



University of Kentucky
UKnowledge

Law Faculty Scholarly Articles

Law Faculty Publications

5-2015

The Limits of Moral Intuitions for Human Rights Advocacy

Andrew K. Woods

University of Kentucky College of Law, andrewkwoods@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/law_facpub



Part of the [Human Rights Law Commons](#)

Right click to open a feedback form in a new tab to let us know how this document benefits you.

Recommended Citation

Andrew K. Woods, *The Limits of Moral Intuitions for Human Rights Advocacy*, 9 L. & Ethics Hum. Rts. 91 (2015).

This Article is brought to you for free and open access by the Law Faculty Publications at UKnowledge. It has been accepted for inclusion in Law Faculty Scholarly Articles by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.

The Limits of Moral Intuitions for Human Rights Advocacy

Notes/Citation Information

The Law & Ethics of Human Rights, Volume 9, Issue 1, Pages 91–111, DOI: 10.1515/lehr-2015-0002 (May 2015).

Andrew K. Woods*

The Limits of Moral Intuitions for Human Rights Advocacy

Abstract: The central ambition of human rights advocacy is to get people to care, who might otherwise not, about the suffering of others. To accomplish this, human rights advocates often appeal to moral intuitions by telling stories that evoke moral outrage, indignation, or guilt. Are these sorts of appeals a good way to promote human rights? The conventional wisdom suggests that they are. But perhaps the conventional wisdom is incomplete – perhaps human rights advocates should treat moral intuitions with skepticism rather than uncritical embrace. In this brief essay, I argue that appeals to moral intuitions are problematic because moral intuitions can lead people to make decisions that are suboptimal from the standpoint of the human rights regime’s goals. I attempt to show, in other words, that one of the great assets of the human rights regime – its ability to harness our strong intuitive reaction to the suffering of others – is also one of its great limitations. To make this argument, I draw from the mind sciences literature on moral decision-making. The latest research in this domain suggests that our moral intuitions are fallible. A number of studies have shown, for example, that moral outrage and indignation can cause people to make decisions that they would not defend under cooler conditions. I focus on three particular sorts of moral judgment biases and explore their implications for human rights advocacy. I then evaluate two different normative claims one might make about these moral judgment biases and offer several concluding thoughts.

DOI 10.1515/lehr-2015-0002

Introduction

The central ambition of human rights advocacy is to get people to care, who might otherwise not, about the suffering of others. To accomplish this, human rights advocates make appeals to moral intuitions; that is, they try to get

*Corresponding author: Andrew K. Woods, University of Kentucky, Lexington, KY, USA, E-mail: andrewkwoods@uky.edu

audiences to feel something about the suffering of others. This is a longstanding practice: historians have noted that the earliest human rights campaigns told stories of individual suffering – often borrowing techniques from nascent art forms like the novel – to encourage audiences to empathize with faraway victims.¹ Today, human rights advocates use new technologies and social media to amplify their message, but they still rely on the same essential methods: stories of wrongdoing that provoke strong feelings of sympathy, guilt, and outrage.²

The most prominent recent example of this can be seen in the Kony 2012 campaign, a viral video project aimed at highlighting the crimes of African rebel Joseph Kony, who was indicted by the International Criminal Court, but remains at large.³ The video tells the story of an American man who travels to Africa, where he meets a young boy fleeing Joseph Kony's forces. The story is intuitively compelling – viewers feel sympathy for Kony's victims and moral outrage over Kony's crimes. The video has enjoyed enormous success: It has been viewed over 100 million times, making it one of the most viral videos of all time, and likely the farthest reaching human rights campaign to date.⁴

Is this a good way to advocate human rights? Conventional wisdom suggests that these sorts of appeals to moral intuitions are useful tools for political mobilization, tools that may be crucial to the success of the human rights regime. As Rorty notes, reason and deliberation will not persuade people to care about distant human rights abuses; human rights advocates must deploy “sentimental education”:

[T]he sort of sad, sentimental story which begins “Because this is what it is like to be in her situation – to be far from home, among strangers”, or “Because she might become your daughter-in-law”, or “Because her mother would grieve for her”. Such stories, repeated and varied over the centuries, have induced us, the rich, safe, powerful, people, to tolerate, and even to cherish, powerless people – people whose appearance or habits or beliefs at

1 See LYNN HUNT, *INVENTING HUMAN RIGHTS* 31–37 (2007) (tracing a brief history of the idea of universal human rights and noting that the concept built on audience experiences of empathy in art forms like the novel).

2 See generally STANLEY COHEN, *STATES OF DENIAL: KNOWING ABOUT ATROCITIES AND SUFFERING* (2001) (describing the tactics that human rights advocates use to arouse sympathy for victims of faraway suffering, a particularly difficult task given that audiences are bombarded daily with images of suffering in the news).

3 See Noam Cohen, *A Video Campaign and the Power of Simplicity*, N.Y. TIMES, Mar. 12, 2012, available at <http://www.nytimes.com/2012/03/12/business/media/kony-2012-video-illustrates-the-power-of-simplicity.html>.

4 See Kony 2012, available at <http://www.youtube.com/watch?v=Y4MnpzG5Sqc>.

first seemed an insult to our own moral identity, our sense of the limits of permissible human variation.⁵

Put most simply, the conventional wisdom suggests that because moral intuitions can be harnessed for political power, human rights advocates should embrace them.

But perhaps the conventional wisdom is incomplete – perhaps human rights advocates should treat moral intuitions with critical skepticism rather than uncritical embrace. While moral intuitions are powerful, they can mislead. In this essay, I argue that appeals to moral intuitions can be problematic because moral intuitions have known blind spots, flaws that have implications for human rights advocacy. I show, in other words, that one of the great assets of the human rights regime – its ability to harness our strong intuitive reaction to the suffering of others, including feelings of guilt, empathy, moral outrage, and indignation – is also one of its great limitations.

To make this argument, I draw from the mind sciences literature on moral decision-making. The latest research in this domain suggests that our moral intuitions are not infallible. In fact, a number of mind scientists have shown that intuitions like moral outrage and indignation can cause people to make sub-optimal decisions – decisions that they would not defend under cooler conditions. And while intuitive appeals about the plight of other humans can at times achieve amazing results – for example, prompting enormous transfers of aid in the wake of earthquakes or Tsunamis – they have been ineffective tools for preventing genocide and other human rights abuses *ex ante*.

Before going any further, a number of clarifications are in order. First, this is not an essay about the use of emotion in human rights discourse; the relationship between emotion and reason is enormously complex and I do not hope to broach the subject here.⁶ Rather, the inquiry is concerned with human rights advocacy that plays on our strong moral *intuitions* about justice – intuitions, I hope to show, that lead to suboptimal human rights policy. By “human rights advocacy” I refer to efforts by individuals and organizations to influence both the public and policymakers to adopt a particular human rights policy. Reliance on moral intuitions can lead policymakers to enact suboptimal human rights policy directly (relying on their own moral intuitions) or indirectly (responding to their constituents who are relying on their moral intuitions). When I say “suboptimal,” I mean suboptimal vis-à-vis the goals of the human rights regime;

⁵ See Richard Rorty, *Human Rights, Rationality and Sentimentality*, in *ON HUMAN RIGHTS* 111 (Stephen Shute & Susan Hurley eds., 1993).

⁶ For a meditation on the powerful and often critical role that emotions can play in morality, see generally MARTHA C. NUSSBAUM, *UPHEAVALS OF THOUGHT* (2001).

I am not concerned here with how intuitions can fail us in absolute moral terms, but rather how they can fail us as a matter of human rights policy.⁷

Second, this is a consequentialist analysis. The human rights regime is guided by a set of international agreements that outline the basic set of universal human rights. For the purposes of this essay, I assume that the goal of the human rights regime is to protect as many peoples' human rights as possible, or to minimize the scope and number of human rights violations that occur. For example, if prosecuting a rebel commander who committed crimes against humanity would destabilize the region and make it likely that new human rights abuses would occur, my approach may counsel against the prosecution. Someone might read this and think "but our intuitions to punish bad actors ought to be obeyed because we have a moral duty to punish human rights violators to the hilt, even if doing so produces more harms than not punishing them." I do not address this or other deontological arguments.⁸ Instead, I try to evaluate the consequentialist claim that a human rights regime that conforms to our moral intuitions can and will produce the most effective human rights policy, all things considered. Specifically, I show where I think moral intuitions – based on what we know from mind science research – are likely to produce suboptimal consequences for the human rights regime on its own terms.

To make my argument, I draw a rough distinction between two sorts of mental processes: automatic and deliberative. I treat intuitions as largely automatic mental activity – what mind scientists call System 1 processes – and I contrast these with more deliberative mental activity, such as reasoning through a math problem, which mind scientists refer to as System 2 processes.⁹ To some extent, of course, this is a false distinction. Damasio, among others, has shown quite persuasively that emotion (typically thought of as a System 1 process) and reason (typically thought of as a System 2 process) are deeply interconnected.¹⁰ But the distinction between automatic, gut-level responses and more deliberative responses can still serve as a useful albeit crude distinction between

⁷ There is of course a vast literature about human intuitions and objective morality, perhaps most famously developed by Henry Sidgwick. See generally, HENRY SIDGWICK, *THE METHODS OF ETHICS* (1890).

⁸ The main reason I do not do this is because – as I have argued elsewhere – many deontological arguments about human rights are in fact "utility of desert" arguments – consequentialist arguments cloaked in deontological language. See Andrew K. Woods, *Moral Judgments and International Crimes: The Disutility of Desert*, 52 VA. J. INT'L L. 633 (2012).

⁹ See Keith E. Stanovich & Robert F. West, *Individual Difference in Reasoning: Implications for the Rationality Debate?*, 23 BEHAV. & BRAIN SCI. 645 (2000) (introducing the terms "system 1" and "system 2" to describe the dual process theory of the mind).

¹⁰ See ANTONIO DAMASIO, *DESCARTES' ERROR: EMOTION, REASON, AND THE HUMAN BRAIN* (1994).

different ways of thinking about human rights. Indeed, many of the most significant findings in experimental psychology and behavioral economics depend on a rough division between thinking “fast” and thinking “slow.”¹¹

This short essay is structured as follows. First, I introduce the idea of moral heuristic failure – the idea that our gut-level intuitions are not always useful guides for complex moral problems like human rights. I then survey three particular sorts of moral heuristic failure – moral outrage, psychic numbing, and naïve realism – and examine the implications of moral heuristic failure for human rights advocacy. This is not a full accounting of the different costs and benefits associated with an intuition-based human rights advocacy, nor is it a playbook for how to reform advocacy to be less intuition-dependent. Rather, the essay merely provides reasons for skepticism about the upward bounds of relying on our moral intuitions to understand complex atrocities. I conclude with a brief discussion of potential strategies for establishing safeguards where we think moral heuristic failure is likely.

The Limits of Moral Intuitions

A growing literature suggests that our automatic moral intuitions rely on heuristics – mental shortcuts that produce moral judgments quickly – that can lead to suboptimal decisions, especially at the level of law and policy.¹² Jonathan Baron demonstrates this with the following example. Imagine that a vaccine carries some risk of death, but the risk is much lower than that posed by the disease being vaccinated against. In this scenario, many parents will opt *not* to give their child the vaccine out of a fear of actively playing a role in their child’s death – a fear that is greater than the fear that the child will die of *not* receiving the vaccine. These parents are acting irrationally against their own wishes: if the parent wants to give the child the greatest chance of survival, it ought to vaccinate. What explains this irrational behavior? The best explanation is that these parents’ judgment is being driven by an otherwise useful heuristic: “do not actively do things to your children that might harm them.”¹³ Unfortunately, in this scenario the heuristic fails because doing “nothing” (not vaccinating) is

¹¹ See generally DANIEL KAHNEMAN, *THINKING, FAST AND SLOW* (2011) (explaining behavioral economics and psychology findings that support the dual-mind theory, which broadly conceives of cognition as either automatic or deliberative, “fast” or “slow”).

¹² See Jonathan Baron, *Nonconsequentialist Decisions*, 17 *BEHAV. & BRAIN SCI.* 1, 1 (1994).

¹³ *Id.*

more dangerous than doing “something that might harm the children” (vaccinating).

Greene et al. have explicitly shown how heuristic failure maps onto the dual-process model of the mind – that is, System 1 and System 2 – using functional magnetic resonance imaging (fMRI) brain scans. They asked subjects a series of questions based on the now-famous trolley car dilemma: if a train is barreling down a track where five people are standing, and the only way to save them is to switch the train’s path to another track where only one man is standing, is it morally acceptable to flip the switch to kill the one and save the five?¹⁴ In this *trolley* dilemma, most people find it morally acceptable to flip the switch to kill the one to save the five.¹⁵ But in a variant on this problem, known as the *footbridge* dilemma,¹⁶ a fat man stands on a footbridge over the train’s track and only by pushing him onto the track, killing him, can the train be stopped and the five saved. Here, experiment subjects and professional philosophers tend to agree that it is not acceptable to kill the one to save the five.¹⁷

Why might people exhibit a consequentialist reading of the *trolley* problem – kill one to save five – but a deontological reading of the *footbridge* case? Greene and colleagues suggest that the best explanation is that people tend toward consequentialism when using System 2, their deliberative mental process, but not when they switch to automatic, heuristic-based System 1 thinking. That is, when subjects do not have strong intuitions about a particular case and must reason through their answer, as they do with the *trolley* example, they give utilitarian answers, finding it acceptable to kill one to save five. But when subjects are given scenarios that trigger strong intuitive feelings – such as the repulsion many people have to pushing someone to their death in the *footbridge* case – they gave non-utilitarian responses, finding it unacceptable to kill one to save five.¹⁸

14 See Philippa Foot, *The Problem of Abortion and the Doctrine of Double Effect*, 5 OXFORD REV. 5 (1967) (describing the *trolley* thought experiment).

15 See Joshua D. Greene, *The Secret Joke of Kant’s Soul*, 3 MORAL PSYCHOL. 35, 42 (2007) (noting that “The consensus among philosophers, as well as people who have been tested experimentally, is that it is morally acceptable to save five lives at the expense of one in this case”) (citations omitted).

16 See Judith Jarvis Thomson, *Killing, Letting Die, and the Trolley Problem*, 59 (2) THE MONIST 204 (1976).

17 See Greene, *supra* note 15, at 42 (“Here the consensus is that it is not okay to save five lives at the expense of one”) (citations omitted).

18 *Id.* at 108 (“intuitive responses [to the moral judgment in the *footbridge* case] drive people to give nonutilitarian response to moral dilemmas that have previously been categorized as ‘personal.’... these intuitive responses are emotional (i.e., constituted or driven by emotions)”).

This difference was detected by brain scans. When they gave the *trolley* and *footbridge* problems to people under fMRI scanners, Greene et al. found that different regions of the mind were activated: in the *footbridge* case, neural activity was seen in the region of the mind usually associated with emotions and social cognition (amygdale and medial surfaces of frontal and parietal lobes) while the *trolley* cases sparked neural activity in the area of the brain used for math and computation (the dorsolateral surfaces of the prefrontal cortex and parietal lobes).¹⁹ This finding was especially robust. Different experiments showed that in some fact patterns, when subjects reach consequentialist outcomes over their deontological alternative, the fMRI recorded cognitive activity in the brain regions associated with such careful deliberation. Greene finds further support for this in the *crying baby* case. In that hypothetical, subjects are told that a group of people are hiding in the basement from soldiers who are currently searching their house. They know that the soldiers will kill them if they are discovered. When an infant starts to cry, its mother is faced with a terrible choice: smother the infant and save the group, or let the baby cry, condemning the group to certain death. Is it morally acceptable for the mother to smother her infant? Greene explains:

Cases like *crying baby* are especially interesting because they allow us to directly compare the neural activity associated with characteristically consequentialist and deontological responses. According to our model, when people say “yes” to such cases (the consequentialist answer), it is because the “cognitive” cost-benefit analysis has successfully dominated the prepotent emotional response that drives people to say “no” (the deontological answer). If that is correct, then we should expect to see increased activity in the previously identified “cognitive” brain regions (the dorsolateral prefrontal cortex and inferior parietal cortex) for the trials in which people say “yes” in response to cases like *crying baby*. This is exactly what we found. In other words, people exhibit more “cognitive” activity when they give the consequentialist answer.²⁰

When subjects were prompted with questions that did not trigger System 1, they reached the consequentialist outcome; when the questions triggered strong moral intuitions, a different portion of the brain lit up and the subjects were much more likely to produce a nonconsequentialist outcome.²¹

19 See Joshua D. Greene et al., *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*, 44 NEURON 389, 389–400 (2004) (reviewing earlier literature and testing a hypothesis and finding support for the claim that moral judgments can occur through two separate processes, one deliberative and one more automatic).

20 See Greene, *supra* note 15, at 45–46.

21 Different portions of the brain register activity that roughly maps onto what Greene calls “cognitive” and “emotional” thought. See *id.* at 40.

This suggests that while our intuitions allow us to make decisions quickly, without deliberation, they can also produce suboptimal decisions. These judgment biases have been explored in a wide range of issues, from vaccines to climate change.²² But they have been largely underappreciated in the human rights context, where they can have grave consequences. The risk of heuristic failure may be particularly pronounced at the international level, where the subject matter – genocide, mass atrocity – may be difficult to comprehend, let alone resolve by intuition alone. The following section outlines three sorts of heuristic failures that are especially worrisome for the human rights regime.

Three Ways Our Intuitions Betray Us

If our moral intuitions can mislead us, this may be especially problematic for the international human rights regime. Consider three sorts of judgment biases that affect human rights advocacy: moral outrage, psychic numbing, and naïve realist thinking.

Moral Outrage

When human rights campaigns stoke moral indignation and outrage, they crowd out deliberative, consequentialist thinking. Psychologists elegantly demonstrated this with an experiment about a hypothetical set of damages awards. Subjects were given a fact pattern detailing the evidence of a tort and asked to determine damage awards – the experimental group received a fact pattern that was designed to evoke strong feelings of moral outrage while the treatment group received a similar but more neutrally worded fact pattern.²³ After reading charged fact patterns that evoked strong intuitions like moral outrage and indignation, respondents set very high damage awards – both when they were told the high damages would have no effect on the company's behavior and, separately, when they were told the high damages would have *harmful* effects, such as forcing the company to stop manufacturing its other socially beneficial products.²⁴ In this scenario, subjects who were outraged were *willing to create a net social harm* in order to satisfy their retributive impulses.

²² See Cass R. Sunstein, *Moral Heuristics*, 28 (4) BEHAV. & BRAIN SCI. 42, 531 (2005).

²³ See Jonathan Baron & Ilana Ritov, *Intuitions about Penalties and Compensation in the Context of Tort Law*, 7 J. RISK & UNCERTAINTY 17 (1993).

²⁴ *Id.*

What explains this desire to punish wrongdoers, regardless of the cost? Both psychological studies and neuroimaging confirm that the desire to punish – and satisfaction with its completion – is most closely linked to intuitive, not deliberative, thought processes.²⁵ We punish because it feels good, and because not punishing (allowing a “culture of impunity” to reign) feels bad.²⁶ Experimental economists have found significant support for this desire to punish in cooperation games – where players are willing to sacrifice their own self-interest in order to punish the violator of a group norm. (Of course, in these scenarios the punishment may very well be consequentialist – to reinforce a group norm against cooperation – but often it is not.²⁷)

This evidence may give us special reason to think that relying on moral intuitions may lead to suboptimal human rights policies. Consider, for example, the individuation of guilt in human rights trials. Human rights courts are tasked with identifying the individuals who bear the greatest responsibility for otherwise diffuse but terrible group crimes.²⁸ From the standpoint of moral intuitions, that individuation is not ideal. Studies have shown that, just as people are more sympathetic and generous toward identifiable victims, they are more punitive with identifiable wrongdoers, even when the identity of the wrongdoer is irrelevant to the wrongness of the act.²⁹ The international criminal regime, in its effort to individuate justice, is in fact creating special distortion effects – unique opportunities for moral outrage to crowd out deliberative thinking.

²⁵ See Greene, *supra* note 15, at 54 (citing neuroimaging study by Alan Sanfey, Jim Rilling, and colleagues that looked at peoples’ brains while they played the ultimatum game, and observed that when unfair offers were made, the anterior insula – the part of the brain associated with anger or disgust – lights up.).

²⁶ *Id.* at 70; psychologically speaking, we punish primarily because we find punishment satisfying (de Quervain et al., 2004) and find unpunished transgressions distinctly unsatisfying (Carlsmith et al., 2002; Kahneman et al., 1998; Sanfey et al., 2003). In other words, the emotions that drive us to punish are *blunt biological instruments*. They evolved because they drive us to punish in ways that lead to (biologically) good consequences. But, *as a by-product of their simple and efficient design*, they also lead us to punish in situations in which no (biologically) good consequences can be expected. Thus, it seems that as an evolutionary matter of fact, we have a taste for retribution, not because wrongdoers truly deserve to be punished regardless of the costs and benefits, but because retributive dispositions are an efficient way of inducing behavior that allows individuals living in social groups to more effectively spread their genes.

²⁷ For an explanation of how this works, see Herbert Gintis et al., *Strong Reciprocity and the Roots of Human Morality*, 21 (2) SOC. JUST. RES. 241 (2008).

²⁸ See Mark A. Drumbl, *Collective Violence and Individual Punishment: The Criminality of Mass Atrocity*, 99 NW. U. L. REV. 539, 581 (2005).

²⁹ See Deborah A. Small & George Loewenstein, *The Devil You Know: The Effects of Identifiability on Punishment*, 18 J. BEHAV. DEC. MAKING 311 (2005).

This is exacerbated by the nature of the crimes on trial. Crimes against humanity, “are particularly odious offences in that they constitute a serious attack on human dignity or grave humiliation or a degradation of one or more human beings.”³⁰ They constitute some of the gravest offenses known to man. As Jon Elster argues, the retributive impulse described by Baron and Ritov is all the stronger in the context of crimes against humanity:

[I]n transitional justice the pure backward-looking argument from desert often has an overwhelming appeal. It can tap into the very strong retributive emotions that are triggered by human rights violations on a scale and of an atrocity far beyond what are found under normal circumstances. It can also tap into the needs of those who did nothing, for whom retribution can be a means to redeeming themselves in their own eyes and, no doubt, in those of others.³¹

Human rights advocates rely on exactly this sort of response in developing their advocacy materials. The gruesomeness of many human rights abuses therefore presents a distinct risk of provoking intuitive outrage and retributivism that can undermine attempts at coolly and deliberately apportioning liability according to the regime’s goals.

Psychic Numbing

Just as the gravity of human rights abuses can cause our intuitions to lead us astray, so too can the scale of rights abuses produce judgment errors. For example, Paul Slovic has identified the phenomenon of “psychic numbing,” whereby people respond less powerfully to news of a tragedy as the scale and complexity of the tragedy increases:

One fundamental mechanism that may play a role in many, if not all, episodes of mass-abuse neglect involves the capacity to experience *affect*, the positive and negative feelings that combine with reasoned analysis to guide our judgments, decisions, and actions. Research shows that the statistics of mass rights violations or genocide, no matter how large the numbers, fail to convey the true meaning of such atrocities. The numbers fail to spark emotion or feeling and thus fail to motivate action. The genocide in Darfur is real, but we do not “feel” that reality. We examine below ways that we might make genocide “feel real” and motivate appropriate interventions. Ultimately, however, we conclude that we cannot only depend on our intuitive feelings about these atrocities. In addition, we must create and commit ourselves to institutional, legal, and political

30 See ANTONIO CASSESE, INTERNATIONAL CRIMINAL LAW 98 (2008).

31 Jon Elster, *Retribution*, in RETRIBUTION AND REPARATION IN THE TRANSITION TO DEMOCRACY 54 (Jon Elster ed., 2006).

responses based upon reasoned analysis of our moral obligations to stop large-scale human rights violations.³²

The smaller the suffering, the greater our intuitive response, and therefore the greater our financial contribution. An experiment by Small et al. explains this effect. Subjects on their way out of a psychological experiment were told they could donate up to \$5 of their earnings from the experiment to the NGO Save the Children. Some respondents were asked to donate in order to feed “Rokia,” a 7-year-old African girl. Others were asked to donate money to save millions of Africans from hunger. And still other respondents were asked to donate to “Rokia,” but were also told of the need to feed millions in Africa. The result was as follows: the average donation to Rokia was more than twice the donation to end hunger for millions; and the mere mention of the hunger of millions had a mitigating effect on the donations to Rokia (Figure 1).³³

This perhaps explains why charities like Save the Children focus on supporting one child, rather than on ending famine. One way to read the Rokia study is that respondents feel uneasy intervening in complicated situations of harm. We can manage the hunger of one girl, subjects might think, but the hunger of a million people may suggest a wider problem; the mere reference to this structural problem puts people off from wanting to feed Rokia. One of the biggest and fastest growing foreign aid charities in the West is the “Smile Train,” which funds surgeries for babies born with cleft palates. Their advertisements – usually full-page or half-page graphics in major Western newspapers – feature

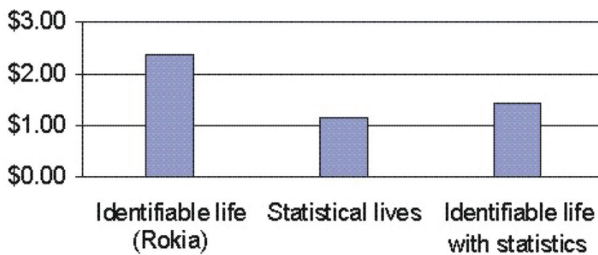


Figure 1: Average donations.

³² Ryan Goodman et al., *Psychic Numbing and Mass Atrocity*, in *THE BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY* 126 (Eldar Shafir ed., 2012).

³³ See Deborah A. Small, George Loewenstein & Paul Slovic, *Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims*, 102 *ORG. BEHAV. & HUM. DEC. PROCESSES* 143, 150 (2007).

photographs of adorable children with mangled mouths. The appeal to sympathy is overwhelming, and the “ask” is relatively simple and effective: “It costs as little as \$250 to give a desperate child not just a new smile, but a new life.”³⁴ Framing – and our intuitive responses to different frames – has a huge effect on how we respond to different human rights issues. Smile Train may be a worthy cause, but one suspects that its huge success has more to do with the gut feelings the service evokes in its audience than with a deliberative distribution of resources aimed at maximizing human rights protection.

These findings about psychic numbing are somewhat confounding. Where moral outrage causes people to care *too much* about a particular scenario, allowing their intuitions to override their deliberative thought processes, here we find that people sometimes care *too little* – that is, they are intuitively turned off by large-scale suffering, and accordingly make suboptimal decisions. Taken together, these findings suggest that there may in fact be an optimal level of concern for human rights abuses – enough concern to avoid psychic numbing, but not so much concern so as to trigger moral outrage.

Naïve Realism

One of the most striking features of the human rights movement has been the widespread appeal and world-wide adoption of rights rhetoric. Even critics of the regime acknowledge the political power of rights language. Political scientist Emilie Hafner-Burton has catalogued the amazing spread of human rights rhetoric in international relations.³⁵ Even when states do not fulfill their international commitments to human rights principles, they speak in human rights terms. Many human rights advocates hope this will lead to lasting change in the long-run. But one underappreciated cost of rights language – the language of deontological rights and wrongs – is that it encourages naïve realist thinking and therefore inhibits conflict resolution.³⁶

Naïve realists – and we are all naïve realists at one time or another – think the world is the way they perceive it. Just as importantly, they think of themselves as reasonable and rational agents who are capable of persuading other

³⁴ See, e.g. the website of the NGO, Smile Train, www.smiletrain.org.

³⁵ Emilie Hafner-Burton, *Sticks and Stones: Naming and Shaming the Human Rights Enforcement Problem*, 62 INT'L ORG. 689 (2008).

³⁶ See Lee Ross & Andrew Ward, *Naïve Realism: Implications for Social Conflict and Misunderstanding*, in VALUES and KNOWLEDGE 103 (Edward S. Reed, Elliot Turiel & Terrence Brown eds., 1996) (describing how naïve realist thinking emboldens rivals and inhibits meaningful negotiation).

reasonable people to see things their way, and that when this does not happen the others must be stupid, biased, or irrational.³⁷ The three central components can best be explained by breaking down the naïve realist thought process, which goes as follows:

1. I see things as they really are.
2. Other fair-minded people will share my views.
3. If someone doesn't share my views, and I can't just explain it to them, that shows that they are lazy, stupid, or biased (irrational).³⁸

Naïve realism offers the central insight that we misread the nature of other people's views – that we underestimate their sincerity. We have a tendency to say, “Oh, people will say anything,” especially when they are enemies or competitors, when in fact they may authentically feel a certain way. Liberals in the United States, for example, tend not to believe that people opposing health care actually think that nationalized medicine would harm Americans, but many of them do; likewise, conservatives suspect that when liberals propose nationalizing health care what they *really* want is an increase in federal power generally.³⁹ Both sides doubt the sincerity of the other; both sides look to explain away the other's statements by doubting their motivations. Members of both sides think they are obviously morally right, and the other side is obviously morally wrong.

Numerous studies confirm this. One experiment by Ross, for example, showed that both pro-Arab audiences and pro-Israeli audiences watching the same news coverage of the Israeli invasion of Lebanon in 1982 thought the coverage was biased against them. Similar results were found with Dartmouth and Princeton fans watching the same football game and judging the fairness of the game. A related and perhaps more troubling study showed that when Palestinians and Israelis were given proposals for a solution to the contentious Israeli settlements, both sides preferred the others' proposals if and only if they thought it was in fact proposed by their side.⁴⁰

The lesson from naïve realism scholarship is that we see our own way of looking at the world – from our obligations to outgroups and enemies (and even

³⁷ See, e.g. LEE ROSS AND RICHARD NISBETT, *THE PERSON AND THE SITUATION* (1991, afterword to 2011 edition) (offering a description of naïve realism thinking and calling it more fundamental than the fundamental attribution error).

³⁸ This first-person schema is taken from Ross & Ward, *Naïve Realism*.

³⁹ For an excellent demonstration of this, see Emily Pronin et al., *The Bias Blind Spot: Perceptions of Bias in Self Versus Others*, 28 *PERS. SOC. PSYCHOL. BULL.* 369 (2002).

⁴⁰ Susan Hackley, Max Bazerman, Lee Ross & Daniel L. Shapiro, *Psychological Dimensions of the Israeli Settlements Issue: Endowments and Identities*, 21 *NEGOTIATION J.* 209 (2005).

to animals, the environment, and more) to our interpretation of history and the entitlements it confers – as reasonable and rational, and we see alternative views as biased, irrational, reflective of flaws in character, and so on. It is not just that moralistic language imperils conflict resolution: It also starts and maintains conflict by emboldening political constituents with no interest in compromise. This suggests that not only are we hard-wired to make irrational decisions about moral quandaries – as when we experience moral outrage and psychic numbing – but we are hard-wired in a way that makes righteous talk costly.

Human rights advocacy encourages strong intuitive feelings about right and wrong, but the naïve realism research suggests that there may be costs to thinking about conflict in such black and white terms. Insofar as rights talk emboldens naïve realist thinking and inhibits compromise, it may imperil conflict resolution, which is a key goal of the human rights regime. Moreover, the lack of conflict resolution may lead to downstream human rights abuses.

Implications for Human Rights Advocacy

Moral intuitions are fallible, and there are at least three sorts of judgment biases (and very likely many more) that are especially costly in the context of human rights advocacy. Human rights advocacy plays a critical role in identifying human rights abuses, promoting human rights norms, and mobilizing political forces to protect human rights.⁴¹ Perhaps no better example of the powerful role that human rights advocates can have is the recent Kony 2012 campaign.⁴² The video is not a mere newsreel of statistics about the suffering in Uganda – statistics that Slovic’s research tells us are unlikely to spark an automatic sympathetic response. Rather, the video tells the story of a white American traveling in Africa where he met a child who was running from Kony’s forces, which killed the boy’s brother. The film’s narrator, Jason Russell, promises the boy that he will help bring Joseph Kony to justice, and the viewer is asked to help as well.⁴³

⁴¹ See Margaret E. Keck & Kathryn Sikkink, *Activists Beyond Borders: Advocacy Networks, in* INTERNATIONAL POLITICS 63 (1998).

⁴² See Karen DeYoung, *Joseph Kony: Barack Obama Sends Planes to Uganda as Hunt for Lord’s Resistance Army Warlord Intensifies*, THE INDEPENDENT, Mar. 24, 2014, available at <http://perma.cc/5KD8-LN6X>.

⁴³ See Kony 2012, *supra* note 4.

The results were spectacular. With over 100 million views, the video was dubbed the most viral video ever,⁴⁴ the US Senate issued a resolution condemning Kony's "crimes against humanity,"⁴⁵ and the United States sent troops and aircraft to help the African Union hunt down Kony, who faces indictment at the International Criminal Court.⁴⁶ While Kony remains at large, the campaign has been enormously successful at generating attention and mobilizing political will in the Western world to focus on human rights abuses in Uganda.

Precisely because it was so successful, the video was celebrated by many human rights activists who hoped to emulate the video's success.⁴⁷ In many respects, this is not surprising at all. The video is only the latest and most successful in a long history of human rights advocacy that seeks to use publicity to "shame and blame" human rights violators. In particular, the video tells a familiar story featuring three roles: a savage, a victim, and a savior.⁴⁸ As Makau Mutua showed long ago, this is a standard narrative for human rights advocacy stories – one that typically elevates the savior and denigrates the victim.⁴⁹ The video's success – indeed, the success of all shame-and-blame stories – is partly due to the fact that they identify a clear victim and a clear villain.⁵⁰

These are powerful attention-getting strategies. They suggest that human rights advocates have devised a way to overcome the risk of psychic numbing with stories and affect-laden images. These strategies work precisely because they counteract our biases against faraway suffering. We are intuitively drawn to sympathetic victims and simplistic stories of good versus evil, and the Kony 2012 video is an example of advocates seeking to tell a complex story in a way that makes it intuitively compelling.

But there are significant shortcomings to this approach. Not all human rights abuses can be made intuitively sympathetic to people, and this raises

44 See Nick Carbone, *Kony 2012*, Time.Com, Dec. 4, 2012, available at <http://entertainment.time.com/2012/12/04/top-10-arts-lists/slide/kony-2012/>.

45 See Press Release, *Senate Condemns Crimes of Joseph Kony and Lord's Resistance Army*, Aug. 3, 2012, available at <http://www.coons.senate.gov/newsroom/releases/release/senate-condemns-crimes-of-joseph-kony-and-lords-resistance-army>.

46 See Helene Cooper, *More U.S. Troops to Aid Uganda Search for Kony*, N.Y. TIMES A7, Mar. 23, 2014, available at <http://www.nytimes.com/2014/03/24/world/africa/obama-is-sending-more-resources-for-joseph-kony-search.html>.

47 Michael Hoffman, *Why "Kony 2012" Video Grabbed 100 Million Views Online*, CHRON. PHILANTHROPY, Mar. 15, 2012, available at <http://perma.cc/VRP6-A8LG>.

48 See Makau Mutua, *Savages, Victims, and Saviors: The Metaphor of Human Rights*, 42 HARV. INT'L. L. J. 1 201 (2001).

49 *Id.*

50 See Hoffman, *supra* note 47 ("In the video, Mr. Kony is portrayed as evil – as if he is a villain in a Batman movie. And if he is the evil villain, then you, the person fighting him, are the hero.")

the prospect of suboptimal resource allocation. In other words, policy decisions are made and resources allocated based solely on a particular issue's photogenic qualities – that is, the extent to which it lends itself to stories that provoke the moral intuitions of a public, moral intuitions that we know can bias our decision making. While the makers of the Kony 2012 video constructed a compelling narrative to tell the story of the atrocities committed by the Lord's Resistance Army, not all human rights abuses lend themselves to such compelling narratives. In fact, one of the most damning critiques of the Kony 2012 campaign was that in order to tell such a compelling narrative, the filmmakers told an incomplete story, and in several cases got the facts wrong.⁵¹ This is obviously problematic. Insofar as the story is factually incorrect, it might mobilize people to take policy steps that are suboptimal.

This is one hypothesis about the harms of relying on moral intuitions to drive advocacy: that it leads to suboptimal resource allocation. We can generate a number of other hypotheses as well. For example, it would be reasonable to think that such human rights advocacy campaigns that appeal to moral intuitions will tend to favor civil and political rights over economic and social rights. While a video can be made to tell the story of a boy who lost his brother and the effort to arrest the bad man who did it, it is much harder to tell an intuitively compelling video about, for example, the social and economic situation that enabled this killing to take place.⁵² Civil rights more readily feature identifiable victims, and identifiable victims are necessary to launch successful (intuitively compelling) advocacy campaigns. It is precisely for this reason that Kenneth Roth, the head of Human Rights Watch, has said that human rights groups should focus on civil and political rights issues rather than economic and social rights issues: because the former lend themselves to the sort of shaming and blaming that works, the kind that appeals to moral intuitions about right and wrong, good and bad.⁵³

We might also predict that human rights advocacy will be shallow rather than deep. That is, advocacy that appeals to moral intuitions is likely to focus on individual human rights violations – which are typically symptoms of larger human rights issues. For example, rather than tell the story of Uganda's shaky political leadership post-colonization, or the history of natural resource conflict in the country, or the story of Sudan's interest in Uganda's destabilization, the

51 See Ethan Zuckerman, *Unpacking Kony 2012*, Mar. 8, 2012, available at <http://www.ethan-zuckerman.com/blog/2012/03/08/unpacking-kony-2012/>.

52 This is perhaps especially true for Western audiences.

53 See Kenneth Roth, *Defending Economic, Social and Cultural Rights: Practical Issues Faced by an International Human Rights Organization*, 26 HUM. RTS. Q. 63, 67 (2004).

Kony 2012 video tells the story of a bad man and a sad boy. It is a massively successful, intuitively appealing, but ultimately shallow campaign. Indeed, the campaign repeated one of the most common wrongheaded narratives of the Ugandan armed conflict, the story of a bad-man rebel, hungry for natural resources and power, a narrative that leaves out many of the complexities of Uganda's internal power struggles.⁵⁴

Advocacy also appears to have an *ex post* bias. That is, in order to be intuitively compelling, advocacy will tend to be backward looking rather than forward looking. Advocacy that seeks to stoke the moral intuitions will point to hard realities – documented instances of rights abuses with clear identifiable victims and perpetrators – rather than vague or distant possibilities. You can get compelling footage of events passed, while you cannot film the future. As a result, it will be concerned with righting the wrongs of the past – with restorative justice – rather than preventing those harms from occurring again.

Finally, advocacy is likely to stoke moral outrage, priming people to be overly punitive and insufficiently willing to compromise, even where compromise means fewer human rights abuses.⁵⁵ Recall the Baron and Ritov studies, showing that otherwise consequentialists become deep retributivists when they are outraged.⁵⁶ Similarly, we might imagine that human rights advocacy that encourages moral outrage will lead people to seek retributive justice for human rights violators rather than a negotiated political compromise. For evidence that this is in fact what happens, consider the fact that when Kony met with negotiators in 2006, he was dissuaded from disarming because of his indictment at the International Criminal Court (ICC), an indictment advocates around the world insisted not be lifted in the name of peace.⁵⁷

These are all sensible – and testable – hypotheses that merit further study. Our moral intuitions are subject to a number of biases, like the bias for cute faces over statistics, and human rights advocacy seeks to exploit these biases for obvious reasons. This can be good: advocates can use this bias to get audiences to care about faraway harms that they might not otherwise care about. But this advocacy can only overcome certain kinds of biases, and it may in fact introduce new biases – such as the bias for civil rights over economic rights, the bias

⁵⁴ See Séverine Autesserre, *Dangerous Tales: Dominant Narratives on the Congo and Their Unintended Consequences*, 111 AFR. AFF. 202 (Spring 2012).

⁵⁵ In many respects, this is the case for the current international criminal regime, which largely prefers punishment to political compromise, and which exhibits a strong retributive impulse. See Woods, *supra* note 8.

⁵⁶ See Baron & Ritov, *supra* note 23.

⁵⁷ See Norbert Mao, *Guest Post: I've Met Joseph Kony and Kony 2012 Isn't That Bad*, Mar. 21, 2012, available at <http://perma.cc/LET9-Q6NN>.

toward punishment and away from settlement, and so on. Not only will future research need to test the foregoing hypotheses, it will need to weigh the costs from these biases – the costs of advocacy that plays to moral intuitions – against the benefits of such advocacy. Without this analysis, we cannot say whether advocacy that appeals to the moral intuitions is a net boon or a net harm to the human rights regime. In the meantime, however, we can attempt to develop strategies for minimizing the likely harms of such advocacy while preserving, if possible, the benefits of appealing to moral intuitions.

Managing Our Moral Intuitions

What should we do with the news that our moral intuitions are imperfect guides for human rights policy? We might imagine two sorts of normative responses. The first response rejects our intuitions. It says that because the moral intuitions can mislead us, they should be distrusted and discouraged at every turn. Call this the anti-intuition position. The anti-intuition position is theoretically attractive because it promises outcomes free from judgment biases. Ideally, with the right information and the right decision-making process, human rights policymakers can allocate resources efficiently and avoid the suboptimal policymaking produced by psychic numbing, moral outrage, and more. Our guts are more feeble than our minds; this claim goes, so we ought to install deliberative procedures and institutions to ensure that intuitions never get in the way of good human rights policy. This approach would likely require abandoning a rights framework, wherever rights talk produced more harm than good, and would instead call for maximizing human welfare.

The problems with this approach are too many to list. Even if better decisions were theoretically possible by taking intuition out of human rights discourse, it would be entirely unrealistic to do so. Human rights advocates and norm entrepreneurs are unlikely to be willing to give up one of their most powerful weapons for social and political mobilization. Moreover, intuitions are complex things; it is implausible that any effort to develop policy without evoking moral intuitions – on matters related to human rights, no less – would succeed. It seems politically impractical, too, to create a system that denies deeply ingrained human intuitions about grave moral questions; in a democracy at least, this approach seems unlikely to last very long. Finally, rights are often described as one-way ratchets. It seems implausible, even if it were desirable, to leave the existing human rights framework behind in favor of some alternative regime aimed at maximizing human welfare.

A second approach is perhaps a bit more realistic. It suggests that moral intuitions are powerful and useful – but they should be deployed in limited contexts and with built-in safeguards. Call this the compatibilist position. This hybrid approach would use deliberative over-rides wherever we anticipate that intuitions are likely to produce suboptimal policy. For example, the psychic numbing literature suggests that there is a great risk that international monitors and the citizens of states required to authorize an intervention will fail to recognize the cost of great losses of life as the number of losses increases.⁵⁸ A shift in default rules might be designed specifically to counteract this problem. Elsewhere my co-authors and I described how one such lock-step proposal might work to prevent the harms of psychic numbing which could inhibit political action in response to an escalating conflict risking genocide:

Say, for example, that valuations of life begin to drop off significantly after 10 deaths. At 10 deaths, a pre-authorized U.N. investigation would automatically be triggered (implementing new reporting methods, as discussed below); at 100 deaths, that investigatory body would immediately acquire certain authorities. These lock-step provisions can be justified on the grounds that any more subjective metric raises the risk of psychic numbing. If such a system could be implemented, it could limit the opportunity for genocidaire-states to stall international intervention under the guise of diplomatic debate.⁵⁹

Triggers have their own risks, of course. They discourage creative thinking and flexible problem solving by quantifying and systematizing complex human political phenomena. Treating the task of monitoring the risk of genocide as a checklist of a series of warning signs may increase the overall comprehensiveness of the monitoring effort, but it also may increase the risk that monitors will miss important events that fall outside the scope of their checklist. In this way, the checklist also risks complacency. But if the checklist is well designed, the risks imposed by this complacency are unlikely to be greater than the risk that without the triggers the regime will be either unresponsive or insufficiently responsive to mass atrocity.⁶⁰ Additionally, just because triggers are deemed a useful mechanism for *authorizing* intervention does not mean that such

⁵⁸ Psychic numbing is the name given to the phenomenon whereby people value lives less as the numbers of lives at risk goes up. See David Fetherstonhaugh, Paul Slovic, Stephen Johnson & James Friedrich, *Insensitivity to the Value of Human Life: A Study of Psychophysical Numbing*, 14 J. RISK & UNCERTAINTY 283 (1997).

⁵⁹ See Goodman et al., *supra* note 32, at 135.

⁶⁰ Consider, as an example, the checklists used to increase the safety of airplane travel worldwide. A number of recurring human errors were causing a significant number of crashes. Today, the number of crashes is reduced significantly because of the implementation of routine, checklist procedures. See MALCOLM GLADWELL, *OUTLIERS: THE STORY OF SUCCESS* (2008).

intervention will be called for in every case. The trigger would merely resolve the question of authority to intervene, without solving the questions of *who* and *how* and *when* to intervene.

This compatibilist approach is not without its problems. Critics will say it does not go far enough and may even be self-defeating. In the psychic numbing scenario, for example, activists use emotion to overcome heuristic failure. But they adopt a strategy that emphasizes relying on moral intuitions, thereby risking other heuristic failures. One way to address this concern is to draw a distinction between relying on moral intuitions in developing human rights policy – setting funding priorities, choosing who to punish and why, etc. – and relying on intuitions in the execution of a particular human rights policy. This is not a perfect distinction, of course, because emphasizing intuitions in the execution of human rights policy might lead to a wider discourse that embraces moral intuitions which could have upstream effects on the development of human rights policy. But the distinction might be useful at a high level of generality.

We can draw a further distinction between different sorts of institutions and actors in the human rights regime. For example, it might be entirely unacceptable for judges to let their intuitions guide sentencing practices in human rights cases – and we might seek to limit the expression of moral intuitions like indignation in human rights trials – but moral intuitions could still be useful in some limited forms of public education and outreach (if not full throated advocacy). Here, too, there is a risk of spillover from one institutional setting to the next. For example, if a judge makes a decision with the aid of careful deliberation and safeguards to limit the risks of moral outrage, but the decision is promoted by human rights advocates as proof of the intuition that the defendant violated moral norms, it can prime the audience to expect policy that matches their moral intuitions.

Conclusion

Where does this leave us? At one extreme, we can reject any proposal that asks us to embrace moral intuitions as perfect guides to solve complex moral questions; and it should be unpersuasive in debates over human rights policy to say that one policy is “intuitively better” than another. At the other extreme, we can reject as impracticable any proposal that requires total suppression of moral intuitions. That leaves us somewhere in the middle. A better approach will allow advocates to make use of the moral emotions, while also taking steps to ensure

deliberative overrides and safeguards exist to correct for heuristic failure where that failure would produce bad human rights policy. Rather than call for “sympathetic education,” – Rorty’s term for getting people to feel sympathy for the suffering of others – policymakers might aim for “moral intuition education,” which would have two aims. First, this sort of education would encourage people to develop strong moral intuitions that align with optimal human rights policy. For example, this might mean dialing down the retributive impulse and strengthening our intuitions about social and economic welfare. Second, this sort of education would ensure that voters, consumers, and policymakers are aware of the limits of their intuitions. Rather than blindly asking people to follow their intuitions, we might hope that human rights policymakers will ask people to evaluate and check their intuitions against their better and more carefully considered judgment.

Acknowledgments: The author thanks the conference participants for valuable feedback on the essay, especially Ram Rivlin for his thoughtful comments.

Note: This essay was presented at the “Human Rights and the Human Minds” conference at the College of Law and Business in Ramat Gan, Israel. The summary of the mind sciences literature on moral decision is adapted from my earlier piece, *Moral Judgments and International Crimes: The Disutility of Desert*, 52 VA. J. INT’L L. 633 (2012). All errors are mine.