

The uncensored truth about morality

Chelsea Schein, Amelia Goranson and Kurt Gray consider why immoral acts always seem to be those that cause harm – especially to children

Moral disagreement is ubiquitous. People argue about the morality of abortion, taxation, immigration, pornography and censorship. But everyone agrees that morality is about harm. Not only do harmful acts seem immoral, but immoral acts seem harmful. Building on this fact, our article outlines a new harm-based theory of morality called ‘dyadic morality’. This theory explains many quirks of moral judgement, such as why pleas of ‘think of the children!’ feature so prominently in moral disagreement. Dyadic morality is supported by experiments in moral psychology, and challenges other popular theories of morality. It also suggests that people – of all political hues and nationalities – have fundamentally the same moral mind, but moral disagreement occurs when people disagree about which acts are harmful.

In December 2014 the British Board of Film Censors (BBFC) banned pornographic materials depicting acts such as consensual spanking, facesitting, urolagnia, and female ejaculation. According to a spokesman for the organisation, the ‘BBFC classification regime is a tried and tested system of what content is regarded as harmful for minors’ and the legislation was aimed at ‘safeguarding children’ (Quinn, 2014). However, critics of the move quickly pointed out that this list is not only ‘arbitrary’, but includes only harmless acts between consenting adults. The BBFC and its critics seem to see the world in two very different ways – one in which pornography is both immoral and harmful to children, and one in which pornography is neither immoral nor harmful. This alignment of perceived immorality and perceived harm is not coincidental, but instead reflects the central role of harm in moral cognition.

The power of harm within morality is intuitive, as many of the gravest moral violations are those that directly cause harm – murder, assault, theft, rape and abuse are acts that are both harmful and universally viewed as immoral. Harm is also central to rhetoric about immorality, as the issue of censorship clearly illustrates. When people want to argue that an act is wrong, they reference its harmfulness, often to children. However, some recent theories of morality (e.g. moral foundations theory: Haidt, 2012) have suggested that this harm-based rhetoric is just that – mere rhetoric. These theories claim that questions of harm are

relevant to only a fraction of moral judgements, and that arguments about harm reflect only convenient post-hoc justifications (Haidt, 2012).

However, the new theory of ‘dyadic morality’ pulls harm from the shadows back into the spotlight (Gray et al., 2012). Dyadic morality suggests that harm is not only the most important factor in moral judgment, but the very core of a universal moral template – a template that both detects harm in diverse moral transgressions (Schein & Gray, 2015) and fills it in when apparently absent (Gray et al., 2014). In this article, we first describe dyadic morality before addressing three apparent challenges to this theory: the intuitive nature of morality (intuitionism); the moralisation of ‘harmless’ transgressions in other cultures (pluralism); and cognitive differences across moral content (modularity).

Categorisation and the moral dyad

In essence, moral judgement is simply about categorisation. The question ‘Is X immoral?’ can be reworded as ‘Does X belong to the category of immorality?’. As decades of research in cognitive psychology suggest, stimuli are categorised by automatically comparing them to a category’s prototype (Murphy, 2004). This prototype (or ‘cognitive template’) emerges from the most common, salient, and important features of the category. For example, the template of ‘bird’ is something small, winged, seed eating and capable of flight. The better a stimulus matches the template, the more robustly it is categorised as part of that group, explaining why sparrows are judged to be birds faster than ostriches.

What is the template of immorality? It is based upon acts that are the most universally condemned, the most evolutionary important, and the most emotionally evocative. Harm fulfills all these criteria. Harmful acts such as murder and assault are condemned across

questions

Why do people morally condemn ostensibly ‘harmless’ violations?
Do liberals and conservatives have fundamentally different moral minds?

resources

Wegner, D.M. & Gray, K. (2016). *The mind club: Who thinks, what feels, and why it matters*. New York: Viking.
The Mind Perception and Morality Lab:
www.mpmlab.org

references

- Baron-Cohen, S. (2011). *The science of evil: On empathy and the origins of cruelty*. New York: Basic Books.
- Cameron, C.D., Lindquist, K.A. & Gray, K. (2015). A constructionist review of morality and emotions: No evidence for specific links between moral content and discrete emotions. *Personality and Social Psychology Review* [Advance online publication]. doi:10.1177/1088868314566683
- Cheng, J.S., Ottati, V.C. & Price, E.D. (2013). The arousal model of moral condemnation. *Journal of Experimental Social Psychology*, 49(6), 1012–1018.
- Gray, K. & Keeney, J.E. (2015). Impure, or just weird? Scenario sampling bias raises questions about the foundation of moral cognition. *Social Psychological and Personality Science* [Advance online publication]. doi:10.1177/1948550615592241
- Gray, K., Schein, C. & Ward, A.F. (2014). The myth of harmless wrongs in moral cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, 143(4), 1600–1615.
- Gray, K., Waytz, A. & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, 23, 206–215.
- Greene, J.D. (2013). *Moral tribes: Emotion, reason, and the gap between us and them*. New York: Penguin.
- Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. New York: Pantheon Books.
- Kohlberg, L. (1969). Stage and sequence:

cultures, strongly impair genetic survival, and consistently evoke powerful negative emotion (Baron-Cohen, 2011). Of course, many harmful acts (e.g. car accidents) are not immoral, so our moral template must have some additional features.

Research finds that our moral template involves two interacting minds – an intentional agent (i.e. perpetrator) harming a suffering patient (i.e. victim; Gray et al., 2012). In other words, the essence of immorality isn't merely 'harm' but 'harm caused by an agent'. This agent can be another person, a corporation, or a government, but it must be an entity perceived to be mentally capable of intention and action. Conversely, the



The BBFC and its critics seem to see the world in two very different ways

patient who receives the harm must be mentally capable of feeling pain and suffering, such as children, or puppies, or the elderly.

The combination of 'intentional agent and suffering patient' – or even more simply 'thinking doer and vulnerable feeler' – gives us a simple formula for understanding the moral world. People are most morally incensed when powerful thinking doers harm powerless vulnerable feelers (e.g. a CEO kicks a baby). This is exactly the combination of minds seen by the BBFC censors, who believed that profit-driven movie producers were harming children. Conversely, people are least morally incensed when powerless vulnerable feelers harm powerful thinking doers (e.g. a baby bites a CEO).

With this formula of 'thinking doer and vulnerable feeler,' we can predict people's moral outrage to any infraction with two questions: How much does the patient/victim suffer? How much does the agent/perpetrator intend the harm? We can also predict a situation's potential for moral outrage by considering the patient's potential for suffering and the agent's potential for thought and action.

Because this moral template involves two interacting minds, it is called the dyadic template, from the Greek work *dyo*, meaning two. Unlike the template for birds (or dogs or furniture), the moral template is a matter of perception. Whether a bird can fly is a matter of fact, but less certain is whether a perpetrator is capable of intentional thought, or a victim is capable of suffering. The inherent ambiguity of other minds means that good people can nevertheless have moral disagreement, because they see different mental capacities in potential perpetrators and victims.

Debates about whether children who kill should go to adult prison hinge on whether they are fully capable of thought and action – an issue of perception as much as fact. Likewise, debates about abortion hinge on whether fetuses are capable of feeling pain – which is also largely a matter of perception. Where

people perceive minds, they perceive the potential for evil.

Moral universals

The dyadic template not only allows for moral differences between people, but also suggests two powerful moral universals. The first is what is harmful seems wrong (Schein & Gray, 2015). As the moral template is rooted in perceptions of harm, the more apparently harmful an act, the more it is judged as immoral: first-degree murder (obviously intentional and harmful) is universally condemned, whereas pornography (ambiguous intention and harm) is less universally condemned, and hence a matter of debate.

Our lab tested the centrality of harm for determining whether or not something is immoral in a number of studies (Schein & Gray, 2015). One asked participants to volunteer the first immoral act that came to mind. If harm is central to moral judgement, people should volunteer something obviously harmful, which is exactly what they did. Over 90 per cent of the acts recalled were dyadic in nature, such as murder, abuse, theft and adultery. Another test asked people to rate the immorality of acts that violated different kinds of norms, for example those that were harmful, unfair, disloyal, disrespectful, or gross. Consistent with a dyadic template, harmful acts were seen as the most immoral. The importance of harm was also observed in automatic judgements, as reaction times to categorise an action as 'immoral' almost perfectly predicted reaction times to categorise an action as 'harmful'. Whether an act is immoral seems best predicted by its perceived harmfulness.

The second moral universal predicted by dyadic morality is what is wrong seems harmful (Gray et al., 2014). Imagine you knew nothing about an action beyond that people believed it to be truly evil. You would assume that this action was harmful – or at least that these people perceived it to be harmful. This

The cognitive-developmental approach to socialization. In T. Mischel (Ed.) *Cognitive development and epistemology* (pp.151–235). New York: Academic Press.

Murphy, G.L. (2004). *The big book of concepts*. Cambridge, MA: MIT Press.

Quinn, B. (2014, 2 December). Pornography law bans list of sexual acts from UK-made online films. *The Guardian*. Retrieved 16 June 2015

from tinyurl.com/noqu7q
Schein, C. & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality & Social Psychology Bulletin*, 41(8), 1147–1163.

Shweder, R.A. (2012). Relativism and universalism. In D. Fassin (Ed.) *A companion to moral anthropology* (pp.85–102). Chichester: Wiley-

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

Young, L. & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, 120(2), 202–214.

perceived harm is an inevitable consequence of a dyadic moral template, which not only determines how stimuli are categorised, but also shapes how stimuli are perceived after categorisation. Like any lens through which we view the world, schemas powerfully shape our perceptions, often without our knowledge.

Consider again the template of birds. Because the concept of flying is central to birds, you will automatically assume that something labelled 'bird' can fly. Because the concept of 'harm' is central to immorality, you will automatically assume that something labelled 'immoral' is harmful. Of course, with birds, you can objectively learn that some birds cannot fly. But because harm is subjective, people can seldom 'objectively' learn that something they see as evil is also harmless. Returning to the example of censorship, no amount of cajoling about the harmlessness of consensual spanking or facesitting is going to change perceptions of those who already perceive it as harmful (and wrong). No one ever says 'It's harmless and wrong'.

This link from wrong to harm (and back) isn't just rhetoric, but automatic and intuitive. In one study from our lab, people read about acts that seemed wrong despite causing no directly physical harm, such as masturbating to a picture of your dead sister (yes, we know it's bizarre – blame our reviewers). Consistent with a dyadic template, participants nevertheless labelled these 'objectively harmless' misdeeds as harmful. Were these perceptions mere effortful justification? No. In fact, people were especially likely to see harm when we impaired their ability to reason by forcing them to answer quickly.

Other studies reveal that these perceptions of harm even creep into other judgements. After reading about 'harmless immoral acts', people were more likely to see sad expressions in the faces of children – an experimental demonstration of why those against pornography inevitably see it as harming children. Importantly, perceptions of harm weren't driven by general feelings of 'badness,' as other judgements about these children were not more negative. Perceived harm leads to judgements of evil, and judgements of evil lead specifically to perceived harm.

Questions for dyadic morality

Although dyadic morality aligns with decades of research on categorisation and historical harm-centric accounts of moral



While it is true that liberals and conservatives often morally disagree, they share similar harm-based moral judgements

judgement (Turiel, 1983), it conflicts with one popular theory of morality – moral foundations theory (MFT: Haidt, 2012). Moral foundations theory argues against the overarching role of harm, instead advocating for distinct moral modules, each corresponding to different kinds of moral content, such as 'purity' or 'fairness'. In these theories, harm is merely one 'little switch' of the moral mind, and not a global template. Advocates of MFT have challenged dyadic morality primarily along three lines: intuitionism, moral pluralism, and content differences. We summarise and address these challenges here, using them as an opportunity to correct common misconceptions about the theory of dyadic morality.

Intuitionism

Historic accounts of moral judgement emphasised the importance of harm, and also of moral reasoning. These theories – such as Kohlberg's stage model (1969) – suggested that moral judgement relied upon careful deliberation and conscious reflection. Reacting against this 'rule of reason', more recent accounts have claimed that morality is typically a matter of intuitive judgements and affective reactions (Haidt, 2001). When people denounce (or fail to denounce) pornography, it is rarely because they have rationally considered the issue. Instead, their moral judgement reflects emotion-based intuitions. As one analogy suggests, our faculty of moral judgement is not an impartial judge weighing the evidence, but rather, it is an impassioned lawyer arguing in favour of its original position (Haidt, 2012).

Significant evidence supports the power of intuition in determining our moral judgement. Moral judgements are formed quickly, are influenced by

"...moral judgement reflects emotion-based intuitions"

incidental emotions, and can sometimes seem contradictory (Greene, 2013). As dyadic morality embraces the power of harm, some have assumed that it also embraces the reign of reason. Nothing could be further from the truth. The role of templates in categorisation – whether moral or non-moral – progresses intuitively and automatically. When people think about birds, they need not consciously ponder about the importance of flight as they categorise sparrows and ostriches. Instead, key template features are used automatically and effortlessly in categorisation.

As we reviewed above, studies reveal that the role of harm in moral judgement is both intuitive and automatic – consistent with a dyadic moral template. People rapidly see harm in moral violations, automatically use this perceived harm in forming their moral judgements, and effortlessly perceive harm in response to immorality (Gray et al., 2014; Schein & Gray, 2015). Of course, people can (and often do) use considerations of harm in subsequent deliberative moral reasoning, but initial perceptions of harm are automatic.

Moral pluralism

Anthropologists have long recognised that different cultures have different morals. In the West, there are no clear prohibitions about what you can eat after a loved one's death. On the other hand, in India, Oriya Hindu Brahmans believe it to be immoral

for the eldest son to eat chicken after his father's death (Shweder, 2012). To account for these differences, theories such as MFT suggest that our basis for morality must extend beyond harm to considerations

of spiritual purity (Haidt, 2012). This perspective is called moral pluralism, because it advocates for a plurality of moral concerns. On the surface, moral pluralism seems to argue against the harm-centric dyadic morality. However, a closer inspection suggest that dyadic morality is actually more consistent with moral pluralism than is MFT.

Although MFT embraces non-Western conceptions of morality, it fails to embrace non-Western notions of harm, rigidly defining it as only direct physical suffering. In contrast, dyadic morality suggests that harm is in the eye of the beholder, which means that even violations of spiritual purity can actually be grounded in concerns about harm. Consider the example of the Brahman son eating chicken – Hindus believe that the eldest son is responsible for processing

the father's 'death pollution' by eating a vegetarian diet (Shweder, 2012). By eating chicken, the son is thereby condemning the father's soul to eternal suffering. In other words, what superficially seems to be just about 'purity' is actually about harm. This idea is also supported by research from our lab, which finds that violations of 'purity' are judged to be simply a certain kind of perceived harm (Gray & Keeney, 2015).

Our research also finds little support for another oft-discussed 'cultural' difference in morality – MFT has suggested that US liberals and conservatives have different 'foundations' (Haidt, 2012). While it is true that liberals and conservatives often morally disagree, they share similar harm-based moral judgements, as revealed by studies in our lab. Conservatives oppose gay marriage not because it seems 'impure', but because they see it as harmful. One notable anti-gay activist believed that allowing gay rights would destroy the American family, throw society into chaos and – of course – harm children. As with the Hindu Brahmins, MFT dismisses these perceptions of harm as illegitimate, whereas dyadic morality embraces 'harm pluralism' – the diversity of perceived harm. With its embrace of both moral and harm pluralism, dyadic morality better embodies the anthropological tradition that MFT seeks to claim for itself.

Content differences

Dyadic morality suggests that – consistent with other forms of judgement – moral judgement is based upon comparison to an overarching template. In contrast, MFT believes that moral judgement is underlain by cognitive modules or 'little switches in the brain' (Haidt, 2012, p.123), each of which has fundamentally 'distinct cognitive computations' (Young & Saxe, 2011, p.203). This module view suggests that violations of one 'content' area (e.g. harmful acts like murder) are processed differently from those of another (e.g. purity acts like facesitting). The module view also suggests that different kinds of moral content are linked to different emotions, such that harm (e.g. murder) is linked to anger, and purity (e.g. facesitting) is linked to disgust.

Despite the common acceptance of these claims in our field (Haidt, 2012), our lab wondered if all was not as it seemed. We examined the research arguing for unique cognition and unique emotions in the content area of 'purity', and found no support for either claim. First, uniqueness requires distinctness,

and judgements of 'harm' and 'purity' are highly correlated ($r = .87$: Gray & Keeney, 2015). Second, previous studies had clear confounds – unaccounted-for third variables – which we suspected gave only the illusion that purity was special.

Studies arguing for unique cognition all used a specific set of scenarios. These scenarios represented harm through acts such as murder and child abuse, and represented purity through acts such as masturbating with a dead chicken or getting a tail via plastic surgery. The savvy reader will recognise that these scenarios differ not only in moral 'content' but also in severity and weirdness. The harm scenarios are punishable by decades in prison, unlike the purity scenarios. The harm scenarios are also much less weird (i.e. atypical) than the purity scenarios. We read about murder everyday in the paper, and can easily imagine motivations for killing another person. Conversely, we seldom read about tail-plastic-surgery, and have difficulty imagining why someone would do this. Our studies reveal that these differences in severity and weirdness account for the apparent uniqueness of purity. With careful experimental controls, the specialness of purity disappears altogether (Gray & Keeney, 2015).

Careful experimental controls also cause the apparent link between purity and disgust to vanish. Previous studies linking these concepts fail to include control conditions with other similar emotions (e.g. anger, fear), or use improper statistical techniques that inflate the appearance of specialness (Cameron et al., 2015). In our comprehensive review, we found only one study that met the threshold for experimental controls, and it found no evidence for a special disgust–purity link (Cheng et al., 2013) – consistent with an overarching dyadic template.

The evolution of moral theory

Our morals change over time. Smoking used to be a matter of personal preference, but now seems morally tinged. Pre-marriage cohabitation used to be deeply sinful, but now seems like a prudent choice before lifelong commitment. Likewise, the scientific understanding of morality changes over time. How people divide norm violations into 'inappropriate' versus 'immoral' was once thought to be a matter of both harm and rational reason. Years later, it was discovered that moral judgement was driven by intuitions and emotion (Haidt, 2001). However, in their rush to abandon reason, these intuitive theories also

largely abandoned harm. Inspired by classic research on categorisation (Murphy, 2004), anthropological pluralism (Shweder, 2012), and modern accounts of cognition (Cameron et al., 2015), dyadic morality has reinstalled harm to its rightful place.

A flurry of recent research reveals that harm is both central to moral judgements (Schein & Gray, 2015), and also shapes our perception of the moral world (Gray et al., 2014). This same research also reveals that moral differences between people do not reflect deep cognitive differences: Whether British or Indian, liberal or conservative, people have the same harm-based template based upon the perception of two minds – and it is these perceptions that drive moral debate. Whether people are pro-life or pro-choice, or for or against gay rights, depends upon the same simple (but ambiguous) question: Do they perceive a mind being harmed?

The perception of harm may be subjective, but it has the hot truth of reality to those who see it. The 'truthiness' of these perceptions suggests that arguing about perceived harm may be as difficult as arguing about the sexiness of fetishes. Depending on whom you ask, watching consenting adults spanking, screaming, and biting each other may be arousing or morally repugnant. As the British Board of Film Censors (BBFC) suggests, the issue of censorship does indeed depend on the question of whether children are harmed. But answers to the question of harm – like those of sexiness – are in the eye of the beholder.



Chelsea Schein
is at the University of North Carolina, Chapel Hill
chelsea.schein@unc.edu



Amelia Goranson
is at the University of North Carolina, Chapel Hill
agoranso@live.unc.edu



Kurt Gray
is at the University of North Carolina, Chapel Hill
kurtgray@unc.edu