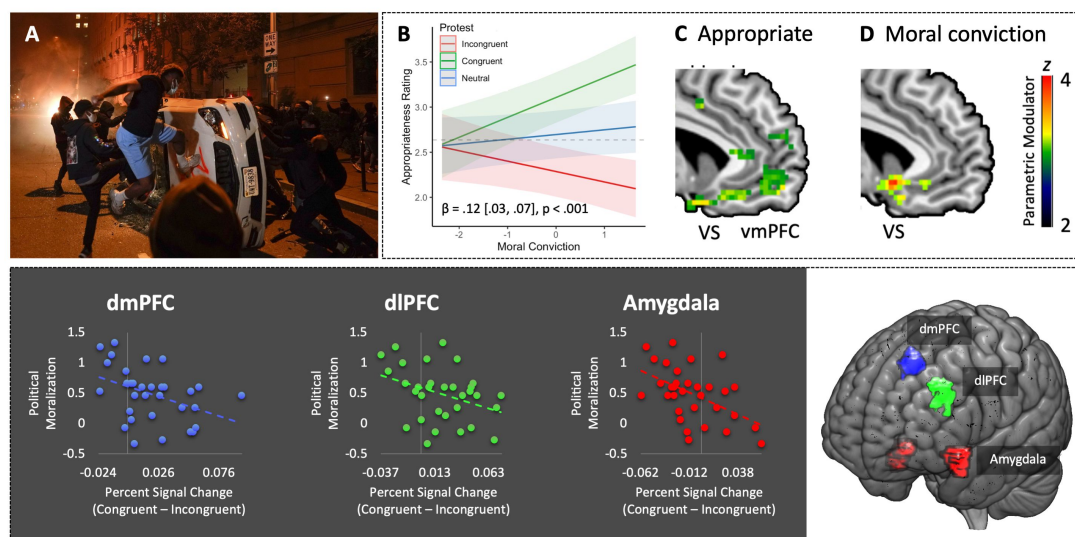


Figure 2: Moral convictions shape the perception and evaluation of violence in the brain circuit, implementing the coding subjective reward value. In the MRI scanner, participants were presented with photographs of violent riots (A) which were either congruent or incongruent with their moral beliefs and political attitudes (B). Parametric increases in neuro-hemodynamic signal in the ventromedial prefrontal cortex (vmPFC) and ventral striatum (VS) predicted evaluations of congruent causes (C) and the intensity of participants' moral convictions (D). The more moralized and legitimate the violence was perceived, the greater a reduction in signal was detected in the dorsomedial prefrontal cortex (dmPFC), dorsolateral prefrontal cortex (dlPFC) and amygdala. Adapted from Workman, Yoder & Decety, 2020).

The activity of striatal neurons reflects reward anticipation, and those in vmPFC encode subjective and motivational values. This makes it possible to establish a link between the cognitive and affective representations necessary to guide social decision-making.

Increased hemodynamic activity is detected in the ventral striatum and vmPFC when

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). <https://doi.org/10.5281/zenodo.11001101>

2024/1 - paris-ias-ideas - Article No.9. Freely available at <https://paris.pias.science/article/the-power-of-moral-conviction-how-it-catalyzes-dogmatism-intolerance-and-violence> - ISSN 2826-2832/© 2024 Decety J.

This is an open access article published under the [Creative Commons Attribution-NonCommercial 4.0 International Public License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

participants assess, perceive, or even imagine performing positive moral actions versus negative moral actions (Decety & Porges, 2011; Yoder & Decety, 2014ab). This circuit also exhibits signals related to the decision to punish a person who has acted immorally, as well as the observation of a punishment inflicted on someone who has acted unfairly in economic games or who violates social norms (Crockett et al., 2013; De Quervain et al., 2004; Stallen et al., 2018). Activation of the ventral striatum and increased functional connectivity of that region with the orbitofrontal cortex is detected in participants enjoying watching mixed martial arts (MMA), an extremely physically violent sport (Porges & Decety, 2013). MMA fighters are real people; the injuries they suffer in the cage and during practice are often severe. Yet, observers derive pleasure and satisfaction from watching their pain. Notably, the striatum receives converging inputs from the cerebral cortex, thalamus, and amygdala, as well as dopamine signals that guide reinforcement learning, the learned associations between a particular state, and the set of actions with the highest expected values (Krauzlis, Bollimunta, Arcizet & Wang, 2014). This will guide the person's behavior. It supports the view that attention arises as a byproduct of circuits centered on the basal ganglia involved in value-based decision-making.

In another study, a Thomas sample of Pakistani Muslims supporting the cause of Kashmir was selected on their beliefs and support for the jihadist group fighting for the integration of this region in Pakistan, then invited to participate in an fMRI study during which they were asked to communicate their willingness to fight and die for a series of values related to Islam (e.g., Sharia should be applied in all Muslim countries) and current international politics on a Likert scale of 7-points (Hamid et al., 2019). As in the study by Workman and colleagues (2020), decisions to support the fight for Islamist values are associated with a parametric increase in activity within the vmPFC and a decreased activity within the dlPFC, as well as weaker functional connectivity between the vmPFC and the dlPFC. These results suggest that beliefs to fight and die for Islamist values engage brain regions associated with subjective value coding (vmPFC) rather than material cost integration (dlPFC) during decision-making, supporting the idea that decisions about costly sacrifices motivated by moral values are not to be mediated by a cost-benefit calculation.

Together, the results of these functional neuroimaging studies indicate that strong moral beliefs about sociopolitical or economic issues increase their subjective value, outweighing the natural aversion to interpersonal wrongs. The more moral convictions are associated with non-negotiable, inviolable, and non-utilitarian values, values that resist trade-offs with economic values (Baron & Spranca, 1997), also called sacred values in political science (Tetlock, 2003), the more they correlate with activity in vmPFC in subjects passively exposed to statements that relate to beliefs and attitudes that they have moralized (Pincus et al., 2014). A study combined, in the same participants, measures on a multidimensional political science data reduction scale to investigate the perception and evaluation of political candidates and fMRI when these participants were asked to read and evaluate 80 political statements (Zamboni et al., 2009). The radical (political) dimension of participants' responses is associated with a parametric increase in activity within the ventral striatum and vmPFC. These regions modulate the online expression of pro-violence attitudes, which is consistent with previous work linking this brain circuit to moral beliefs and political radicalism.

When socioeconomic and political issues are highly moralized, individuals to whom they are presented exhibit a pattern of cortical information processing that is prioritized over weakly moralized issues at several stages including early automatic attention. This is reflected by the components of ERPs, especially medial frontal negativity and early posterior negativity (Yoder & Decety, 2022). Furthermore, these quick responses predict a reduction in social conformity (i.e., being concerned with what other people may think). Changes in alpha and beta spectral power indicate increased attention and engagement with moralized content. Together, these electroencephalographic results indicate that strong moral conviction alter the susceptibility to social influence by altering cognitive and emotional processing to prioritize information that relates to previously moralized beliefs. They align with research on attitude strength showing that strong attitudes are resistant to change and are easily accessible (Young & Fazio, 2013).

Coalitional affiliation shapes appraisals of events concerning these groups and motivates emotional responses, rewards and value processing and resulting moral judgment accordingly, especially in intergroup conflicts.

The functional architecture of moral conviction

Beliefs and attitudes that are held with strong moral convictions are psychologically distinct from other beliefs. Moral conviction refers to a strong and absolute belief that something is right or wrong, moral or immoral (Bauman & Skitka, 2009). In addition, when a moral belief is perceived as grounded in objective facts, independent from contexts and cultures, it implies that there is only one right answer as to whether something is morally right or wrong, good or bad. Such absolute moral realism, combined with a greater perception of consensus within one's community or alliance, is associated with feelings of superiority, an inflexible mindset, and intolerance (Wright & Pözler, 2021). Framing an action as moral leads to more extreme and faster judgments (Van Bavel et al., 2012).

Moral conviction can be defined as a set of implicit and explicit representations that incorporates two distinct dimensions: cognitive (moral vs. nonmoral) and emotional (strong vs. weak intensity). A series of experiments conducted by Wright and colleagues (2008) indicate that the cognitive dimension is sufficient to predict most of the negative interpersonal responses to divergent moral attitudes, including greater intolerance, less sharing, and greater distancing from people with divergent attitudes. Emotional intensity, on the other hand, plays a motivational force when combined with moral beliefs, exacerbating their influence (Dancy, 1993; Días, 2023).

Strong moral convictions manifest in the qualities of durability and impact, which are indicators of attitude strength (Krosnick & Petty, 2014). The first aspect of durability is the persistence or stability of the belief, reflecting the degree to which it remains unchanged over time. The second aspect of durability is resistance, which refers to the ability to withstand and react⁹. Similarly, the attitudinal impact influences information processing and judgments in the sense that certain kinds of information come to mind quickly and render certain decisions. Thus, strong moral convictions are more likely to bias information processing activity and judgments that are weak (Petty & Cacioppo, 1981). Finally, moral convictions can guide behavior, and strong ones are more likely to do so than weak ones.

Because people are objectivist when judging moral beliefs to be true or false, they tend to be intolerant of what they consider to be morally deviant behavior, regardless of the preferences of others, and with little room for compromise (Skitka, 2010). As noted earlier, moral beliefs held with strong moral conviction are seen as non-negotiable and objective universal mandates, dictating what everyone “should” do on the justification it is for the greater good. People experience strong negative reactions when others violate these expectations and may resort to violence against those who question or do not share their views or values. The 19 young perpetrators of the 9/11 attacks were responding to perceived threat to their core moral values (i.e., encroaching secularism and Western values and globalization that were threatening their extreme fundamentalist beliefs) when they developed moral mandates, which led to their suicidal actions (Skitka & Mullen, 2002). Individuals who blow up abortion clinics or participate in violent protests against political reforms or for animal welfare may have very different political leanings but are fundamentally similar in several general psychological dimensions (Thomas et al., 2019). Furthermore, strong moral convictions have essential social and political consequences, such as increased intolerance of others with different beliefs, difficulty in resolving conflicts, increased political participation, polarization of attitudes, a willingness to accept violent means to achieve goals that correspond to the values associated with one's beliefs, punishment of those who violate the norms and immunity to habitual pressures to obey authority or follow the law (Thomas et al., 2019).

All over the world, people are increasingly polarized. This polarization is exacerbated by moralization. As individuals perceive their political attitudes as an expression of their moral convictions, the bitter division once confined to the political arena seeps into nearly every issue of social importance, be it the Black Lives Matter movement or veganism. In the United States, divisions over social and economic issues—government, race, stem cell research, immigration, national security, environmental protection, gender identity, and social justice—are reaching record levels of polarization with irreconcilable views. Perceived politicization undermines public support for institutions (e.g., police, criminal justice system, public health, public media, etc.), willingness to defer to their expertise, and increased skepticism toward academic disciplines (Clark et al., 2023).

Issues that people see as moral are more likely to be closed to compromise and are particularly linked to motivations to engage politically, either to proactively defend what

they believe to be right and good, or to fight reactively against what they believe to be fundamentally wrong.

When people who are morally convinced are confronted with societal attitudes that do not align with their moral values, some may find it intolerable enough to justify violent actions against those who challenge their beliefs and values, in order to send unambiguous signals to members of their group and members of other groups (Pauls et al., 2022). Virtue signaling is a powerful force for social regulation by creating common knowledge around a moral issue that people would otherwise ignore. This view aligns well with theoretical models in evolutionary psychology, emphasizing the importance of morality in regulating the behavior of individuals in contexts of group dynamics and in coalition processes (Boyer, 2020).

Moral beliefs provide the psychological underpinnings and justification for many attitudes that can become polarized and can motivate extreme actions intended to achieve morally desirable ends. Moral convictions have several specific characteristics (1-4) and consequences (5-6) (Figure 3).

1. **Objectivity and universality:** Perceiving moral beliefs as objective facts, grounded on a fundamental truth or reality and universally applicable across time and places increases people's inclination to render judgments in absolute evaluations (Van Bavel et al., 2012). They show higher attitudinal intolerance and dogmatism than individuals who are anti-realists in their moral grounding (Wright & Pölzler (2022). For instance, stronger moral convictions about legalized abortion, same-sex marriage, and mandatory vaccination are associated with greater perceived objectivity and universal generalizability (Morgan & Skitka, 2020). Certainty is the catalyst that turns attitudes into action, making people more willing to express their opinions and to persuade others to adopt their beliefs (Tomala & Rucker, 2015). Endorsement of moral objectivism provides people with a rigid, unambiguous, and definitive set of rules and expectations that are applicable across diverse social situations and contextual circumstances (Stanley et al., 2021). Absolute beliefs bring order against a random and chaotic social environment, and reduce anxiety (Kay et al., 2009).
2. **Perceived consensus:** The more agreement people perceive there to be within their community about whether a particular belief, value, or practice, the more likely they are to ground it objectively (Wright et al., 2014). Perceived consensus exerts a causal role on the perception of objectivity about a particular moral issue (Goodwin & Darley, 2012). A higher level of moral conviction about an issue and a greater perception of consensus within one's community drive the relationship with increased dogmatism and intolerance. An analysis of 19 million tweets during the 2015 Baltimore protests demonstrates that people are more likely to endorse violence when a given issue is moralized, and this effect is moderated by the degree to which they believe others share their values (Mooijman et al., 2018). The risk of violent protest is not only simply a function of moralization but also the perception that others agree with one's moral position, which can be strongly influenced by social media dynamics. Social conformity changes people's preferences and alters neural representations of value

assigned to all kinds of stimuli (things we like, values and ideas), manifested as the modulated engagement of the ventral striatum and ventromedial prefrontal cortex (Wu et al., 2016).

3. **Stability over time:** The more people perceived a moral basis for an attitude, the more consistent attitude reports were at two different times, separated by up to two months (Luttrell & Togans, 2021). Both the sense of objectivity and emotionality accompanying moralized attitudes may contribute to their stability. Moralization may lead people to develop “echo chambers” or attitudinally homogeneous social networks (Dehghani et al., 2016) and aligned social identities (Martin & Yurukoglu, 2017). In such networks, people’s views are supported by others and rarely challenged by those around them, allowing for these attitudes to persist over time. Social media amplifies these effects, as they involve a high density of interactions with others and instantly rewards and reinforces the information or opinions that are shared. This is particularly the case for moral outrage, which motivates punishment of moral transgressions and catalyzes collective action, but can also lead to the chilling of public speech and the spreading of disinformation (Brady et al., 2021).
4. **Association with emotions:** Moral convictions are associated with higher levels of autonomic nervous system physiological excitability (Garrett, 2019), as well as strong positive and negative emotional responses toward those who agree or disagree with their beliefs and attitudes (Ryan, 2014; Yoder & Decety, 2022). Using a large sample of messages on Twitter (n = 563,312), one study identified a critical role of emotion when it came to the diffusion of strong moral beliefs in online social networks (Brady et al., 2017). The presence of moral-emotional words in messages increased their transmission by approximately 20% per word. The effect of moral-emotional language was observed over and above distinctly moral and distinctly emotional language. Furthermore, the valence of moral beliefs is a robust predictor of their objectivity (Goodwin & Darley, 2012). Beliefs about negative, immoral actions are perceived to be more objective than beliefs about positive, moral actions. Moral convictions are more connected with hate when compared with disliked objects, even after adjusting for negativity and attitude strength (Pretus et al., 2023).
5. **Motivation and obligation:** Moral convictions are associated with a strong commitment to engage in activism and collective action (Sabucedo et al., 2018). They generate strong motivational states. For instance, a series of studies examined the motivation power of moral beliefs and found a dissociation between moral beliefs and action (Díaz et al., 2023). They measured participants’ moral beliefs, moral emotions, and moral choices about health recommendations during the Covid-19 pandemic and about fairness beliefs and fairness behavior in a Dictator game. Moral emotions (the emotional component of conviction rather than the cognitive component) were predictive of participants’ efforts to comply with official health recommendations.
6. **Unwillingness to compromise:** Moralized attitudes such as abortion or immigration tend to generate all of the hallmarks of political polarization—an unwillingness to compromise, lower tolerance for disagreement, and greater desired social distance from disagreeing others (Clifford, 2019). Morally convicted attitudes are insensitive to new information and lead people to reject arguments grounded in cost/benefit analysis (Ryan, 2019). Hence, moralized issues can spur citizens into action while reducing prospects for a democratic resolution to the debate when groups are polarized (Mooney, 1999).

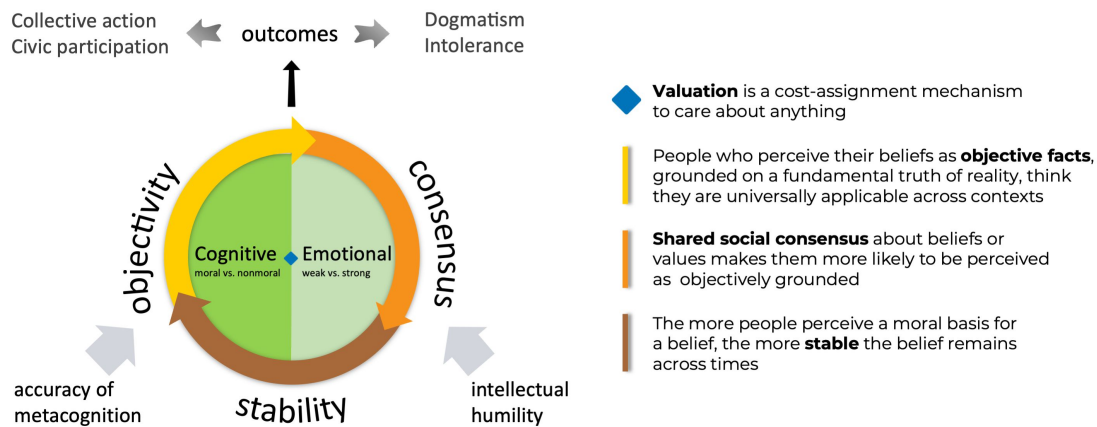


Figure 3: Functional architecture of moral conviction. This nested model characterizes the components (cognitive, emotional, valuation) of moral conviction, which are flexibly reinforced by several characteristics (objectivity, stability & social consensus). Values are the powerhouse of moral convictions. They exert a motivational force that varies in direction and intensity directing behavior toward desirable outcomes. Moral convictions motivate behavior (outcomes) and affect how people act in predictable ways. Certainty –the confidence people have in their beliefs–is deeply influenced by several social factors including consensus, and not necessarily by objective facts. Strong and dogmatic opinions may be the consequence of a cognitive style that includes lower metacognitive sensitivity, the extent to which someone can accurately distinguish their correct from incorrect responses in a variety of domains, including perception, decision-making, and memory. Conversely, intellectual humility, the sense that one’s own knowledge and perspective are limited and thus can be incorrect, reduces polarization by weakening overly dogmatic and morally convicted beliefs.

Unlike many social rules (e.g., proper greetings in different situations), which are often matters of convention and vary by location, moral rules are generally considered by an individual to be universal and binding. People are often objectivistic about their moral principles (that is, they believe that a moral rule is either totally true or false). They do not tolerate behavior that deviates from it, regardless of other people's beliefs, without room for nuance. Believing a disagreement involves moral beliefs and values is a powerful predictor of intolerance in people of all ages, which results in dogmatism, less flexible attitudes towards divergent beliefs and values, and less willingness to help, share resources with, or even sit close to those who have them (Wright et al., 2008). Moral convictions motivate the will to support and even participate in violent collective actions if these seem compatible with the subjective values of people in the coalition. They tend to elicit strong moral emotions, such as anger, contempt, outrage, guilt, and shame. As an example, morally convicted political issues on social security and gun control are immune to new information about policy effects and lead citizens to reject arguments grounded in cost/benefit reasoning (Ryan, 2019). This deontological mindset

is dangerous in leading actions (and their goals) to be evaluated in a way that is detached from their consequences.

It is of paramount importance to note that according to research, it is not morality per se that makes people intolerant and narrow-minded, but rather the view that a moral issue is objectively grounded (moral realism), which tends to be accompanied by feeling more convinced about it—about its moral rightness or wrongness—in conjunction with perceived consensus from their community (Goodwin & Darley, 2008; Wright & Pölzler, 2022). For example, refusing to provide help to people who obviously need it, or helping terminally ill patients to end their lives. This is consistent with the view that moral beliefs negatively affect interpersonal tolerance primarily because they involve issues viewed as being objective, non-negotiable, and self-evident rather than as authority-contingent (Turiel, 1983).

It is important to note that some individuals can hold a strong moral conviction against conformity pressure. Empirical evidence shows these people resist group norms, especially in studies regarding social issues. They become even more willing to speak out against a belief, value, or practice they view as deviant when they believe they hold a minority opinion. For instance, a study found that Israelis were sometimes more willing to express their opinions on Palestine when the prevailing particular public culture was antagonistic rather than sympathetic to their views (Shamir, 1997). Counter-conformity among those with moral conviction emerges on intentions to speak out. However, when people are in a position to be publicly identified with their attitude, the evidence disappears (Hornsey, Smith & Begg, 2007). This seems consistent with the general adage that actions speak louder than words. The former are more costly in terms of value-signaling and more difficult to fake. Nevertheless, several real-life incidents demonstrate that morally convicted individuals may align their actions with their words, even if they have to pay a hefty price. Kim Davis, a former clerk of Rowan County in eastern Kentucky, is an apt example. She rose to national prominence for refusing to issue same-sex marriage licenses in 2015, arguing such actions violated her moral belief that marriage is between a man and a woman, and expressing the view that the Bible is infallible. Interestingly, Davis was elected as a Democratic candidate for county clerk at a time when the majority (57% of Americans and 65% on the Democratic side) favored same-sex marriage. Davis was briefly jailed on contempt of court charges for refusing to

issue the licenses. She was later released when her staff issued the marriage certificates without her name on the form. In September 2023, Davis was ordered to pay \$100,000 to a same-sex couple she denied a marriage license. She was earning the equivalent of \$67,400 in wages at the time. Davis lost her 2018 County Clerk reelection, despite her switch to the Republican side. Since, she has become the subject of numerous mockeries and satirical works on television and social media.

Hence, there is no need to call for morality to be abolished, rather be mindful of some detrimental consequences such as intolerance for diverging values, beliefs, and practices. Sometimes, moral disagreements are fundamental, and intolerance seems justified and legitimate. But sometimes many disagreements can be traced to the distorting effects of emotions, attitudes, and interests that are inevitably bound up with moral issues.

The process of moralization

Moral values are central to human identity. They are distinct from other attitudes and opinions in several ways. Our most cherished moral beliefs are those associated with values because they have a force of attraction or repulsion. Moralized beliefs and attitudes exert a powerful motivation because they are associated with values that exert force with direction and intensity and thus guide people to pursue specific desired goals and outcomes. More formally, Lewin (1951) developed the field theory suggesting that values have the status of power fields, which represent potential energy and desire. When activated by opportunity or threat, they produce goals. Goals have the status of force fields. They represent kinetic energy or actual desire. Goal-directed motives are vectors in a force field. They reflect the desire of the person to attain a positively-valenced goal, associated with pleasure, or to avoid a negatively-valenced one, associated with pain. Behavior is a movement within the present set of force fields. Thus, the relationships between values, goals, motivation, and behavior are specified (Batson, 2015).

The process of moralization converts preferences into values, whereby objects, activities, attitudes, or behaviors that were previously morally neutral acquire a moral component and a more intense emotional charge (Rozin et al., 1997). The association between

beliefs, attitudes, and subjective values acts as a critical signal because the latter guide decisions and motivate the individual to act. Motivation concerns the processes that transform anticipated rewards into behavioral activity. When social or economic issues, whether about GMOs, a vaccination campaign, a tax reform or immigration policy are moralized, they are perceived as objective statements of fact that are universally applicable to others, regardless of their agreement, and are more resistant to persuasion. Moralization transforms what was once the realm of social preferences into values that possess intensity and direction. In doing so, this moralizing process influences inter-individual transmission, increases the likelihood of internalization, elicits a more robust emotional response, and mobilizes support for these causes. Moral beliefs can become absolute, or universal standards, which means that compromises are more difficult regardless of the cost to themselves or others (Clifford, 2019).

The process that encodes subjective values to potential actions necessary to survive (like food or running away from predators) is a fundamental biological mechanism. It is a cost-assignment biological principle, which allows one to assign values and motivates behavioral choices according to these values (Levy & Glimcher, 2011). All stimuli with these reward properties share a particular property: they reinforce behavior. Nature has endowed biological calculations (or computations) with a measure of their value that plays a pivotal role in guiding choices and decisions. While even bacteria value their short-term future and recent past, humans have the particularity of assigning values to abstract objects like ideas, which can act as reward signals. The brain uses dopamine neurons to compute a reward prediction error signal, the difference between received and predicted reward (Schultz et al., 1997). Dopamine fluctuations do not represent just one thing, nor do they play a unique role but instead signal multiple kinds of information. These signals, evolutionary designed to find food, water, and mates, are also used to pursue abstract ideas. It does this by reusing an evolutionarily old reward-chasing system but plugging some abstract ideas into the “reward” slot (Montague, 2007). This last feature allows ideas to act as rewards. Ideas gain the power of rewards and instantly become meaningful to the whole brain, especially for learning and decision-making algorithms.

Both moral and non-moral values motivate one to obtain certain goals or desirable outcomes. They are also influenced by the specific culture in which one lives, and are

encoded in similar brain circuits involved with rewards and punishments (Miller & Cushman, 2018) including the ventromedial prefrontal cortex (vmPFC), orbitofrontal cortex, striatum, anterior and posterior cingulate cortex, and lateral intraparietal cortex (Glimcher, 2011). Social reinforcers, important to learning moral norms, appear to update value representations in the brain via the same processes as non-social rewards (Jones et al., 2011). The motivational value attached to choices to approach and avoid, prosocial and antisocial outcomes, may have distinct domain-specific neural representations on a single scale, and these distinct representations for different kinds of rewards converge to a single common-scale value representation in the vmPFC, what neuro-economists have called “common currency” (Ruff & Fehr, 2014). There is some evidence that while subjective moral values are computed according to similar computational principles as financial values, they are represented in domain-specific brain areas that differ markedly from those involved in financial valuation (Ugazio et al., 2021). In that study, participants chose between saving the lives of a larger number of people at the expense of sacrificing one person's life and not harming one person and letting the group die. Moral values were represented by signals in the temporoparietal junction (TPJ), the dorsolateral prefrontal cortex (dlPFC), and the anterior insula. Thus, some of the neural processes that assign value to human lives seem dissociable from those that assign value to money.

Arbitrary goals can be plugged into the reward circuit. Abstract ideas generated by the prefrontal cortex can act directly as high-priority reward signals from the perspective of prediction error systems. These ideas then act directly as a reward signal to dopaminergic neurons, which then combine information from other brain regions to predict this new reward (Montague, 2007). Instead of simply associating goals with predictive reward stimuli, goals become rewards. In some cases, goals can become "high-status rewards," in the same way that food and sex are primary rewards necessary for survival and reproduction. These abstract rewards are not required for survival and reproduction, but they can be experienced as such, and individuals act in ways that make sense to them. For instance, selective activation of the reward circuit was detected in Danish Christians who were praying silently while being scanned with fMRI (Schjødt et al., 2008). This pattern of neural activity was only found in participants who pray regularly.

This biological mechanism accounts for why people can give their lives for all sorts of abstract ideas such as social justice, fairness, or religious beliefs. Homo sapiens are a unique species in their capacity to deny their most basic survival instincts for arbitrary ideas, like starving to death for a political or religious aim. Addiction exists here, too, including destructive behaviors motivated by moral and abstract values.¹⁰

The ventromedial prefrontal cortex and the ventral striatum are key regions of a neural system that assign and code subjective values to a wide variety of objects, both concrete (e.g., food or sex) or abstract (e.g., monetary), and that reinforces us to act. The values coded in the vmPFC influence the activity of other cortical regions involved in attentional priority, choice evaluation and the generation of behavior. Among these, one distinguishes the cognitive regions that allow reasoning and the development of an action plan and the motor regions that allow the control of the muscles necessary to carry out this action. The rewards that our brain receives both shape our affective representations (a value associated with each context in which we have received a reward) and shape our repertoire of behaviors, as we tend to reproduce actions that have been followed by a reward.

This link between abstract ideas, values, moral convictions, and the reward circuit is clearly demonstrated by functional magnetic resonance imaging (fMRI) experiments. In a study with Mormons, the sensation of “feeling the holy spirit,” which is central to their devotional practice, was reproducibly associated with activation of the ventral striatum, vmPFC, and prefrontal attentional circuitry (Ferguson et al., 2018). Interestingly, the activation of the striatum precedes the peak of spiritual feeling by 1 to 3 seconds, which is for distinct activities (reading prayers or holy scriptures). The association of abstract ideas and brain reward circuits may interact with prefrontal processing of attention and emotions, suggesting a mechanism by which religious or doctrinal concepts may become intrinsically rewarding and motivate people's behavior.

In short, evolution has endowed us with a sensation, pleasure, which allows the brain to approximate, in advance, how good or useful something is for us, generally what increases our fitness. This system can assign values to objects or situations that we imagine, including abstract ideas or objects such as ideological or religious attitudes and

beliefs. Moralization increases their subjective value, which motivates and reinforces behaviors responsible for obtaining this reward and association with pleasure.

Metacognition and dogmatism

Cognitive rigidity is frequently an outcome characteristic of moral conviction. Once an issue is highly moralized, it is likely to be perceived as an absolute belief about right or wrong. And this belief is held with high subjective confidence. High confidence in a belief leads to confirmation bias, making people unwilling to seek out corrective information and being dogmatic regardless of accuracy (Rollwage et al., 2018; Toner, Leary & Asher, 2013). Underlying cognitive processes, under the umbrella term of metacognition, can explain variation in people's capacity to reflect on, evaluate and control mental functioning in a variety of useful ways (Fleming, 2024). Metacognition allows individuals to effectively monitor, evaluate and recognize when they may be right or wrong and adjust their representations accordingly (Mazancieux et al., 2023; Proust, 2013). Specific measures of metacognition can estimate the statistical relationship between confidence judgment and objective performance, which is known as metacognitive sensitivity (Fleming, 2024).

Prior work in decision-making shows that metacognitive sensitivity depends on mechanisms that facilitate the monitoring and revision of confidence in prior choices (Van den Berg et al., 2016). Evidence from multiple studies in cognitive psychology, social psychology, clinical psychology, and political science indicates that strong and dogmatic opinions – such as those that are moralized – may be, at least in part, the consequence of a cognitive style that includes lower metacognitive sensitivity. As already indicated above, it is important to note that moralized attitudes and beliefs are more rigid than social conventions (Huebner et al., 2010). This is not specific to the moral domain. The more certain people are of a belief, regardless of its objective correctness, the greater its influence on their decisions, and the more willing they are to express their opinions and try to persuade others to adopt their own views.

Studies using a variety of measures of mental flexibility, perceptual bias, and overconfidence indicate that low metacognitive abilities contribute to a dispositional

tendency toward dogmatic intolerance, whether of ideological or political beliefs (on both sides of the political spectrum) or religious (Yoder & Decety, 2022; Zmigrod et al., 2020; Ferguson, 2021). The opposite trait, cognitive flexibility, is associated with tolerance and acceptance of viewpoints from dissimilar others (Van Prooijen & Krouwel, 2019). Additionally, research using cognitive anchoring tasks indicates that political extremists exhibit greater belief superiority and are more likely to reject outside information than moderate individuals (Brandt, Evans, & Crawford, 2015). For instance, in the case of the EU refugee crisis, political extremists from both the right and the left show greater certainty of judgment about their knowledge of the crisis, independent of their actual knowledge, than moderate individuals (van Prooijen et al., 2018). The psychological literature on right-wing and left-wing authoritarianism indicates that both entail hostility towards ‘different’ others, a willingness to wield in-group authority to coerce behavior, cognitive rigidity, aggression and punitiveness towards out-groups, a focus on status hierarchies and moral absolutism (Costello et al., 2022; Osborne et al., 2023).

Cognitive inflexibility – the inability to adapt to new or changing environmental conditions – is characterized by a difficulty in switching from one way of thinking to another and appears to play a major role in the promotion and maintenance of extreme opinions and beliefs that are associated with radicalization and terrorist actions (Decety, Pape & Workman, 2018; Victoroff, 2005). As discussed above, attitudes based on moral convictions differ from other attitudes because they are likely to be perceived as absolute, universally applicable, relatively independent of authority, and may even lead to greater acceptance of violent solutions (Skitka & Morgan, 2014). A growing body of empirical work suggests that dogmatic intolerance reflects a stable cognitive style that promotes inflexibility and distinguishes those individuals holding radical beliefs from those who are more flexible or open to debate (Rollwage et al., 2018). For example, people with high dogmatic intolerance may come from both ends of the American political spectrum, but exhibit a common ideological rigidity and a refusal to accept other viewpoints with a sense of moral superiority (Toner et al., 2013). A study using high-density electroencephalography (EEG) in college students found that their metacognitive style moderates the impact of moralization on social conformity, as reflected in early and late evoked potentials when participants perceived and then decided to support violent protests (Yoder & Decety, 2022). Additionally, participants

with extreme moral beliefs about socioeconomic and political issues during this EEG study exhibited an impairment in their metacognitive sensitivity regarding low-level perceptual discrimination judgments in a psychophysical task that allows quantitative measurement of the degree of association between accuracy and confidence.¹¹ Specifically, participants who were more radical in their moral beliefs showed less insight into the correctness of their choices (more objective errors), overconfidence in their perceptual judgments, and less updating of their confidence when they were presented with correction clues (feedback) after their decision had been made. These findings point to a generic resistance to recognizing and revising erroneous beliefs as a potential driver of moral conviction.

Weaker metacognitive ability reflects either lack of resolution or of calibration (overconfidence) and contributes to a dispositional tendency toward dogmatic intolerance, as revealed by several laboratory studies (Rollwage et al., 2018; Zmigrod et al., 2021). In support of this proposition, a series of five studies with different large samples of participants and psychophysical measures carried out in Germany indicates that the more extreme a person's beliefs (positive or negative) in both moral and nonmoral domains are extreme, the more they are convinced of them and vice versa (Heinzelmann & Tran, 2022). This relationship can contribute to moralism, dogmatism, ostentatious virtue signaling, and polarization, which in turn can be epistemically and morally problematic.

The link between dogmatism and metacognition seems mediated by reduced interest in getting additional information about a perceptual stimulus at a small cost before making a judgment about it. A study employed a perceptual decision-making task in two independent general population samples of liberals and conservatives (Schulz et al., 2020). The authors found that more dogmatic participants were less likely to seek out new information to refine an initial perceptual decision, leading to a reduction in overall belief accuracy despite similar initial decision performance. In other words, the relation between dogmatic beliefs and poor domain-general metacognition is that both phenomena are associated with a lack of interest in getting more information, even when it will help with making a better decision. The study also found that individuals on both the far left and far right of the political spectrum showed enhanced dogmatism.

However, this increase in dogmatism is more marked for those on the far right. A linear

relationship between political orientation and political belief superiority reveals that individuals on both the far left and far right show heightened beliefs in the superiority of their respective policy positions, but more so on the far left. Finally, a positive linear relationship was found between political belief superiority and dogmatism, indicating more dogmatic individuals also tended to be more confident in the superiority of their specific political conviction. Another study examined the role of metacognition when it comes to taking in new information about climate change in two large samples of individuals in the US (de Beukelaer et al., 2023). The results showed that poorer metacognition on a domain-general (perception) task was associated with less willingness to change beliefs in the face of bad news about the climate, as was being more dogmatic or more conservative.

Neuroscience studies employing EEG and functional MRI illuminate the mechanisms involved in metacognition and the relationship between error monitoring and confidence judgment. An error positivity (EP, 250-350 ms) is detected following incorrect choices in perceptual decision-making (Boldt & Yeung, 2015). This event-related component is modulated as a function of decision confidence. Its amplitude is predictive of subtle shifts in confidence. This EEG component can be reliably used as a neural index for error awareness as it covaries with subjective confidence. Activity measured via fMRI in the dorsal ACC and pre-supplementary motor area tracks perceived errors that may lead people to change their mind (Fleming et al., 2018). There is evidence that these neural signals of low confidence can generalize across distinct cognitive domains (Morales et al., 2018). Failure to update beliefs based on an error signal can lead to a failure to discard previously held beliefs, also called a bias against disconfirmatory evidence (Bronstein & Cannon, 2017).

Domain-general signals in the ventral striatum and vmPFC are also modulated by confidence level. Activity in the ventral striatum reflects the degree of confidence (Hebart et al., 2016). Neurons in the vmPFC are involved in calculating the expected value of choice outcomes during many neuroeconomic tasks (Kable et al., 2007; Pessiglione & Delgado, 2015; Shapiro & Grafton, 2020). These studies suggest that the vmPFC encodes additional characteristics of the decision process, such as the degree of confidence of the subjects in their choice. In one fMRI study, participants were asked to make judgments, such as estimating a person's age from a photo, how enjoyable they

found a food, or assessing the likelihood of France winning the next soccer World Cup. Participants were also asked to indicate the extent to which they were confident about the accuracy of their judgments (Lebreton et al., 2015). The results showed that confidence and reaction times varied quadratically with assessment across a wide range of different tasks. The same quadratic relationship between a participant's rating and their confidence in their answer was also detected in a signal increase in the vmPFC as a function of the rating index. The authors proposed that trust cues may have a distinct role in metacognitive monitoring in other prefrontal regions that modulate their coupling strength with the vmPFC as a function of trust, but not necessarily as a function of value.

Beyond having more difficulty in accurately judging one's own performance, recent work has also identified a mechanism linking dogmatic beliefs with a poorer ability to use metacognition to guide decisions. Individuals who score high in dogmatism are more reluctant to pay a small quantifiable cost to access information that will lead to a more accurate decision on a perceptual task, particularly when confidence is low (Schulz et al., 2020). These studies suggest that people high in dogmatism exhibit a domain-general reluctance to process information that may challenge existing beliefs that they know may be inaccurate. The results echo prior work showing that individuals higher in dogmatism are often less inclined to update beliefs (e.g., about climate change; De Beukelaer et al., 2023).

People tend to take in new information when they believe the information to be valuable, and avoid new information that they believe has negative value. According to one framework (Sharot & Sunstein, 2020), decisions about whether to seek new information emerge from a combined valuation based on instrumental utility (increases or decreases in extrinsic rewards due to new knowledge), hedonic utility (positive or negative affect generated by new knowledge), and cognitive utility (the degree to which new information leads to a feeling of understanding about the world). Note that cognitive utility is positive when new information either fits into an existing mental model or when a challenge to existing beliefs leads to the creation of an updated mental model. Cognitive utility is zero when new information is orthogonal to existing mental models, and can be negative when new information challenges one's mental model but there is a disinclination to update the model. Subsequent work has shown that the

degree to which people rely on each type of utility (instrumental, hedonic, and cognitive) when making choices about information uptake differs between individuals and can be quantified as a relatively stable trait, even across different domains of information (Kelly & Sharot, 2021).

Some research indicates an inverse relationship between analytical reasoning and dogmatism (Friedman & Jack, 2018). Studies in cognitive science suggest that higher levels of analytical intelligence and manipulations that involve analytical thinking encourage the revision of beliefs by enabling an individual to evaluate contradictory claims and consider alternatives (Pennycook et al., 2012; 2016). Additionally, individuals with greater fluid intelligence, which is strongly correlated with working memory capacity (Stuppel et al., 2013; Toplak et al., 2011), are better able to keep in mind alternative possibilities and therefore are more able to deal with uncertainty (Coutinho et al. 2015). These considerations suggest that analytical reasoning ability is likely to be associated with the flexibility of beliefs and the ability to revise them. However, it is unlikely that analytical reasoning alone is the only psychological factor related to the flexibility of beliefs. Indeed, many beliefs are inherently social, and the literature reviewed above clearly involves social-emotional processes.

It seems established that lower cognitive flexibility and metacognition levels are linked to strong moral convictions, dogmatic attitudes, and excessive self-confidence. Across all ideologies and political spectra, people who hold extreme social, political, or religious attitudes, including ideologically motivated violence against others, have a consistent psychological profile (Zmigrod et al., 2021). This research was based on a large sample of participants who completed an extensive set of well-established cognitive tasks and self-report surveys focusing on self-regulation and personality characteristics. Nevertheless, it remains necessary to determine the direction of the causality. Are people with a predisposition for low metacognitive abilities more likely to adopt strong moral convictions and have dogmatic attitudes, regardless of the domain (aesthetics, attitudes, politics, economics, social, religious)? Or, do people who acquire moral convictions become more rigid and less capable of conscious regulation of their mental representations? What is the mediating influence of coalitional psychology on cognitive rigidity? Studies examining the impact of metacognitive sensitivity training on extreme beliefs would provide information about the direction of causality, and if the training

reduces extremism, it would indicate a route through which moralization could be reduced.

Finally, it is worth mentioning the relationship between metacognitive abilities, humility, and self-esteem. Favorable impressions of oneself may not be an unmitigated good from the perspective of society if they lead to violence (Baumeister et al., 1996). Politically motivated violence, including terrorism, is often correlated with and preceded by strongly favorable self-regard, a sense of moral superiority, confidence and arrogance, and the perception that these views are threatened by others. Conversely, a growing body of research has demonstrated the relevance of intellectual humility, that is, recognizing the limits and fallibility of one's knowledge, in helping buffer people from some of their more authoritarian, dogmatic, and other biased proclivities (Porter et al., 2022). In the political domain, intellectual humility is associated with less affective polarization concerning political and religious groups. Still, individuals are unlikely to act in an intellectually humble manner when motivated by strong convictions or when their political, religious, or ethical values seem to be challenged (Krumrei-Mancuso et al., 2020).

Can moralization be countered?

Moralization is a social-psychological process through which morally neutral issues take on moral significance. Moralization can be a positive force and lead to good outcomes like social progress. Yet it can be detrimental to individuals and to society as a whole by exacerbating social conflict and political polarization. Understanding the cognitive, emotional, and social processes and mechanisms underlying moral convictions is essential to guide how to best respond to dogmatic views when necessary.

Certainty, or the confidence people have in their beliefs, contributes to their durability and impact on behavior, particularly when they think others share their views (Cheatham & Tormala, 2015). There is some evidence about the degree to which highly moralized attitudes are more resistant to counter-attitudinal persuasive messages than attitudes that are low in moral conviction (Luttrell et al., 2016). Whether individuals with morally convicted attitudes are persuadable appears to vary as a function of the specific

persuasive message. Morally convicted attitudes appear to be resistant to non-moralized counter-attitudinal messages (e.g., those that frame arguments in terms of pragmatic concerns) or messages that emphasize consequences (e.g., harms and benefits). In keeping with theories of persuasion and behavior change suggesting that people are persuaded more by subjective value and less by facts (Falk & Scholtz, 2018), some research has shown that moralized attitudes have greater malleability in response to moralized counter-attitudinal messages, that is, arguments that are framed using deontological or rules-based messaging (Luttrell et al., 2019, Ryan, 2019), or messages that emphasize counter-attitudinal anger and disgust (Clifford, 2019). There are also suggestions that people's moral convictions are attenuated when they are at odds with their financial interests (Bastian et al., 2015), except for sacred values, which are immune to material and financial tradeoffs (Tetlock et al., 2017).

When moralization seems unjustified and inappropriate, Kraaijeveld and Jamrozik (2022), using concrete examples relevant to public health, suggest engaging in metaethical analysis of moralization processes in conjunction with paying close attention to the relevant facts/data. This metaethical perspective attempts to analyze and understand the metaphysical, epistemological, semantic, and psychological presuppositions and commitments of moral thought and practice. The authors then propose a corrective process, which involves actively working against the forces that lead to memorialization, like avoiding or opposing emotional frames eliciting disgust and anger.

Framing strategies can be successfully designed to counter moralized attitudes or beliefs. Moral frames are persuasive and moralize people's attitudes, whereas nonmoral frames are persuasive and demoralize people's attitudes. A series of studies conducted by Kodapanakkal and her colleagues (2022) employed persuasion messages (moral, nonmoral, and control) that opposed big-data technologies on crime surveillance technologies and hiring algorithms. Participants were asked to report their support for the technology and the degree to which their attitude was based on a moral conviction. Across the studies, as expected, moral frames and nonmoral frames were equally persuasive. However moral frames increased the strength of people's moral convictions and made them less willing to compromise, whereas nonmoral frames decreased the strength of people's moral convictions. The authors also found that anger and disgust

potentially drove moralization and that considering how financially costly a technology is drives demoralization.

On a positive note, the process of moralization is reversible. Something in the moral domain can gradually cease to be so, and become a mere preference and vice versa (Rozin, 1999). In the United States, for instance, attitudes to inter-racial marriage, marijuana use, and homosexuality have become a matter of personal preference rather than morals, at least for most citizens.

Conclusion

Harmonious and productive group living requires adaptive systems for encouraging, maintaining, and regulating social order. Morality has evolved to facilitate cooperation between individuals whose survival depends on one another, by coordinating their efforts to foster mutual interests. Morality also serves a critical role in bringing into focus issues of harm and fairness—and the potential for people to be harmed and mistreated—so that we can make course corrections to protect and promote our individual and collective well-being (Wright & Pölzler, 2021).

Counterintuitively, moralization is not always associated with positive social outcomes. Strong moral beliefs can promote overconfidence, dogmatism, and intolerance for divergent views, especially when people are moral objectivists combined with a high degree of conformity with their coalition.

When an issue is moralized, it is likely to receive more attention from individuals and institutions but it also comes with the risk of inflaming people's righteous indignation, making rational debate and compromise more difficult and likely to be polarized. Oftentimes, moral convictions are deontological to the extent that they treat specific moral norms or attitudes as absolutes, motivating people to pursue these beliefs at the expense of all else, including personal gain (Bastian et al., 2015). An apt illustration is the opposition to legalizing euthanasia, which is perceived by many people as murder, even if the person has a terminal illness and is begging for help to end their life. Despite this opposition, a majority of individuals (73%) in the US do not moralize this

action. Similarly, 69% of Americans say abortion should be legal, and 85% say in at least certain circumstances. 80% of French people consider euthanasia and physician-assisted suicide to be end-of-life care.¹² In contrast, euthanasia has been decriminalized in 10 countries and is legal in 11 jurisdictions in the US, but is still prohibited in France.

We naturally tend to unite in groups and reciprocally influence each other to act in conformity with certain moral beliefs. This is consistent with the view that our moral psychology evolved as an extension of the adaptations underlying coalitional psychology and adaptations for social assortment and exclusion (Cosmides & Tooby, 2010). Biased representations of the social world emerge from this inevitable trait of human nature, beneficial to survival within one's coalition. Such information processing biases are amplified by moralization, by selectively guiding attention to information that reinforces existing beliefs (Tetlock, 2002). In other words, tribal biases lead to motivated reasoning, selectively seeking out information that supports the interests of one's coalition and avoiding or filtering out why could harm it. This is clearly the case with engagement in social information networks, which form "echo chambers" by promoting ideological homophily (Boutyline & Willer, 2016). These types of cognitive biases are problematic for two reasons. All post-Enlightenment liberal societies place a high premium on rationality and individual freedom and do not tolerate obvious manifestations of group favoritism (Clark et al., 2019). Moreover, democratic governments have to coordinate coalitions with differing and conflicting interests to properly function for the common good. Intolerance for divergent beliefs, values, and practices can quickly breed in homogenous groups where members share the same views (Wright & Pölzler, 2021). Many forms of political violence, including extreme ones like war, are morally motivated and culturally legitimized. Individuals' support of war depends on their perceptions of the righteousness of violence rather than the strategic efficacy of warfare (Ginges & Atran, 2011). Values can operate as moral imperatives that delineate which actions and policies, including killing, are right or wrong (Halevy et al., 2015). Similarly, values can impact which opinions are acceptable, which books can be published or used in schools, etc. Major advances in the field of social neuroscience show that moral convictions acquire the power of rewards and become meaningful to the rest of the brain. They motivate and reinforce collective actions and

are supercharged by group dynamics and social validation. Since coalitional programs

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). <https://doi.org/10.5281/zenodo.11001101>
2024/1 - paris-ias-ideas - Article No.9. Freely available at <https://paris.pias.science/article/the-power-of-moral-conviction-how-it-catalyzes-dogmatism-intolerance-and-violence> - ISSN 2826-2832/© 2024 Decety J.

This is an open access article published under the [Creative Commons Attribution-NonCommercial 4.0 International Public License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

evolved to promote the self-interest of the coalition's membership (in dominance, status, legitimacy, resources, moral force, etc.), even coalitions whose organizing ideology originates (ostensibly) to promote human welfare are at risk of sliding into extreme forms of oppression, as such existing in complete contradiction to the putative values of the group (Tooby, 2017).

A better scientific understanding of attitudes and moralization dynamics is valuable to social scientists and policymakers to identify and support pathways away from extremism and strengthen and sustain engagement in benevolent forms of collective action. When we see acts of hostile actions, we should start by asking ourselves if they are motivated by moral considerations, what social relations they aim to regulate or group status improvement they aim to achieve, and what social coalition is promoted and benefits from such actions. It may be wise, when possible, to avoid moralizing an issue, whether it is climate change, refugees, prison rape, vaccination, gender affirmative care, police brutality, the protection of dolphins, or eating meat, because once it is moralized, it becomes more difficult to engage in rational cost-benefit analysis. Moralized attitudes tend to generate all of the hallmarks of political polarization—an unwillingness to compromise, lower tolerance for disagreement, strong emotions, and hostile forms of collective action such as protests and riots (Zaal et al., 2011). Often, government agencies regulate products that some people believe are morally wrong. While cost-benefit calculations can be challenging when strong moral convictions are involved, it should still be possible to use valuations that reflect how much people are willing to pay to see their moral beliefs vindicated or to reduce the level of psychological harm they feel (Posner & Sunstein, 2017). More specifically, Posner and Sunstein argue that, if agencies monetize the welfare losses that come from violation of moral commitments, they will be in a better position to decide on the severity of regulations in a nonarbitrary way.

Acknowledgments: I am grateful to Qiongwen Cao, Michael S. Cohen, Annalise Halverson, Joanna Li, and Abbigail Schwartz (at the University of Chicago), Joëlle Proust (CNRS-Institute Jean Nicod, Paris) and Cecilia Heyes (University of Oxford) for helpful comments on an earlier version of the paper. This article was supported by a fellowship from the Paris Institute for Advanced Study.

Box 2. Glossary of terms and key concepts

- **Attitude** refers to a set of emotions, beliefs, and behaviors toward a particular person, object, or event. Attitudes are composed of three dimensions: strength, accessibility, and ambivalence. Attitudes can be explicit and implicit.
- **Coalition** or alliance accounts for individuals behaving in ways to enhance each other's welfare.
- **Dopamine** is a neurotransmitter that is both excitatory and inhibitory in function, as well as a neuromodulator involved in reward and motivation. Transient mesolimbic dopamine release events energize reward-seeking and encode reward value.
- **Moral conviction** is a strong and absolute belief that something is right or wrong, moral or immoral.
- **Moral realism** is the view that morality is grounded in objective facts, independent from the beliefs, attitudes, and practices of individuals and cultures.
- **Morality**, through the internalization of social norms, coordinates interpersonal exchanges, facilitates coexistence and cooperation, and allows to find a balance when individual interests conflict with collective interests. People are motivated by morality because it is advantageous at the individual level – a non-zero-sum game.
- **Moralization** converts preferences into values and as a result, increases the likelihood of internalization, invokes greater emotional response, draws attention, and mobilizes coalition members.
- **Normative influence** prompts individuals to conform to group norms to fit in, gain acceptance, and feel good.

- **Norms** can be categorized as descriptive or injunctive. Descriptive norms are regularities of behaviors corresponding to what most people do in a given situation. Injunctive norms are behavioral expectations that are enforced by sanctions.
- **Reward system** is a group of interconnected regions that determine the valence of a stimulus, and signal whether it should be approached or avoided. Structures that are part of the reward system are found along the primary dopamine pathways, including the ventral tegmental area that projects to the nucleus accumbens, which is part of the ventral striatum. The reward circuit, considered to be embedded within the cortico-basal ganglia network, is a central component for developing and monitoring motivated behaviors.
- **Neural valuation system** includes the ventral striatum and ventromedial prefrontal cortex (vmPFC) in which neuronal activity is linked to preference judgments and valuation in decision-making. The vmPFC and dorsal anterior cingulate cortex code reward signals across a wide range of value-based choice contexts, including the subjective value of information as well as anticipatory signals indicating that a reward is expected later on. Social reputation and approval are powerful social rewards.
- **Social exchange heuristic** corresponds to a set of assumptions about social interaction viewed as a form of reciprocal cooperation. People attempt to reduce costs and maximize rewards. Then, they base the likelihood of developing a relationship with someone on the anticipated outcomes. Individuals intuitively cooperate or defect depending on which behavior is beneficial in their interactions.
- **Social influence** refers to the phenomenon where an individual's behaviors, opinions, or beliefs change as a result of their network ties, often becoming more similar to those with whom they are connected.

- **Valuation** is a cost-assignment mechanism. Valuation puts meaning into neural computations and is used to guide behavior. It is a signal for energetic investment. Humans have a unique capacity to value abstract and arbitrary objects.
- **Value** is a motivational force (attraction toward something or repulsion away) representing potential energy or desires. It has direction and intensity. Value is an integration of factors that affect how desirable a reward is, including its magnitude, the probability of receiving it, and the effort needed to obtain it. When activated by threats or opportunities, values produce goals that lead to action. The neural representation of value is implemented in ventromedial prefrontal cortex and ventral striatum.
- **Ventral striatum** is the primary input nucleus of the basal ganglia. The ventral portion of the striatum is activated by rewards and receives input from the dopamine system. The magnitude of activity in the ventral striatum reflects the value of reward-predicting stimuli.
- **Ventromedial prefrontal cortex** is a subregion of the prefrontal cortex, interconnected with subcortical and cortical regions, activated by stimuli that anticipate a reward. Neurons in this region code the subjective value (the specific needs and preferences of the individual). Values coded in the vmPFC influence the activity of cortical regions involved in generating behavior.
- **Virtue signaling** corresponds to conspicuously displaying opinions intended to demonstrate one's good character or the moral correctness of one's position on a particular issue. It is a powerful evolved force for social regulation by creating common knowledge around a moral issue that people would otherwise ignore. This works even when there is no guarantee the individuals who are sending the signals are particularly virtuous or committed to the cause.

Bibliography

Aknin, L. B., & Whillans, A. V. (2021). Helping and happiness: A review and guide for public policy. *Social Issues and Policy Review*, 15(1), 3–34.

Alexander, R. D. (1974). The evolution of social behavior. *Annual Review of Ecology and Systematics*, 5(1), 325–383.

Asif, M., & Weenink, D. (2022). Vigilante rituals theory: A cultural explanation of vigilante violence. *European Journal of Criminology*, 19(2), 163–182.

Atran, S. (2010). *Talking to the enemy: faith, brotherhood, and the (un) making of terrorists*. Harper Collins.

Barkow, J. H. (1991). Précis of Darwin, sex and status: Biological approaches to mind and culture. *Behavioral and Brain Sciences*, 14, 295–334.

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

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

Batson, C. D. (2015). *What's wrong with morality*. Oxford University Press.

Bauman, C. W., & Skitka, L. J. (2009). Moral disagreement and procedural justice: Moral mandates as constraints to voice effects. *Australian Journal of Psychology*, 61(1), 40–49.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529.

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: the dark side of high self-esteem. *Psychological Review*, 103(1), 5–33.

Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. *PloS One*, 7(11), 48693.

Bavel, J. J., FeldmanHall, O., & Mende-Siedlecki, P. (2015). The neuroscience of moral cognition: From dual processes to dynamic systems. *Current Opinion in Psychology*, 6, 167–172.

Bentahila, L., Fontaine, R., & Pennequin, V. (2021). Universality and cultural diversity in moral reasoning and judgment. *Frontiers in Psychology*, 12, 764360.

Berntson, G. G., Norman, G. J., Hawkley, L. C., & Cacioppo, J. T. (2012). Evolution of neuroarchitecture, multi-level analyses and calibrative reductionism. *Interface Focus*, 2(1), 65–73.

Beukelaer, S., Vehar, N., Rollwage, M., Fleming, S. M., & Tsakiris, M. (2023). Changing minds about climate change: a pervasive role for domain-general metacognition. *Humanities and Social Sciences Communications*, 10(1), 46.

Bhanji, J. P., & Delgado, M. R. (2014). The social brain and reward: social information processing in the human striatum. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5(1), 61–73.

Black, D. (1983). Crime as social control. *American Sociological Review*, 48, 34–45.

Bloom, P. B. N., Arkan, G., & Courtemanche, M. (2015). Religious social identity, religious belief, and anti-immigration sentiment. *American Political Science Review*, 109(2), 203–221.

Böhm, R., Rusch, H., & Baron, J. (2020). The psychology of intergroup conflict: A review of theories and measures. *Journal of Economic Behavior & Organization*, 178, 947–962.

Boldt, A., & Yeung, N. (2015). Shared neural markers of decision confidence and error detection. *Journal of Neuroscience*, 35(8), 3478–3484.

Bostyn, D. H., & Roets, A. (2017). An asymmetric moral conformity effect: Subjects conform to deontological but not consequentialist majorities. *Social Psychological and Personality Science*, 8(3), 323–330.

Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, 38(3), 551–569.

Bowles, S. (2016). *The moral economy*. Yale University Press.

Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *Proceedings of the National Academy of Science USA*, 100, 3531–3535.

Boyd, R. (2018). *A different kind of animal - How culture transformed our species*. Princeton University Press.

Boyer, P., Firat, R., & Leeuwen, F. (2015). Safety, threat, and stress in intergroup relations: A coalitional index model. *Perspectives on Psychological Science*, 10(4), 434–450.

Boyer, P. (2018). *Minds make societies*. Cambridge University Press.

Boyer, P. (2020). Morality, valuation and coalitional psychology. *American Journal of Bioethics - Neuroscience*, 11(4), 287–289.

Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318.

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33), 5641.

Brandt, M. J., Evans, A. M., & Crawford, J. T. (2015). The unthinking or confident extremist? Political extremists are more likely than moderates to reject experimenter-generated anchors. *Psychological Science*, 26, 189–202.

Branscombe, N. R., Wann, D. L., Noel, J. G., & Coleman, J. (1993). In-group or out-group extremity: Importance of the threatened social identity. *Personality and Social Psychology Bulletin*, 19(4), 381–388.

Bronstein, M. V., & Cannon, T. D. (2017). Bias against disconfirmatory evidence in a large nonclinical sample: Associations with schizotypy and delusional beliefs. *Journal of Experimental Psychopathology*, 8(3), 288–302.

Burns, M. P., & Sommerville, J. A. (2014). I pick you’’: The impact of fairness and race on infants’ selection of social partners. *Frontiers in Psychology*, 5, 93.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin*, 115, 401–423.

Cacioppo, J. T., Berntson, G. G., & Decety, J. (2010). Social neuroscience and its relationship to social psychology. *Social Cognition*, 28(6), 675–685.

Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology*, 20(13), 1165–1170.

Carpenter, J. (2008). *Social preferences*. *The new Palgrave dictionary of economics* (pp. 1–5). Palgrave Macmillan.

Caviola, L., & Faulmüller, N. (2014). Moral hypocrisy in economic games—how prosocial behavior is shaped by social expectations. *Frontiers in Psychology*, 5, 897.

Cheatham, L., & Tormala, Z. L. (2015). Attitude certainty and attitudinal advocacy: The unique roles of clarity and correctness. *Personality and Social Psychology Bulletin*, 41(11), 1537–1550.

Cichocka, A., & Dhont, K. (2018). The personality bases of political ideology and behavior. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The SAGE Handbook of personality and individual differences* (Vol. 3, pp. 323–351). Sage Publications, Ltd.

Clark, C. J., Liu, B. S., Winegard, B. M., & Ditto, P. H. (2019). Tribalism is human nature. *Current Directions in Psychological Science*, 28(6), 587–592.

Clark, C., Isch, C., Everett, J., & Shariff, A. (2023). *Politicization undermines trust in institutions, even among the ideologically aligned public*. <https://orcid.org/0000-0002-3083-9179>

Clifford, S. (2019). How emotional frames moralize and polarize political attitudes. *Political Psychology*, 40(1), 75–91.

Contreras-Huerta, L. S., Baker, K. S., Reynolds, K. J., Batalha, L., & Cunnington, R. (2013). Racial bias in neural empathic responses to pain. *PloS One*, 8(12), 84001.

Cosmides, L., Tooby, J., & Barkow, J. (1992). Introduction: Evolutionary psychology and conceptual integration. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 3–15). Oxford University Press.

Cosmides, L., & Tooby, J. (2010). Groups in mind: The coalitional roots of war and morality. In H. Høgh-Olesen, C. Boesch, & L. Cosmides (Eds.), *Human Morality and Sociality* (pp. 191–234). Palgrave Macmillan.

Costello, T. H., Bowes, S. M., Stevens, S. T., Waldman, I. D., Tasimi, A., & Lilienfeld, S. O. (2022). Clarifying the structure and nature of left-wing authoritarianism. *Journal of Personality and Social Psychology*, 122(1), 1–36.

Coutinho, M. V., Redford, J. S., Church, B. A., Zakrzewski, A. C., Couchman, J. J., & Smith, J. D. (2015). The interplay between uncertainty monitoring and working memory: Can metacognition become automatic? *Memory and Cognition*, 43(7), 990–1006.

Cowell, J., & Decety, J. (2015). Precursors to morality in development as a complex interplay between neural, socioenvironmental, and behavioral facets. *Proceedings of the National Academy of Sciences*, 112(41), 12657–12662.

Critcher, C. R., Inbar, Y., & Pizarro, D. A. (2013). How quick decisions illuminate moral character. *Social Psychological and Personality Science*, 4(3), 308–315.

Crockett, M. J., Apergis-Schoute, A., Herrmann, B., Lieberman, M. D., Müller, U., Robbins, T. W., & Clark, L. (2013). Serotonin modulates striatal responses to fairness and retaliation in humans. *Journal of Neuroscience*, 33(8), 3505–3513.

Curry, O. S., Mullins, D. A., & Whitehouse, H. (2019). Is it good to cooperate? Testing the theory of morality-as-cooperation in 60 societies. *Current Anthropology*, 60(1), 47–69.

Díaz, R. (2023). *Do moral beliefs motivate action? Ethical Theory and Moral Practice* (pp. 1–19).

Dancy, J. (1993). *Moral reasons*. Blackwell.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *NeuroImage*, 23(2), 744–751.

Decety, J., Echols, S. C., & Correll, J. (2010). The blame game: The effect of responsibility and social stigma on empathy for pain. *Journal of Cognitive Neuroscience*, 22(5), 985–997.

Decety, J., & Cacioppo, J. T. (2011). *The Oxford handbook of social neuroscience*. Oxford University Press.

Decety, J., & Porges, E. C. (2011). Imagining being the agent of actions that carry different moral consequences: an fMRI study. *Neuropsychologia*, 49(11), 2994–3001.

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.

Decety, J., Pape, R., & Workman, C. I. (2018). A multilevel social neuroscience perspective on radicalization and terrorism. *Social Neuroscience*, 13(5), 511–529.

Decety, J. (2019). Les origines naturelles du sens moral chez le petit enfant. *Devenir*, 31(3), 207–230.

Decety, J. (2021). Why empathy is not a reliable source of information in moral decision-making. *Current Directions in Psychological Science*, 30(5), 425–430.

Decety, J., Steinbeis, N., & Cowell, J. M. (2021). The neurodevelopment of social preferences in early childhood. *Current Opinion in Neurobiology*, 68, 23–28.

Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N. J., & Graham, J. (2016). Purity homophily in social networks. *Journal of Experimental Psychology: General*, 145(3), 366–375.

Delton, A. W., & Krasnow, M. M. (2015). Adaptationist approaches to moral psychology. In J. Decety & T. Wheatley (Eds.), *The moral brain* (pp. 19–34). MIT press.

Den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. (2016). A common mechanism underlies changes of mind about decisions and confidence. *Elife*, 5, 1–21.

DeScioli, P., Massenkoff, M., Shaw, A., Petersen, M. B., & Kurzban, R. (2014). Equity or equality? Moral judgments follow the money. *Proceedings of the Royal Society: Biological Sciences*, 281(1797),

Decety, J. (2024). The power of moral conviction: How it catalyzes dogmatism, intolerance, and violence. In *Proceedings of the Paris Institute for Advanced Study* (Vol. 1). <https://doi.org/10.5281/zenodo.11001101>

2024/1 - paris-ias-ideas - Article No.9. Freely available at <https://paris.pias.science/article/the-power-of-moral-conviction-how-it-catalyzes-dogmatism-intolerance-and-violence> - ISSN 2826-2832/© 2024 Decety J.

This is an open access article published under the [Creative Commons Attribution-NonCommercial 4.0 International Public License \(CC BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

20142112.

DeScioli, P., & Kurzban, R. (2018). Morality is for choosing sides. In K. Gray & J. Graham (Eds.), *Atlas of moral psychology* (pp. 177–185). Guilford Press.

Diaz, S., & Bailey, J. M. (2023). Rapid onset gender dysphoria: Parent reports on 1655 possible cases. *Journal of Open Inquiry in the Behavioral Sciences*. <https://doi.org/10.58408/issn.2992-9253.2023.01.01.00000012>

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., & Zinger, J. F. (2019). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*, 14(2), 273–291.

Duhaime, E. P. (2015). Is the call to prayer a call to cooperate? A field experiment on the impact of religious salience on prosocial behavior. *Judgment and Decision Making*, 10(6), 593–596.

Eagan, S. P. (1996). From spikes to bombs: The rise of eco-terrorism. *Studies in Conflict & Terrorism*, 19(1), 1–18.

Edelman, B. (2009). Markets: Red light states: Who buys online adult entertainment? *Journal of Economic Perspectives*, 23(1), 209–220.

Engelmann, J. M., & Rapp, D. J. (2018). The influence of reputational concerns on children's prosociality. *Current Opinion in Psychology*, 20, 92–95.

Enke. (2019). Kinship, cooperation and the evolution of moral systems. *The Quarterly Journal of Economics*.

Eriksson, K., Vartanova, I., Ornstein, P., & Strimling, P. (2021). The common-is-moral association is stronger among less religious people. *Humanities and Social Sciences Communications*, 8(1), 1–8.

Falk, E., & Scholz, C. (2018). Persuasion, influence, and value: Perspectives from communication and social neuroscience. *Annual Review of Psychology*, 69, 329–356.

Fazio, R. H. (1995). Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility. In R. E. Petty & J. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 247–282). Erlbaum.

Ferguson, M. A., Nielsen, J. A., King, J. B., Dai, L., Giangrasso, D. M., Holman, R., & Anderson, J. S. (2018). Reward, salience, and attentional networks are activated by religious experience in devout Mormons. *Social Neuroscience*, 13(1), 104–116.

Fernbach, P. M., Light, N., Scott, S. E., Inbar, Y., & Rozin, P. (2019). Extreme opponents of genetically modified foods know the least but think they know the most. *Nature Human Behaviour*, 3(3), 251–256.

Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, 370, 533–536.

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

Fleming, S. M., Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, 21(4), 617–624.

Fleming, S. M. (2024). Metacognition and confidence: A review and synthesis. *Annual Review of Psychology*, 75.

Forgas, J. P., Haselton, M. G., & Hippel, W. (2007). *Evolution and the social mind: Evolutionary psychology and social cognition*. Routledge/Taylor & Francis Group.

Fourie, M. M., Subramoney, S., & Gobodo-Madikizela, P. (2017). A less attractive feature of empathy: Intergroup empathy bias. In M. Kondo (Ed.), *Empathy: An Evidence-based Interdisciplinary Perspective* (pp. 45–61). InTech.

Fousiani, K., Yzerbyt, V., Kteily, N. S., & Demoulin, S. (2019). Justice reactions to deviant ingroup members: Ingroup identity threat motivates utilitarian punishments. *British Journal of Social Psychology*, 58(4), 869–893.

Friedman, J. P., & Jack, A. I. (2018). What makes you so sure? Dogmatism, fundamentalism, analytic thinking, perspective taking and moral concern in the religious and nonreligious. *Journal of Religion and Health*, 57, 157–190.

Gómez, J. M., Verdú, M., González-Megías, A., & Méndez, M. (2016). The phylogenetic roots of human lethal violence. *Nature*, 538(7624), 233–237.

Garrett, K. N. (2019). Fired up by morality: The unique physiological response tied to moral conviction in politics. *Political Psychology*, 40(3), 543–563.

Garrett, K. N., & Bankert, A. (2020). The moral roots of partisan division: How moral conviction heightens affective polarization. *British Journal of Political Science*, 50(2), 621–640.

Gelfand, M. J., Harrington, J. R., & Jackson, J. C. (2017). The Strength of Social Norms Across Human Groups. *Perspectives on Psychological Science*, 12(5), 800–809.

Georgiev, A. V., Klimczuk, A. C., Traficante, D. M., & Maestripieri, D. (2014). When violence pays: a cost-benefit analysis of aggressive behavior in animals and humans. *Evolutionary Psychology*, 11(3), 678–699.

Ginges, J., Hansen, I., & Norenzayan, A. (2010). Religious belief, coalitional commitment, and support for suicide attacks. *Evolutionary Psychology*, 8(3), 346–349.

- Ginges, J., & Atran, S. (2011). War as a moral imperative (not just practical politics by other means). *Proceedings of the Royal Society B*, 278(1720), 2930–2938.
- Gintis, H. (2011). Gene–culture coevolution and the nature of human sociality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1566), 878–888.
- Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition*, 106(3), 1339–1366.
- Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology*, 48(1), 250–256.
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, 23, 101–124.
- Grossman, M. (2023). *Lost in trans nation*. Skyhorse.
- Habib, C. (2019). *La question* (L. Débat, Trans.). Gallimard.
- Haidt, J. (2011). *The righteous mind: why good people are divided by politics and religion*. Pantheon.
- Halevy, N., Kreps, T., Weisel, O., & Goldenberg, A. (2015). Morality in Intergroup Conflict. *Current Opinion in Psychology*, 6, 10–14.
- Hamid, N., Pretus, C., Atran, S., Crockett, M. J., Ginges, J., Sheikh, H., & Vilarroya, O. (2019). Neuroimaging ‘will to fight’ for sacred values: an empirical case study with supporters of an Al Qaeda associate. *Royal Society Open Science*, 6(6), 181585.
- Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers reach to antisocial others. *Proceedings of the National Academy of Sciences*, 108, 19931–19936.
- Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science*, 22(3), 186–193.
- Hamlin, J. K. (2015). The infantile origins of our moral brain. In J. Decety & T. Wheatley (Eds.), *The moral brain* (pp. 105–122). MIT Press.
- Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J. D. (2016). The relationship between perceptual decision variables and confidence in the human brain. *Cerebral Cortex*, 26(1), 118–130.

Heinzelmann, N., & Tran, V. (2022). Extremists are more confident. In *Erkenntnis* (pp. 1–26).

Henderson-King, D., Henderson-King, E., Bolea, B., Koches, K., & Kauffman, A. (2004). Seeking understanding or sending bombs: Beliefs as predictors of responses to terrorism. *Peace and Conflict: Journal of Peace Psychology*, 10(1), 67–84.

Henrich. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Heyes, C. (2024). Rethinking norm psychology. *Perspectives on Psychological Science*, 1, 12–38.

Hippel, W. (2018). *The social leap*. Haper Collins.

Hoffman, M., Yoeli, E., & Nowak, M. A. (2015). Cooperate without looking: Why we care what people think and not just what they do. *Proceedings of the National Academy of Sciences of the USA*, 112, 1727–1732.

Holbrook, C., Fessler, D. M., Sparks, A. M., Johnson, D. L., Samore, T., & Reed, L. I. (2023). Coalitionality shapes moral elevation: evidence from the US Black Lives Matter protest and counter-protest movements. *Royal Society Open Science*, 10(3), 220990.

Hornsey, M. J., Smith, J. R., & Begg, D. (2007). Effects of norms among those with moral conviction: Counter-conformity emerges on intentions but not behaviors. *Social Influence*, 2(4), 244–268.

Huang, S., & Han, S. (2014). Shared beliefs enhance shared feelings: religious/irreligious identifications modulate empathic neural responses. *Social Neuroscience*, 9(6), 639–649.

Huebner, B., Lee, J., & Hauser, M. (2010). The moral-conventional distinction in mature moral competence. *Journal of Cognition and Culture*, 10, 1–6.

Huppert, E., Cowell, J. M., Cheng, Y., Contreras-Ibáñez, C., Gomez-Sicard, N., Gonzalez-Gadea, M. L., & Decety, J. (2019). The development of children's preferences for equality and equity across 13 individualistic and collectivist cultures. *Developmental Science*, 22(2), 12729.

Iwasaki, M. (2023). Social preferences and well-being: theory and evidence. *Humanities and Social Sciences Communications*, 10(1), 1–13.

Jackson, J. C., Halberstadt, J., Takezawa, M., Kongmeng, L., Smith, K. M., Apicella, C., & Gray, K. (2023). Generalized morality culturally evolves as an adaptive heuristic in large social networks. *Journal of Personality and Social Psychology – Attitudes and Social Cognition*, 125, 1207–1238.

Jin, K. S., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences*, 114(31), 8199–8204.

Johnson, M. K., Rowatt, W. C., & LaBouff, J. P. (2012). Religiosity and prejudice revisited: In-group favoritism, out-group derogation, or both? *Psychology of Religion and Spirituality*, 4(2), 154–168.

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., & Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *Journal of Neuroscience*, 31(37), 13039–13045.

Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children's third-party punishment of selfishness. *Proceedings of the National Academy of Sciences*, 111(35), 12710–12715.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10(12), 1625–1633.

Kay, A. C., Whitson, J. A., Gaucher, D., & Galinsky, A. D. (2009). Compensatory control: Achieving order through the mind, our institutions, and the heavens. *Current Directions in Psychological Science*, 18(5), 264–268.

Kelly, C. A., & Sharot, T. (2021). Individual differences in information-seeking. *Nature Communications*, 12, 7062.

Killen, M., Smetana, J. G., & in Chief, R. M. L. (2015). Origins and development of morality. In M. E. Lamb (Ed.), *Handbook of child psychology and developmental science* (7th ed., Vol. 3, pp. 701–749). Wiley-Blackwell.

Kiyonari, T., Tanida, S., & Yamagishi, T. (2000). Social exchange and reciprocity: confusion or a heuristic? *Evolution and Human Behavior*, 21(6), 411–427.

Kleinfeld, R. (2023). *Polarization, democracy, and political violence in the United States: What the research Says* (pp. 4–74). Carnegie Endowment for International Peace.

Kodapanakkal, R. I., Brandt, M. J., Kogler, C., & Beest, I. (2022). Moral frames are persuasive and moralize attitudes; nonmoral frames are persuasive and de-moralize attitudes. *Psychological Science*, 33(3), 433–449.

Kraaijeveld, S. R., & Jamrozik, E. (2022). Moralization and mismoralization in public health. *Medicine, Health Care and Philosophy*, 25(4), 655–669.

Krauzlis, R. J., Bollimunta, A., Arcizet, F., & Wang, L. (2014). Attention as an effect not a cause. *Trends in Cognitive Sciences*, 18(9), 457–464.

Krebs, D. L. (2008). Morality – An evolutionary account. *Perspectives on Psychological Science*, 3, 149–172.

Krosnick, J. A., & Petty, R. E. (2014). Attitude strength: An overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength* (pp. 1–24). Psychology Press.

Krumrei-Mancuso, E. J., & Newman, B. (2020). Intellectual humility in the sociopolitical domain. *Self and Identity*, 19(8), 989–1016.

Kundu, P., & Cummins, D. D. (2013). Morality and conformity: The Asch paradigm applied to moral decisions. *Social Influence*, 8(4), 268–279.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, 98(26), 15387–15392.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, 18(8), 1159–1167.

Lee, J., & Holyoak, K. (2020). But he's my brother: The impact of family obligation on moral judgments and decisions. *Memory & Cognition*, 48(1), 158–170.

Levy, D. J., & Glimcher, P. W. (2011). Comparing apples and oranges: using reward-specific and reward-general subjective value representation in the brain. *Journal of Neuroscience*, 31, 14693–14707.

Lewin, K. (1951). *Problems of research in social psychology. Field theory in social science: Selected theoretical papers*. Harpers.

Littman, L. (2018). Parent reports of adolescents and young adults perceived to show signs of a rapid onset of gender dysphoria. *PLoS One*, 13(8), 0202330.

Lu, H. J., & Chang, L. (2016). Resource allocation to kin, friends, and strangers by 3-to 6-year-old children. *Journal of Experimental Child Psychology*, 150, 194–206.

Luttrell, A., & Togans, L. J. (2021). The stability of moralized attitudes over time. *Personality and Social Psychology Bulletin*, 47(4), 551–564.

Malhotra, D. (2010). When) are religious people nicer? Religious salience and the “Sunday Effect” on pro-social behavior. *Judgment and Decision Making*, 5(2), 138–143.

Maniscalco, B., & Lau, H. C. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430.

Marean, C. W. (2015). The most invasive species of all. *Scientific American*, 313(2), 32–39.

Marie, A., Altay, S., & Strickland, B. (2023). Moralization and extremism robustly amplify myside sharing. *PNAS nexus*, 2(4), 078.

Marques, J. M., & Yzerbyt, V. Y. (1988). The black sheep effect: Judgmental extremity towards ingroup members in inter- and intragroup situations. *European Journal of Social Psychology*, 18, 287–292.

Martínez, C. A., Prooijen, J. W., & Lange, P. A. (2022). Hate: Toward understanding its distinctive features across interpersonal and intergroup targets. *Emotion*, 22(1), 46–63.

Martin, G. J., & Yurukoglu, A. (2017). Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9), 2565–2599.

Mastroianni, A. M., & Gilbert, D. T. (2023). The illusion of moral decline. *Nature*, 618, 782–789.

McManus, R. M., Kleiman-Weiner, M., & Young, L. (2020). What we owe to family: The impact of special obligations on moral judgment. *Psychological Science*, 31(3), 227–242.

Meier, S., & Stutzer, A. (2008). Is volunteering rewarding in itself? *Economica*, 75(297), 39–59.

Mercier, H. (2011). What good is moral reasoning? *Mind & Society*, 10, 131–148.

Miller, R., & Cushman, F. (2018). Moral values and motivations: How special are they? In K. Gray & J. Graham (Eds.), *Atlas of Moral Psychology* (pp. 59–69). Guilford Press.

Montague, R. (2007). *Your brain is almost perfect*. Penguin Group.

Mooijman, M., Hoover, J., Lin, Y., Ji, H., & Dehghani, M. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, 2(6), 389–396.

Mooney, C. (1999). The politics of morality policy: Symposium editor's introduction. *Policy Studies Journal*, 27(4), 675–680.

Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *Journal of Neuroscience*, 38(14), 3534–3546.

Morgan, G. S., & Skitka, L. J. (2020). Evidence for meta-ethical monism: Moral conviction predicts perceived objectivity and universality across issues. *Annual Meeting of the Society for Personality and Social Psychology*.

Noë, R., & Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behavioral Ecology and Sociobiology*, 35, 1–11.

- Norenzayan, A. (2014). *Big gods: how religion transformed cooperation and conflict*. Princeton University Press.
- Nucci, L. P. (1996). Morality and personal freedom. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 41–60). Erlbaum.
- Nucci, L. P. (2001). *Education in the moral domain*. Cambridge University Press.
- Pauls, I. L., Shuman, E., Zomeran, M., Saguy, T., & Halperin, E. (2022). Does crossing a moral line justify collective means? Explaining how a perceived moral violation triggers normative and nonnormative forms of collective action. *European Journal of Social Psychology*, 52(1), 105–123.
- Pennycook, G., Cheyne, J. A., Seli, P., Koehler, D. J., & Fugelsang, J. A. (2012). Analytic cognitive style predicts religious and paranormal belief. *Cognition*, 123(3), 335–346.
- Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2016). Atheists and agnostics are more reflective than religious believers: Four empirical studies and a metaanalysis. *PLoS One*, 11(4), 0153039.
- Pessiglione, M., & Delgado, M. R. (2015). The good, the bad and the brain: neural correlates of appetitive and aversive values underlying decision making. *Current Opinion in Behavioral Sciences*, 5, 78–84.
- Petersen, M. B. (2015). Evolutionary political psychology: On the origin and structure of heuristics and biases in politics. *Political Psychology*, 36, 45–78.
- Petersen, M. B., Osmundsen, M., & Tooby, J. (2022). The evolutionary psychology of conflict and the functions of falsehood. In D. C. Baker & E. Suhay (Eds.), *The politics of truth in polarized America concepts* (pp. 131–151). Oxford University Press.
- Petty, R. E., & Cacioppo, J. T. (1981). *Attitudes and persuasion: Classic and contemporary approaches*. Wm. C. Brown.
- Petty, R. E., & Wegner, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Dual-process theories in social psychology* (pp. 41–72). McGraw-Hill.
- Pincus, M., LaViers, L., Prietula, M. J., & Berns, G. (2014). The conforming brain and deontological resolve. *PLoS One*, 9(8), 106061.
- Porges, E. C., & Decety, J. (2013). Violence as a source of pleasure or displeasure is associated with specific functional connectivity with the nucleus accumbens. *Frontiers in Human Neuroscience*, 7, 447.
- Porter, T., Elnakouri, A., Meyers, E. A., Shibayama, T., Jayawickreme, E., & Grossmann, I. (2022). Predictors and consequences of intellectual humility. *Nature Reviews Psychology*, 1(9), 524–536.

- Posner, E. A., & Sunstein, C. R. (2017). Moral commitments in cost-benefit analysis. *Virginia Law Review*, 103, 1809–1860.
- Pretus, C., Hamid, N., Sheikh, H., Gómez, Á., Ginges, J., Tobeña, A., & Atran, S. (2019). Ventromedial and dorsolateral prefrontal interactions underlie will to fight and die for a cause. *Social Cognitive and Affective Neuroscience*, 14(6), 569–577.
- Pretus, C., Ray, J. L., Granot, Y., Cunningham, W. A., & Bavel, J. (2023). The psychology of hate: moral concerns differentiate hate from dislike. *European Journal of Social Psychology*, 53(2), 336–353.
- Prooijen, J. W., Krouwel, A. P. M., & Emmer, J. (2018). Ideological responses to the EU refugee crisis: The left, the right, and the extremes. *Social Psychological & Personality Science*, 9, 143–150.
- Prooijen, J. W., & Krouwel, A. P. (2019). Psychological features of extreme political ideologies. *Current Directions in Psychological Science*, 28(2), 159–163.
- Proust, J. (2013). *The philosophy of metacognition*. Oxford University Press.
- Quervain, D. J. F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., & Fehr, E. (2004). The neural basis of altruistic punishment. *Science*, 305, 1254–1258.
- Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, 118(1), 57–75.
- Rathje, S., Bavel, J. J., & Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), 2024292118.
- Reagan, L. J. (2022). *When abortion was a crime: women, medicine, and law in the United States*. University of California Press.
- Repucci, S., & Slipowitz, A. (2022). *Freedom in the world 2022: The annual survey of political and civil liberties*. Freedom House.
- Richerson, P. J., & Boyd, R. (2005). *Not by genes alone*. Chicago University Press.
- Richerson, P. J., Boyd, R., & Henrich, J. (2010). Gene-culture coevolution in the age of genomics. *Proceedings of the National Academy of Sciences*, 107, 8985–8992.
- Rilling, J. K., DeMarco, A. C., Hackett, P. D., Thompson, R., Ditzen, B., Patel, R., & Pagnoni, G. (2012). Effects of intranasal oxytocin and vasopressin on cooperative behavior and associated brain activity in men. *Psychoneuroendocrinology*, 37(4), 447–461.

Rollwage, M., Dolan, R. J., & Fleming, S. M. (2018). Metacognitive failure as a feature of those holding radical beliefs. *Current Biology*, 28, 4014–4021.

Rothschild, Z. K., Abdollahi, A., & Pyszczynski, T. (2009). Does peace have a prayer? The effect of mortality salience, compassionate values, and religious fundamentalism on hostility toward out-groups. *Journal of Experimental Social Psychology*, 45(4), 816–827.

Roughgarden, J. (2004). Evolution, gender, and rape. *Ethology*, 10, 76–77.

Royzman, E. B., McCauley, C., & Rozin, P. (2005). From Plato to Putnam: four ways to think about hate. In R. J. Sternberg (Ed.), *The psychology of hate* (pp. 3–35). American Psychological Association.

Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and becoming a vegetarian: The transformation of preferences into values and the recruitment of disgust. *Psychological Science*, 8(2), 67–73.

Rozin, P. (1999). The process of moralization. *Psychological Science*, 10(3), 218–221.

Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, 15(8), 549–562.

Ryan, T. J. (2014). Reconsidering moral issues in politics. *The Journal of Politics*, 76(2), 380–397.

Ryan, T. J. (2019). Actions versus consequences in political arguments: Insights from moral psychology. *The Journal of Politics*, 81(2), 426–440.

Sabucedo, J. M., Dono, M., Alzate, M., & Seoane, G. (2018). The importance of protesters' morals: Moral obligation as a key variable to understand collective action. *Frontiers in Psychology*, 9, 418.

Sarkar, A., & Wrangham, R. W. (2023). Evolutionary and neuroendocrine foundations of human aggression*. *Trends in Cognitive Sciences*, 27(5), 468–493.

Saroglou, V., Delpierre, V., & Dernelle, R. (2004). Values and religiosity: A meta-analysis of studies using Schwartz's model. *Personality and Individual Differences*, 37(4), 721–734.

Schjødt, U., Stødkilde-Jørgensen, H., Geertz, A. W., & Roepstorff, A. (2008). Rewarding prayers. *Neuroscience Letters*, 443(3), 165–168.

Schmidt, M. F., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PloS One*, 6(10), 23223.

- Schmidt, M. F., Rakoczy, H., & Tomasello, M. (2019). Eighteen-month-old infants correct non-conforming actions by others. *Infancy*, 24(4), 613–635.
- Schnepper, R., Blechert, J., & Stok, F. M. (2022). Reception of health messages: effects of stigmatization and forcefulness. *Journal of Public Health*, 44(2), 387–393.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275(5306), 1593–1599.
- Schulz, L., Rollwage, M., Dolan, R. J., & Fleming, S. M. (2020). Dogmatism manifests in lowered information search under uncertainty. *Proceedings of the National Academy of Sciences*, 117(49), 31527–31534.
- Schwartz, S. H. (2014). National culture as value orientations: Consequences of value differences and cultural distance. In V. Ginsburgh & D. Thorsby (Eds.), *Handbook of the Economics of Art and Culture* (pp. 547–586). Elsevier.
- Shamir, J. (1997). Speaking up and silencing out in face of a changing climate of opinion. *Journalism & Mass Communication Quarterly*, 74(3), 602–614.
- Shapiro, A. D., & Grafton, S. T. (2020). Subjective value then confidence in human ventromedial prefrontal cortex. *PloS One*, 15(2), 0225617.
- Sharot, T., & Sunstein, C. R. (2020). How people decide what they want to know. *Nature Human Behavior*, 4, 14–19.
- Shinada, M., Yamagishi, T., & Ohmura, Y. (2004). False friends are worse than bitter enemies: Altruistic punishment of in-group members. *Evolution and Human Behavior*, 25(6), 379–393.
- Siegel, S. (1986). Alcohol and opiate dependence: Re-evaluation of the Victorian perspective. In H. D. Cappell, F. B. Glaser, Y. Israel, H. Kalant, W. Schmidt, E. M. Sellers, & R. C. Smart (Eds.), *Research advances in alcohol and drug problems* (Vol. 9, pp. 279–314). Plenum Press.
- Simas, E. N., Clifford, S., & Kirkland, J. H. (2020). How empathic concern fuels political polarization. *American Political Science Review*, 114(1), 258–269.
- Simpson, B., & Willer, R. (2015). Beyond altruism: Sociological foundations of cooperation and prosocial behavior. *Annual Review of Sociology*, 41, 43–63.
- Skitka, L. J., & Mullen, E. (2002). The dark side of moral conviction. *Analyses of Social Issues and Public Policy*, 7, 35–41.
- Skitka, L. J., Bauman, C. W., & Mullen, E. (2008). Morality and justice: An expanded theoretical perspective and empirical review. In K. A. Hegtvedt & J. Clay-Warner (Eds.), *Justice* (pp. 1–27). Emerald Group Publishing.

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, 4(4), 267–281.

Skitka, L. J., & Morgan, G. S. (2014). The social and political implications of moral conviction. *Political Psychology*, 35, 95–110.

Skitka, L. J., Hanson, B. E., Morgan, G. S., & Wisneski, D. C. (2021). The psychology of moral conviction. *Annual Review of Psychology*, 72, 347–366.

Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, 23, 196–204.

Slovic, P., Mertz, C. K., Markowitz, D. M., Quist, A., & Västfjäll, D. (2020). Virtuous violence from the war room to death row. *Proceedings of the National Academy of Sciences*, 117(34), 20474–20482.

Smetana, J. G. (2013). Moral development: The social domain theory view. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology: Body and mind* (pp. 832–863). Oxford University Press.

Sripada, C. S., & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind, Volume 2: Culture and cognition* (pp. 280–301). Oxford University Press.

Stallen, M., Rossi, F., Heijne, A., Smidts, A., Dreu, C. K., & Sanfey, A. G. (2018). Neurobiological mechanisms of responding to injustice. *Journal of Neuroscience*, 38(12), 2944–2954.

Stanley, M. L., Marsh, E. J., & Kay, A. C. (2021). Structure-seeking as a psychological antecedent of beliefs about morality. *Journal of Experimental Psychology: General*, 149(10), 1908–1918.

Staub, E. (2005). The Origins and evolution of hate. In R. J. Sternberg (Ed.), *The psychology of hate* (pp. 51–66). American Psychological Association.

Stuppel, E. J. N., Gale, M., & Richmond, C. (2013). Working memory, cognitive miserliness and logic as predictors of performance on the cognitive reflection test. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Swann, W. B., Jr., Gómez, A., Buhrmester, M. D., LópezRodríguez, L., Jiménez, J., & Vázquez, A. (2014). Contemplating the ultimate sacrifice: Identity fusion channels pro-group affect, cognition, and moral decision-making. *Journal of Personality and Social Psychology*, 106, 713–727.

Swann, W., Buhrmester, J., & D, M. (2015). Identity fusion. *Current Directions in Psychological Science*, 24, 52–57.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Nelson-Hall.

- Tetlock, P. E. (2002). Social functionalist frameworks for judgment and choice: intuitive politicians, theologians, and prosecutors. *Psychological Review*, 109(3), 451–471.
- Tetlock, P. E. (2003). Thinking the unthinkable: Sacred values and taboo cognitions. *Trends in Cognitive Sciences*, 7(7), 320–324.
- Tetlock, P. E., Mellers, B. A., & Scoblic, J. P. (2017). Sacred versus pseudo-sacred values: How people cope with taboo trade-offs. *American Economic Review*, 107(5), 96–99.
- Thomas, E. F., Bury, S. M., Louis, W. R., Amiot, C. E., Molenberghs, P., Crane, M. F., & Decety, J. (2019). Vegetarian, vegan, activist, radical: Using latent profile analysis to examine different forms of support for animal welfare. *Group Processes & Intergroup Relations*, 22(6), 836–857.
- Ting, F., Buyukozur Dawkins, M., Stavans, M., & Baillargeon, R. (2020). Principles and concepts in early moral cognition. In J. Decety (Ed.), *The social brain: A developmental perspective* (pp. 41–65). MIT Press.
- Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology*, 44(3), 187–194.
- Tomasello, M. (2016). *A natural history of human morality*. MIT Press.
- Toner, K., Leary, M. R., Asher, M. W., & Jongman-Sereno, K. P. (2013). Feeling superior is a bipartisan issue. *Psychological Science*, 24, 2454–2462.
- Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). Oxford University Press.
- Tooby, J., & Cosmides, L. (2010). Groups in Mind: The coalitional roots of war and morality. In H. Høgh-Olesen (Ed.), *Human morality & sociality: Evolutionary & comparative perspectives* (pp. 91–234). Palgrave MacMillan.
- Tooby, J. (2017). Coalitional instincts. *The Edge*. <https://www.edge.org/response-detail/27168>
- Toorn, J., Ellemers, N., & Doosje, B. (2015). The threat of moral transgression: The impact of group membership and moral opportunity. *European Journal of Social Psychology*, 45, 609–622.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics and biases tasks. *Memory and Cognition*, 39(7), 1275–1289.
- Tormala, Z. L., & Rucker, D. D. (2015). How certainty transforms persuasion. *Harvard Business Review*.

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

Turiel, E. (2006). The development of morality. In N. Eisenberg (Ed.), *Handbook of child psychology: Social, emotional and personality development* (pp. 789–857). Wiley.

Turiel, E. (2012). *The culture of morality: Social development, context, and conflict*. Cambridge University Press.

Ugazio, G., Grueschow, M., Polania, R., Lamm, C., Tobler, P., & Ruff, C. (2022). Neuro-computational foundations of moral preferences. *Social Cognitive and Affective Neuroscience*, 17(3), 253–265.

Vaughn, D. A., Savjani, R. R., Cohen, M. S., & Eagleman, D. M. (2018). Empathic neural responses predict group allegiance. *Frontiers in Human Neuroscience*, 12, 302.

Victoroff, J. (2005). The mind of the terrorist: A review and critique of psychological approaches. *Journal of Conflict Resolution*, 49(1), 3–42.

W, G. P. (2011). *Foundations of neuroeconomics analysis*. Oxford University Press.

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy*, 11(3), 271–294.

Wilson, E. O. (1998). *Consilience*. Random house.

Workman, C. I., Yoder, K. J., & Decety, J. (2020). The dark side of morality – Neural mechanisms underpinning moral convictions and support for violence. *American Journal of Bioethics - Neuroscience*, 11(4), 269–284.

Wright, J. C., Cullum, J., & Schwab, N. (2008). The cognitive and affective dimensions of moral conviction: Implications for tolerance and interpersonal behaviors. *Personality and Social Psychology Bulletin*, 34(11), 1461–1476.

Wright, J. C., McWhite, C. B., & Grandjean, P. T. (2014). The cognitive mechanisms of intolerance: Do our meta-ethical commitments matter? In J. Knobe, T. Lombrozo, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (pp. 28–61). Oxford University Press.

Wright, J. D. (2016). More religion, less justification for violence: A cross-national analysis. *Archive for the Psychology of Religion*, 38(2), 159–183.

Wright, J. D., & Khoo, Y. (2019). Empirical perspectives on religion and violence. *Contemporary Voices*, 1–26.

- Wright, J. C., & Pölzler, T. (2022). Should morality be abolished? An empirical challenge to the argument from intolerance. *Philosophical Psychology*, 35(3), 350–385.
- Wright, J. C. (2023). The psycho-social function of moral conviction. In A. Finger & M. Wagner (Eds.), *Bias, belief, and conviction in an age of fake facts* (pp. 117–133). Routledge.
- Wu, H., Luo, Y., & Feng, C. (2016). Neural signatures of social conformity: A coordinate-based activation likelihood estimation meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, 71, 101–111.
- Xygalatas, D. (2013). Effects of religious setting on cooperative behavior: A case study from Mauritius. *Religion, Brain & Behavior*, 3(2), 91–102.
- Xygalatas, D., Klocová, E. K., Cigán, J., Kundt, R., Maño, P., Kotherová, S., & Kanovsky, M. (2016). Location, location, location: effects of cross-religious primes on prosocial behavior. *International Journal for the Psychology of Religion*, 26(4), 304–319.
- Yamagishi, T., & Mifune, N. (2009). Social exchange and solidarity: In-group love or out-group hate? *Evolution and Human Behavior*, 30(4), 229–237.
- Yoder, K. J., & Decety, J. (2014). The good, the bad, and the just: Justice sensitivity predicts neural response during moral evaluation of actions performed by others. *Journal of Neuroscience*, 34(12), 4161–4166.
- Yoder, K. J., & Decety, J. (2014). Spatiotemporal neural dynamics of moral judgment: A high-density ERP study. *Neuropsychologia*, 60, 39–45.
- Yoder, K. J., & Decety, J. (2018). The neuroscience of morality and social decision-making. *Psychology, Crime & Law*, 24(3), 279–295.
- Yoder, K. J., & Decety, J. (2022). Moral conviction and metacognitive ability shape multiple stages of information processing during social decision-making. *Cortex*, 151, 162–175.
- Young, A. I., & Fazio, R. H. (2013). Attitude accessibility as a determinant of object construal and evaluation. *Journal of Experimental Social Psychology*, 49(3), 404–418.
- Zaal, M. P., Laar, C. V., Ståhl, T., Ellemers, N., & Derks, B. (2011). By any means necessary: The effects of regulatory focus and moral conviction on hostile and benevolent forms of collective action. *British Journal of Social Psychology*, 50(4), 670–689.
- Zamboni, G., Gozzi, M., Krueger, F., Duhamel, J. R., Sirigu, A., & Grafman, J. (2009). Individualism, conservatism, and radicalism as criteria for processing political beliefs: a parametric fMRI study. *Social Neuroscience*, 4(5), 367–383.
- Zhu, P., Liu, W., Zhang, X., Li, M., Liu, G., Yu, Y., & Zhou, X. (2023). Correlated evolution of social organization and lifespan in mammals. *Nature Communications*, 14(1), 372.

Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2020). The partisan mind: Is extreme political partisanship related to cognitive inflexibility? *Journal of Experimental Psychology: General*, 149(3), 407–418.

Zmigrod, L., Eisenberg, I. W., Bissett, P. G., Robbins, T. W., & Poldrack, R. A. (2021). The cognitive and perceptual correlates of ideological attitudes: a data-driven approach. *Philosophical Transactions of the Royal Society B*, 376, 20200424.

Footnotes

1 : On May 23rd, 2023, the Tel Aviv International Salon canceled the launch of Abigail Shrier's book "*Irreversible Damage: The Transgender Craze Seducing Our Daughters*" in response to transgender activists (The Jerusalem Post – Lahav Harkov). Years before, when the book was released in the US, Amazon would not allow Shrier to advertise it, and employees of the book-selling giant petitioned for it not to be sold at all. It was twice removed from Target's shelves. A lawyer from the American Civil Liberties Union wrote on Twitter "*stopping the circulation of this book and these ideas is 100% a hill I will die on.*" The book makes the case that social contagion plays a significant role in the epidemic of rapid gender dysphoria seen during the past decades, especially for young girls. Several large-scale empirical studies support the social contagion hypothesis (Diaz & Bailey, 2023; Littman, 2018). None of the views expressed by Shrier are disrespectful to transgenders. There is not a drop of animosity in the book. [!\[\]\(35e4f762fc1cfea5610d92e2d225d5b4_img.jpg\)](#)

2 : A naturalistic perspective rejects the dualistic ontology of a natural/human opposition, mental states/neurophysiological states, or nature/culture. It transcends polar oppositions and seemingly clear-cut dichotomies. Methodologically, it involves attempting to limit philosophical inquiries to empirical methods whose validity can be established and vindicated by the scientific process. And as a corollary, it excludes metaphysical entities which do not follow physical laws. There is a continuum between nature and culture, and all social and cultural phenomena built on biological properties. Culture is a natural fact, not independent from nature. The most widely distributed traits of culture confer Darwinian advantage on the genes that predispose them (Wilson, 1998). Moreover, and contrary to what is expressed by social scientists who hold dualist anti-naturalist, a naturalistic perspective does not seek to reduce the social sciences to biology. It offers the social sciences a set of theoretical and methodological tools that can be used in explaining social phenomena, and allows to integrate knowledge across levels of organization. All social phenomena imply that information is captured and processed by the individuals who are involved (i.e. information processing computations, representations, and valuations that put meaning into computations). Understanding how information processing is done and its causal role to behavior can only help us explain more objectively and predict more accurately how social and cultural processes unfold, and overall enrich our knowledge. Molar constructs such as those developed by the social sciences provide a means of understanding highly complex activity without needing to specify each individual action's simplest components, thereby providing an efficient means of describing the behavior of a complex system (Decety & Cacioppo, 2011). [!\[\]\(d84e7ea36f695d92cb39ec32c307ac93_img.jpg\)](#)

3 : In her review, Roughgarden writes that: "Zoologists may not know much about the comparative anthropology of sexual coercion, and scientists may not know much about proven techniques in rape-education programs. To bridge these disciplinary gaps, this book is a necessary handbook for anyone from all disciplines, anthropology through zoology, concerned with sexual coercion. The book is readable, fast-paced, often humorous, and the chapters hang together, overall a great success. A success in scholarship that is, not politics. Evolution, Gender and Rape preaches to the choir. No press conferences heralded its release, nor were the authors featured on Oprah. TP's crime is political, not scholarly." [!\[\]\(feabb98897b440bc8695a03336a6e2df_img.jpg\)](#)

4 : Authoritarianism fosters attitudes and behaviors that serve to enhance sameness and minimize diversity of people, ideology, and behavior in one's social environment. It entails both coercion of different others and approval of the use of group authority (e.g., coercion by the state), which may entail legal discrimination, limits on free speech and association, and state regulation of moral behavior (Costello et al., 2021). Across the globe, there is a resurgence of authoritarianism (largely right-wing) and a global decline in citizens' political rights and civil liberties (Repucci & Slipowitz, 2022). [!\[\]\(9dfdaff1d86ba3c1f8353b4d1b61b8c5_img.jpg\)](#)

5 : The construct of social preferences is used by economists to describe the human tendency to not only care about one's own material payoff, but also the reference group's payoff or/and the intention that leads to the payoff (Carpenter, 2008). Social preferences are studied extensively in behavioral and experimental economics and social psychology. Types of social preferences include altruism, fairness, reciprocity, inequality and inequity aversion.↵

6 : In game theory, a non-zero-sum game is a situation in which one participant's gain is not necessarily another's loss – in which the gains and losses do not sum to zero. Trade is a non-zero-sum game since seller and buyer can both improve their positions. Morality-as-cooperation draws on the theory of non-zero-sum games to identify distinct problems of cooperation and their solutions. The theory predicts that specific forms of cooperative behavior—including helping kin, helping your group, reciprocating, being brave, deferring to superiors, dividing disputed resources, and respecting prior possession—will be considered morally good wherever they arise (Curry et al., 2019).↵

7 : We adopt heuristics to avoid incurring cognitive costs associated with deliberation. In a world with repeated interactions, it is usually worthwhile to cooperate, and therefore, individuals may adopt heuristics, such as always cooperate with members of their coalition (Hoffman et al., 2015). Deliberative decision-making leads to slower reaction time. There is evidence that intuitive cooperators are trusted more than reflective cooperators. For instance, in an experiment eliciting moral judgments, participants who read vignettes about people who returned lost wallets judged those who returned the wallets without hesitation more positively than those who hesitated (Chritcher et al., 2013).↵

8 : The idea of the social exchange heuristic assumes that making such an error is more adaptive in social exchange situations than making the logically correct decisions (Kiyonari, Tanida, & Yamagishi, 2000). It is true that the one who makes such an error in a Prisoners' Dilemma (PD) experiment loses an opportunity to make additional money. However, the consequence of missing the opportunity to attain mutual cooperation in a real life social exchange would often be much more serious than missing a few extra dollars in a one-shot PD experiment. The social exchange heuristic “biases” us to seek mutual cooperation in social exchange, diverting us from attempts to exploit exchange partners. Whether this “bias” improves the level of our adaptive success or reduces it depends on the relative importance of attainment of mutual cooperation vis-à-vis one-sided exploitation in social exchange.↵

9 : In social psychology, attitudes refer to general and enduring evaluative (good/bad, harmful/beneficial, wise/foolish) predispositions toward a stimulus or category of stimuli (e.g., person, object, issue, position, group) (Petty & Cacioppo, 1981). Attitudes serve important social and psychological functions and can influence decisions and behaviors. They serve as convenient summaries for one's beliefs, emotions, and preferences regarding issues, objects, and people. They facilitate the differentiation of hostile from hospitable stimuli. They help us to know what to expect when exposed to a stimulus. They reduce the stress of decision-making (Cacioppo & Berntson, 1994; Fazio, 1995; Petty & Wegner, 1998). However, one feature that makes attitudes distinct from moral convictions is their capacity to change in light of new information, goals, and challenges. Conversely, a conviction is an unshakable and absolute belief that something is right or wrong without necessarily proof or evidence.↵

10 : Addiction is certainly too strong a word in this context, or at least debatable. For the World Health Organization, addiction is a neurobiological disease, defined as a compulsion to use a drug and the onset of withdrawal symptoms when consumption stops. At first restricted to drug and alcohol misuse, the concept of addiction has been subsequently extended to “pathological” behaviors such as exercise addiction, compulsive buying, sexual hyperactivity, and computer/video playing and gambling. This list is open. The important aspect across domains is that addiction is self-destructive.↵

11 : Metacognitive sensitivity refers to the proportion in which a positive confidence is reported for a correct performance (i.e., to the within participant correspondence between confidence and accuracy

in individual trials). The meta-d' model is widely used. It offers a performance-controlled metric of metacognitive sensitivity, by estimating the level of first-order performance (d') that would have given rise to the observed confidence data under a signal detection theoretic model (Maniscalco & Lau, 2012). The ratio meta- d'/d' thus provides a performance-controlled metric of metacognitive capacity (also referred as metacognitive efficiency).[↵](#)

12 : The main difference between euthanasia and assisted suicide is who performs the final, fatal act. Euthanasia refers to active steps taken to end someone's life to stop their suffering and the "final deed" is undertaken by someone other than the individual, for example, a physician. Assisted suicide is about helping someone to take their own life at their request – in other words, the final deed is undertaken by the person themselves. Euthanasia is legal in Australia, Belgium, Canada, Colombia, Luxemburg, Netherlands, Spain, and in 11 states in the US.[↵](#)