Accepted version

Published in *Empirically Engaged Evolutionary Ethics*,

edited by Johan de Smedt and Helen De Cruz (Springer Synthese Library, 2021)

https://doi.org/10.1007/978-3-030-68802-8_8

The Cultural Evolution of Extended Benevolence

1 Extended Benevolence in Darwin's *Descent of Man*

In *The Descent of Man* (1879)[1], Charles Darwin theorizes the history of "the moral sense" and anticipates its future. He writes:

> "As man advances in civilization, and small tribes are united into larger communities, the simplest reason would tell each individual that he ought to extend his social instincts and sympathies to all the members of the same nation, though personally unknown to him. The point being once reached, there is only an artificial barrier to prevent his sympathies extending to the men of all nations and races." (Darwin 1879: 147)

Darwin ups the ante a few lines down the page. He suggests that human sympathies can and will extend beyond our own species:

> "Sympathy beyond the confines of man, that is, humanity to the lower animals, seems to be one of the latest moral acquisitions…This virtue, one of the noblest with which man is endowed, seems to arise incidentally from our sympathies becoming

[1] Full title: *The Descent of Man, and Selection in Relation to Sex*.

more tender and more widely diffused, until they are extended to all sentient beings." (Ibid.)

Darwin's comments were mostly speculative. Yet as I shall argue, his account of the evolution of the "moral sense" has turned out to be remarkably prescient. My effort here will be to update Darwin's outline of the emergence of a human sympathetic capacity that extends to all nations, races, and even to all sentient beings. I shall call this form of sympathy *extended benevolence*. In the following, I cite *cultural* evolutionary mechanisms to explain the emergence and spread of extended benevolence. I will discuss ways that extended benevolence could arise through forces of cultural evolution known as *adapted transmission biases*.

2 Darwin on the "Moral Sense"

In trying to answer the question of how extended benevolence might have evolved, Darwin offers a worthwhile starting point. He famously suggests in the *Descent* that the human "moral sense," as he called it, evolved via natural selection, emotionality, habit, community rules of conduct, instruction, and reason. Let us review the details of Darwin's evolutionary account of the moral sense. In so doing, we will see how Darwin grasped many of the most crucial insights that allow a cultural evolutionary theory to meet the challenge of explaining extended benevolence.

Darwin describes the "moral sense" in the opening paragraph of chapter 4 of the *Descent*:

"…the moral sense or conscience…is summed up in that short but imperious word *ought*, so full of high significance. It is the most noble of all the attributes of man,

> leading him without a moment's hesitation to risk his life for that of a fellow-creature; or after due deliberation, impelled simply by the deep feeling of right or duty, to sacrifice it in some great cause." (Darwin 1879: 120)

In this passage, Darwin uses the term "moral sense" interchangeably with "conscience." He links the moral sense to normative attitudes expressed (in English) through the word "ought," and he cites the moral sense as a motivation for altruistic behavior.

Darwin outlines four stages in the evolution of the moral sense, with natural selection most prominently driving the first stage. In the first stage, an animal acquires "social instincts," which leads it "to take pleasure in the society of its fellows, to feel a certain amount of sympathy with them, and to perform various services for them" (Darwin 1879: 121). In Darwin's view, sympathy is a chief motivation behind the altruistic "services" that animals perform for others. Various animals, including birds, dogs, monkeys, and humans, feel love and sympathy for others. In particular, they feel sympathetic pain when in the presence of the pain of another individual. For animals in the first stage, sympathy does not extend to all members of the same species, but only to others in the same "association" (Ibid.). Darwin tries to explain the evolution of the social instincts through an early appeal to group selection. Sympathy, he suggests, likely proliferated due to natural selection between different "communities" of the same species, since "those communities, which included the greatest number of the most sympathetic members, would flourish best, and rear the greatest number of offspring" (Darwin 1879: 130).

In the second stage of the evolution of the moral sense, some animals gain the ability to remember their past actions (Darwin 1879: 121). With this ability, animals come to remember past moments in which they experienced a conflict between their social instincts

and their "instincts of self-preservation," such as instincts to pursue food and sex (Darwin 1879: 136). Darwin mentions that human beings are unique in feeling *regret* and *shame* brought on by nagging memories of past instances when one acted against one's social instincts (Darwin 1879: 135 – 136, 138). Shame is a painful feeling prompted by the experience and memory of others' disapproval of one's own behavior. Such disapproval tends to be elicited by behavior that serves one's own interests at the expense of others (Darwin 1879: 136, 138). As a result, human beings have some inclination not to repeat past actions in which they satisfied their self-preserving instincts rather than their social instincts. This inclination, Darwin adds, is conscience: "for conscience looks backwards, and serves as a guide for the future" (Darwin 1879: 138).

Darwin explains the second stage with reference to at least two mechanisms—habit and natural selection. He maintains that conscience can be strengthened through habit into a capacity for "self-command" (Darwin 1879: 139 - 140). An individual possessing self-command would be accustomed to acting in accordance with his or her social instincts "instantly," and "without struggle" (Darwin 1879: 139). Apart from being acquired through habit, Darwin emphasizes that self-command may also be inherited (Darwin 1879: 140). Darwin's rationale for this claim appears to be that conscience depends on shame, and shame in turn depends on sympathy. Sympathy, we saw, is theorized by Darwin to be a product of natural selection on groups (Darwin 1879: 136, 138).

The third stage in Darwin's account of the evolution of the moral sense follows the advent of language. It occurs when "the common opinion of how each member ought to act for the public good, would naturally become in a paramount degree the guide to action" (Darwin 1879: 122). Darwin observes that the "imperious word 'ought'" implies an

awareness of a rule of conduct, the violation of which will be met with social disapproval (Darwin 1879: 140). To avoid the shame elicited by this disapproval, humans will tend to comply with rules of conduct formulated and enforced by common opinion. The common opinion is expressed through language—at first in speech, later in writing (Darwin 1879: 146). Accordingly, people can learn about rules of conduct through instruction: they can listen to or read the words of other people who explicitly articulate the rules. Further, rules of conduct can be learned by example: people can observe which specific behaviors performed by someone elicit approval and disapproval among others in the community (Darwin 1879: 146, 149, 157).

In the fourth and last stage of Darwin's evolutionary history of the moral sense, "reason" brings about extended benevolence (Darwin 1879: 141 – 143). Darwin proposes that "as small tribes are united into larger communities, the *simplest reason* would tell each individual that he ought to extend his instincts and sympathies" to unfamiliar strangers, and when this point is reached, there is only "an artificial barrier" to prevent human sympathies from reaching "beyond the confines of man…until they are extended to all sentient beings" (Darwin 1879: 147, emphasis added). Darwin calls "[s]ympathy beyond the confines of man" a virtue; indeed, he says it is "one of the noblest with which man is endowed" (Ibid., cf. Darwin 1879: 151). Additionally, as soon as extended benevolence "is honoured and practiced by some few men, it spreads through instruction and example to the young, and eventually becomes incorporated in public opinion" (Darwin 1879: 147).

Darwin's account of the evolution of the moral sense ends with the emergence of extended benevolence. As we advance in "intellectual power," as we become more adept at tracing the remote consequences of our actions, as we sympathize more with others, and as

we learn more "from habit, following on beneficial experience, instruction and example," our sympathies ultimately become "more tender and widely diffused, extending to men of all races, to the imbecile, maimed, and other useless members of society, and finally to the lower animals" (Darwin 1879: 149).

In this speculative history, Darwin characterizes the moral sense as an assemblage of components. It consists of (1) a capacity to make normative judgments (i.e., *ought* judgments); (2) a set of "social instincts," particularly sympathy, which can motivate altruistic behavior; (3) a disposition to obey rules prescribed by community opinion; (4) a rational capacity to anticipate how the consequences of actions and practices affect the welfare of one's social group; and (5) a rational capacity to extend one's sympathies to unfamiliar others in spite of "artificial" or arbitrary differences one has with them. As I will discuss in sections 4 and 5 below, Darwin's analysis of the components of the moral sense has fared remarkably well in the light of contemporary research.

3 Extended Benevolence: Behaviors, Institutions, and Attitudes

Before we consider how extended benevolence might have evolved, we could do with more clarity on what it is. Extendedly benevolent *behaviors and institutions* treat the good of all human beings, or even all sentient beings, as having some degree of moral significance. This section highlights several human behaviors and institutions which may be described as extendedly benevolent.

Political regimes that practice equal respect for the legal human rights of individuals are one type of extendedly benevolent institution. *Legal human rights* are legal rights protected for all human beings within a jurisdiction. Legal human rights are *equally respected* when a governmental body protects them to the same degree for all human

beings in the relevant jurisdiction. Equal respect for legal human rights is an instance of extended benevolence, since all human rights-bearers are treated as having equal moral standing by the state.

Particularly since the 20th century, there has been substantial progress in protections of human rights. Political scientist Christopher Fariss has shown that since the early 1980s, human rights to *physical integrity* have been increasingly respected by governments throughout the world (Fariss 2014). Physical integrity rights include human rights not to be subjected to political kidnapping, arbitrary imprisonment, battery, torture, execution, politicide, and genocide.

Another extendedly benevolent institution is democracy. Democracies exhibit extended benevolence to the degree that every adult citizen is able to influence political outcomes. Democracies distribute political power more equally among adult citizens than other systems of government do. More than other political regimes, democracies ensure free and fair elections, the freedom to organize political movements, the freedom to express political opinions, an independent and impartial judiciary, and most of all, the power to vote. In *Freedom in the World*, an annual report published by Freedom House, countries are ranked according to these characteristics and other measures of political equality. The report shows that democracies consistently outperform other political systems on these measures (see Freedom House 2020).[2]

[2] To be sure, existing democracies do not institute perfect political equality. All too often, the wealthy have disproportionate power to influence politicians' decisions. Many democracies disenfranchise adult citizens convicted of a criminal offense. And, democracies typically prohibit minors and non-citizen adults from voting. The point is only that political equality among adult citizens is achieved to a greater extent in democracies than in nondemocracies.

At the turn of the 20th century, there were more autocracies than democracies in the world. By the turn of the 21st century, democracies outnumbered autocracies (Roser 2020). Political scientist Daniel Treisman examined a composite of four authoritative measures[3] used to classify a country's political system for every year between 1800 and 2016 (Treisman 2018). The proportion of the world's democracies underwent volatile growth through this period, with both steep rises and precipitous falls in the 20th century. Despite those ups and downs, Treisman observed a clear overall pattern: a rising tide of democracy in which the global proportion of democracies reached "at or near an all-time high" of around 59% by 2016.

Scholars are debating whether democracy is on the verge of decline.[4] Some data are indeed troubling. The 2020 report of the Varieties of Democracy (V-Dem) Project found that, for the first time since 2001, democracies no longer made up a majority of countries (Lührmann et al. 2020). In 2019, 48% of countries in the world were democracies, and democracies were home to only 46% of the world's population. I am not in a position to speak to whether this is only a short-lived dip or a sustained backsliding of democracy.[5] Instead, my concern will be to explain, from a cultural evolutionary perspective, how the form of extended benevolence manifested by democracy came to be as widespread as it is now.

Extended benevolence can also be observed in the treatment of non-human animals. Since the 19th century, there has been a steady rise of laws prohibiting the exploitation of animals in dozens of countries (Waldau 2011: 106 – 108). For instance, in 2005 Australia

[3] Polity, Freedom House, the Boix-Miller-Rosato code, and V-DEM.
[4] For a discussion of the debate surrounding the "new pessimism" about democracy, see Welzel et al. (2019).
[5] The 2020 V-Dem Report also notes that the recent decline in democracy has mobilized resistance: pro-democracy protests reached an all-time high in 2019.

banned any experiment on nonhuman apes that is not in the interest of the animal itself. In 2000, the High Court of Kerala, India, ruled that under Article 21 of the Indian Constitution, circus animals were "beings entitled to a dignified existence" (Waldau 2011: 108). In 2015, an Argentinian court declared an orangutan named Sandra a "nonhuman being" entitled to basic rights to life, freedom, and protection from harm (Giménez 2015). Even though these events don't quite amount to treating animals and humans *equally*, they nonetheless display a form of extended benevolence that treats animals as beings worthy of protection and concern.

Many people possess *normative attitudes* that may be described as extendedly benevolent. There is, moreover, compelling evidence that these attitudes play a causal role in bringing about extendedly benevolent behaviors and institutions.

Data from the United States suggest that people in the animal rights movement were driven by normative commitments to achieve legal protections for animals against suffering, death, and exploitation at human hands (Waldau 2011). The sociologist James M. Jasper (1997) found that "moral shocks" play a key role in recruiting people to join animal rights protests. Moral shocks are events which raise "such a sense of outrage in people that they become inclined toward political action" (Jasper 1997: 106). Jasper and his team collected questionnaires from over 300 protestors who attended two animal rights demonstrations in 1988. When asked to rate the importance of a list of factors that drew them into the animal rights movement, 72% of the respondents rated "Things you have read" as *very important* (Jasper 1997: 175-176). Jasper observes that "[p]eople were recruited by an animal rights literature filled with powerful images designed to shock," such as cats with electrodes planted in their heads and white rabbits with puss-filled eyes from

cosmetics testing (Ibid.). For instance, Jasper's team interviewed an animal rights activist who testified to being deeply affected by the texts and images documenting experiments done on animals. "[T]hat's gotta stop," he vowed (Jasper 1997: 176). Other studies suggest that vegetarianism and veganism can be substantially attributed to people's normative attitudes of concern for the rights or welfare of animals. In a 2002 telephone survey of 400 vegetarians in the U.S., 10% cited *animal rights* as their reason for being vegetarian. According to a 2012 survey of 145 vegetarians (aged 18 - 25) in the U.S., 67% of the respondents cited *ethics* as their reason (Cooney 2018: loc. 1233).

In addition, extendedly benevolent normative attitudes have been a powerful contributing cause of the institutionalization of democracy and human rights. This much has been shown by sociologists Christian Welzel, Ronald Inglehart, and their collaborators (see Inglehart & Norris 2003; Inglehart & Welzel 2005; Welzel 2013; Inglehart 2018). Welzel, in particular, found strong correlations between a cluster of normative attitudes that he calls *emancipative values*, on the one hand, and human rights and democracy, on the other. Generally, a person who accepts emancipative values will tend to emphasize the importance of freedom of choice and equality of opportunity for all persons (Welzel 2013: loc. 4818 – 4831). To measure the acceptance of emancipative values in a given country's population, Welzel relies on the World Values Survey (WVS).[6] Welzel uses the following items on the WVS as indicators of whether the respondents hold emancipative values (Welzel 2013: loc. 1989):

[6] The goal of the WVS is to collect data on the beliefs and attitudes of people around the world. Since it was launched in 1981, the WVS has polled 150,000 people in 100 countries containing 90 percent of the world's population (Welzel 2013: 58; Inglehart 2018: 5). It collects statistically representative samples of all residents living in every country surveyed.

- WVS respondents are taken to value *freedom of choice* if they *agree* that independence and imagination, but *not* obedience, are desirable qualities in children, or if they express tolerance of abortion, divorce, and homosexuality.
- Respondents are taken to value *equality of opportunity* if they express *disagreement* with the idea that education is more important for a boy than for a girl; or they *disagree* that men should have priority over women to get a job when jobs are scarce; or they *disagree* that men make better political leaders than women.
- WVS respondents' normative attitudes are viewed as valuing *equality of opportunity* if they assign a high priority to protecting freedom of speech, or to giving people more say in important government decisions, or to giving people more say about how things are done at their jobs and in their communities.

Welzel argues that changes in the popular acceptance of emancipative values are powerful causes of legal human rights and democracy. To support this thesis, Welzel cites strong and statistically significant correlations between his measure of emancipative values, on the one hand, and measures of institutional protections of human rights and democracy, on the other. To measure human rights and democracy, Welzel relies on a *citizen rights index* and a *women's rights index* (Welzel 2013: Appendix 8, 9). He examines approximately 50 countries which were surveyed at least twice by the World Values Survey over a period of at least 10 years. Ultimately, Welzel discovers a strong, positive, and significant association between (1) changes in the proportion of people in a country who accept emancipative values over a time period of at least a decade, and (2) the country's scores on citizen rights and women's rights measured at the end of that decade-long period (Welzel

2013: loc. 7197 – 7395). Welzel's work reveals that extendedly benevolent institutions such as legal human rights and democracy owe their existence in large part to certain normative attitudes—namely, emancipative values. Emancipative values themselves are properly regarded as extendedly benevolent attitudes, given their emphasis on equality of opportunity and freedom of choice for all persons (Welzel 2013: loc. 5020, 5227).

Summing up, behaviors and institutions that can be described as extendedly benevolent are widespread. The cases in point were legal human rights, democracy, and the protection of animal rights and welfare. These behaviors and institutions can be substantially explained by extendedly benevolent normative attitudes, such as emancipative values or a belief in the moral standing of animals. Altogether, these phenomena are emblematic of the capacity for human social instincts and sympathies to extend, as Darwin had predicted, to "all nations and races," and even "beyond the confines of man."[7]

4 Extended Benevolence Evolving

At this point I begin to advance a cultural evolutionary explanation for extended benevolence. In their book *Not by Genes Alone* (2005), evolutionary anthropologists Robert Boyd and Peter J. Richerson set out the nuts and bolts of their influential theory of cultural evolution. *Culture*, as they define it, is "*information capable of affecting individuals' behavior that they acquire from other members of their species through teaching, imitation, and other forms of social transmission*" (Richerson & Boyd 2005: 5). *Cultural variants* are elements of cultural information; they include ideas, knowledge, beliefs, values, skills, and

[7] I do not claim that extended benevolence will remain as widespread as it is forever. The recent rise of nationalist-populism in the West might augur the demise of extended benevolence. It is too soon to tell. My aim is merely to establish that the existence of extended benevolence can be explained from a Darwinian—i.e., cultural evolutionary—perspective.

attitudes (Richerson & Boyd 2005: 5 - 6, 63). Different populations of people exhibit differences in language, custom, moral belief systems, technologies, and art because they adopt different cultural variants (Richerson & Boyd 2005: 6).

Cultural evolution, as Boyd and Richerson define it, is change in the relative frequencies of different cultural variants within a population over time (Richerson & Boyd 2005: 59 – 60). Boyd and Richerson identify several causes, or *forces*, of cultural evolution (Richerson & Boyd 2005: 68 – 69). Among those cultural evolutionary forces are *transmission biases*, which are features of human psychology that make people more likely to adopt some cultural variants than others (Richerson & Boyd 2005: 68). Boyd and Richerson distinguish between three transmission biases (Richerson & Boyd 2005: 69). First, there is *content-based bias*, which operates when individuals are more likely to learn or remember some cultural variants than others due to their content (Ibid.). Boyd and Richerson add that "[c]ontent-based bias can result from calculation of costs and benefits associated with alternative variants" (Richerson & Boyd 2005: 69). Second, there is *frequency-based bias*, in which individuals choose to adopt a cultural variant based on how frequent it is in the surrounding community (Ibid.). And third, there is *model-based bias*, in which individuals choose to adopt a cultural variant as a result of observing the attributes of other people who have adopted the variant. A model-based bias known as *prestige bias* may motivate an individual to adopt a cultural variant merely because the most prestigious, high-status individuals in the relevant society have adopted it. Alternatively, a model-based bias called *success bias* may guide an individual to adopt a cultural variant for the reason that others who've adopted it are relatively successful in some way—more wealthy, healthy, happy, etc.

Boyd and Richerson's framework can be used to explain the cultural evolution of extendedly benevolent institutions. Human rights institutions and animal welfare protections, in particular, can be regarded as assemblages of cultural variants that have been increasingly adopted in many societies. It is uncontroversial that these social phenomena have spread through social transmission. Moreover, there are quantitative measures of both human rights and animal protections. Philosopher Jonathan Birch stresses that cultural variants need to be measured quantitatively for researchers to do the essential work of making mathematical models of cultural evolution (Birch 2017: 196). Fariss developed a Human Rights Protection Score on the basis of several other indices (Fariss 2014). Also, the Animal Protection Index, published by the organization World Animal Protection, scores countries according to their demonstrated commitment to promote animal welfare through policy and legislation.

Transmission Biases and Human Rights

Social scientists have pinpointed content-based and frequency-based transmission biases that are causing humans rights to proliferate worldwide. Consider, for instance, the work of political scientist Brian Greenhill (2015). Greenhill finds that "over time, states adopt similar human rights practices to those of the other states with whom they share IGO memberships" (Greenhill 2015: 14). *IGOs* are inter-governmental organizations whose members are representatives of sovereign countries (Greenhill 2015: 5 – 6, 60). Well-known IGOs include the United Nations and the European Union, while others include the Gulf Cooperation Council (GCC), the West African Health Organization, and the International Coral Reef Initiative. Greenhill relies on the Correlates of War 2 International Governmental Organizations Data Set, which provides data on 495 IGOs between the years 1815 to 2005

(Greenhill 2015: 60). Greenhill's analysis demonstrates that among IGOs whose cultures strongly expect their member states to respect human rights norms, the human rights records of member states tend to improve within the first few years of joining the IGO (Greenhill 2015: ch. 3). This occurs because diplomats or policymakers from member countries who operate in the IGOs are influenced by their exposure to the human rights cultures of the organizations. They then go on to influence policymaking in their home countries (Greenhill 2015: 46 - 51).

By Greenhill's account, *acculturation* is one mechanism through which co-members of IGOs become more similar in their human rights adherence. In an acculturation process, "an actor changes his or her beliefs and behaviors in order to conform to the norms of a new social environment" (Greenhill 2015: 44 – 45). Acculturation is different from *material inducement*, in which an agent changes behavior to comply with someone else's demands so as to reap material awards or avoid material sanctions. It's also distinct from *persuasion*, in which an agent undergoes a change in beliefs after thoughtfully deliberating over information conveyed by others (Greenhill 2015: 39, 43). Acculturation is driven by two of the transmission biases emphasized by Boyd and Richerson: frequency-based bias and model-based bias. Meanwhile, persuasion would qualify as a content-based bias.

Greenhill shows that a frequency-based bias is at work when countries adopt the same human rights practices as their IGO partners. He measures the human rights performance of countries by means of the Physical Integrity Rights (PIR) index. The PIR gives states an annual score which represents the frequency of human rights violations—namely, torture, political imprisonment, extrajudicial killing, and disappearances—that take place within each state in a given year (Greenhill 2015: 62). Greenhill then tests for an association

between states' PIR scores and their *IGO context*. IGO context is another measure of Greenhill's design which is roughly a weighted average of the PIR scores of all the IGO partners of a focal state (cf. Greenhill 2015: 64 – 70). For the 154 countries he surveys (covering the period from 1982 to 2006), Greenhill finds a positive, statistically significant, and relatively robust correlation between states' PIR scores and their IGO context from the previous year. When controlling for other factors such as GDP per capita, democracy, and trade dependence, if a focal state's IGO partners had lower PIR scores in a given year, the state tended to have lower PIR scores the following year, and if a focal state's IGO partners had higher PIR scores in a given year, the state tended to have higher PIR scores the following year (Greenhill 2015: 72 – 76). Greenhill's results are indicative of a frequency-based bias which causes a country to reduce the number of its human rights violations after participating in an IGO network composed largely of partner countries that have low numbers of violations.

In the work of another political scientist, we see how content-based transmission biases help to explain why countries comply with international human rights treaties. In her book *Mobilizing for Human Rights* (2009), Beth A. Simmons argues that a country's ratification of a human rights treaty increases the expected utility of *mobilizing for human rights*—i.e., of joining a mass movement to demand the fulfillment of the rights promised in the human rights treaty (Simmons 2009: 138 – 153).[8] Simmons argues, first, that a country's ratification of human rights treaties increases the probability that mobilizing for human

[8] International human rights treaties are international legal agreements in which the governments ratifying the agreement commit to respecting the human rights of their people. They include the International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), the International Covenant on Civil and Political Rights (ICCPR), the International Covenant on Economic, Social and Cultural Rights (ICESCR), the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW), the Convention against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (CAT), and the Convention on the Rights of the Child (CRC).

rights will successfully strengthen a country's protections of human rights. Ratification of human rights treaties increases the probability of successful mobilization by attracting more allies to the country's human rights movement, by enhancing the perceived legitimacy of the human rights movement, and by expanding the legal and political strategies that the human rights movement can employ to achieve broader human rights protections (Simmons 2009: 144 – 147).

Second, Simmons argues that a country's ratification of human rights treaties increases the utility, or value, of human rights protections for the people within the country. Legal frameworks, including treaties, perform an "educative role" by changing individuals' perceptions of their own identities and interests (Simmons 2009: 140). Simmons cites the work of social anthropologist Sally Engle Merry, whose research describes how individuals can incorporate transnational human rights into their already-held values and perspectives (Merry 2006). When people understand and reflect on the content of a human rights treaty, they may come to think of themselves as being entitled to the rights codified in the treaty (Simmons 2009: 141-143). As a result, exposure to the content of a human rights treaty may increase the utility of human rights protections for people who come to perceive themselves as rights-bearers.

The *expected utility* of human rights mobilization is the product of the utility of human rights protections and the probability of successfully realizing human rights protections. Simmons's theory predicts that actual compliance with ratified human rights treaties will be greater in countries where the expected utility of human rights mobilization is higher. When this happens, there will be more mobilization, and thus more political

pressure placed by citizens on governments to comply with the human rights treaties they have ratified.

Simmons posits that the expected utility of human rights mobilization is highest in countries that are transitioning from an autocratic political system to a partially democratic one (Simmons 2009: 150-153). Many of these partially democratic transitional regimes (PDTRs) are just beginning to emerge from a condition where there had been extensive political repression. Hence there is more demand—i.e., high utility—for human rights protections within these regimes. But because PDTRs are newly and partially democratic, they also have institutional mechanisms—such as the ballot, a free press, and an independent judiciary—that incentivize governments to be responsive to citizens' demands. Hence there is a reasonably high probability that mobilizing for human rights will lead to real improvements in the human rights performance of PDTRs.

The predictions of Simmons's expected utility hypothesis are borne out by the data: in PDTRs where the expected utility of human rights mobilization is hypothesized to be highest, ratification of human rights treaties is most strongly associated with improved human rights protections. Simmons compares the human rights performance of stable autocracies, stable democracies, and PDTRs. She finds that PDTRs that ratified the Convention Against Torture (CAT) are much more likely to reduce their incidence of torture than PDTRs that did not ratify. As Simmons observes, "[r]atification of the CAT is associated with almost a 40 percent increase in the likelihood that a country will improve by one category on the torture scale" (Simmons 2009: 276). Also, among PDTRs, ratification of the International Covenant on Civil and Political Rights (ICCPR) is associated with an 11 percent improvement in a country's average religious freedom score (Simmons 2009: 176). And

furthermore, ratification of the ICCPR by PDTRs is associated with fairer domestic trials for up to five years (Simmons 2009: 185).

The mechanism posited by Simmons's expected utility hypothesis is a content-based bias, since it involves agents selecting novel cultural variants as a result of a cost-benefit calculation. Citizens in partially democratic transitional regimes value the rights codified in human rights treaties, and when assessing whether to mobilize collectively for stronger human rights protections, they deem themselves to have a good enough chance of success under the political circumstances to make mobilizing a better prospect than not mobilizing. Here, stronger human rights protections and the status quo can be considered alternative cultural variants. Citizens in the relevant regimes—the PDTRs—assess the relative costs and benefits of pursuing novel cultural variants versus staying with the status quo, and they opt in favor of the former.

Transmission Biases and Animal Welfare

We can look to other social science research for insight into how transmission biases explain the cultural evolution of extendedly benevolent institutions that protect nonhuman animals. Plausibly, norms that prohibit cruelty to animals were favored in cultural evolution by certain content-based transmission biases. Some of these content biases were likely rooted in people's capacities for sympathy and perspective-taking. In *Sentimental Rules* (2004), philosopher Shaun Nichols traces a process of growing public opposition to animal cruelty in Western European societies during the nineteenth century. By the late nineteenth century, animal protection laws prohibiting animal blood sports and other abuses became commonplace throughout the United Kingdom and Europe (Nichols 2004: loc. 1879 - 1883). But why did laws against animal cruelty become so popular in that particular moment?

Strikingly, Nichols notes that anti-cruelty laws were championed by pet owners "who seem to have developed heightened sensitivity to the plight of animals" (Nichols 2004: loc. 1887). This "heightened sensitivity" should not be surprising, considering that pet owners are well-practiced in taking the perspective of, and sympathizing with, their pets. Indeed, a questionnaire study of Scottish primary school children found that an emotional attachment to pets predicts a concern for the welfare of *all* animals—not just pet animals, but also farm animals and wild animals (Hawkins et al. 2017).

Of course, since pet ownership predates the 19th century, it cannot be the whole explanation for extensions in benevolence toward animals that began in that century. James M. Jasper and Dorothy Nelkin discuss other factors in their book *The Animal Rights Crusade* (1992) (see also Jasper 1997: 162 – 165). They suggest that European and American attitudes toward nonhuman animals have changed gradually since the 16th century. In this period, a bourgeoisie inhabiting industrialized towns and cities grew, while the share of people practicing agriculture shrank. An agricultural way of life fostered the perception that animals were mere resources to be exploited. Sure enough, people in agricultural societies did have pets and formed emotional attachments to them, but these affectionate bonds coexisted with the economic use and consumption of livestock. With urbanization and industrialization, a declining fraction of the population directly exploited animals as a resource. As Jasper notes, people "hunted less, had fewer fields to plow, and raised fewer animals to slaughter" (Jasper 1997: 163). Instead, more and more people incorporated pets into their tight emotional circles, cherishing them as beloved companions. Additionally, advances in science in the 18th and 19th centuries, such as Darwin's theory of common descent, made the similarities between animals and humans more salient in people's minds

(Ibid.). These sociological developments allowed for feelings of sympathy and affection to gradually displace a callous, exploitative approach to animals.

Nichols cites the spread of norms against animal cruelty as evidence in favor of his *affective resonance* theory of cultural evolution. According to this theory, "norms prohibiting actions that are likely to elicit negative affect, 'affect-backed norms,' will have an advantage in cultural evolution" (Nichols 2004: loc. 2020). In other words, people have a defeasible preference to adopt and follow norms that do not elicit negative affect. Nichols suggests that norms protecting animal welfare are "affect-backed" in the sense that they spare people from experiencing aversive emotions caused by an awareness of the suffering of animals. If so, we should expect to see these norms become more widespread as people increasingly sympathize with and take the perspective of animals. This is indeed what took place in Europe between the 16th and 19th centuries.

5 The Moral Sense as an Assemblage of Adapted Transmission Biases

The research reviewed thus far illustrates Boyd and Richerson's cultural evolutionary framework. That framework offers a powerful explanation for the diffusion of extendedly benevolent institutions and behaviors due to frequency-based and content-based transmission biases. In this section, we'll see that the transmission biases driving the cultural evolution of extended benevolence can themselves be explained in the light of cultural evolutionary theory. As the philosopher Tim Lewens (2015) observes, Boyd and Richerson theorize that some transmission biases are *adapted* in the sense that they evolved because they enhanced the reproductive success of our hominin ancestors (Lewens 2015: 17; Richerson & Boyd 2005: 7-8, 71, 196 – 197). Following Boyd and Richerson, I maintain that the moral sense is an assemblage of such *adapted* transmission biases.

As we saw in section 2, Darwin thought of the human moral sense as a complex cognitive-motivational system made up of five components. The psychologist Michael Tomasello (2016, 2018a, 2018b, 2020) has an account of how the "sense of moral obligation" might have evolved. This sense of obligation is a foundation for at least the first three of the components that Darwin ascribes to the moral sense: namely, sympathy, our disposition to follow community rules, and our capacity to express normative "ought" judgments about people's behavior. Hereafter, community rules will be denoted as "norms."

Tomasello theorizes that the sense of obligation is a motivation that evolved in *Homo sapiens* psychology because it facilitated large-scale cooperation among individuals who were not genetically related to one another. If Tomasello's account is correct, the sense of obligation is an adapted transmission bias that inclines people to carry out cooperative behaviors and preserve cooperative arrangements.

In Tomasello's theory, the sense of moral obligation evolved in two major transitions. First, a little less than 2 million years ago, a global cooling and drying period caused land-dwelling monkeys to migrate into the habitats of early hominins of the *Homo* genus. The resulting competition over food forced some early *Homo* to scavenge carcasses killed by other animals. But eventually, early *Homo* populations—perhaps *Homo heidelbergensis* some 400,000 years ago—began to forage for food cooperatively in face-to-face, dyadic interactions (Tomasello 2018a: 662, 664; Tomasello 2018b). Cooperation was so essential to survival that natural and social selection pressures favored individuals who possessed psychological dispositions to cooperate. Not having collaborators was a sure way to die. On the other hand, individuals who could prove themselves to be reliable

cooperation partners were selected as collaborators and mates, and this brought significant fitness advantages.

The moral psychology that emerged from this first transition was a "*second-personal morality*," which Tomasello defines as "the tendency to relate to others with a sense of respect and fairness based on a genuine assessment of both self and others as equally deserving partners in a collaborative enterprise" (Tomasello 2018b; see also Tomasello 2018a: 665). Second-personal morality includes a capacity to feel sympathy for someone who is or prospectively could be a partner in cooperation. But Tomasello emphasizes that second-personal morality also consists of a sense of *fairness* (Tomasello 2018a: 664 – 665; Tomasello 2020: 5-6; Tomasello 2016: loc. 808 – 823). The sense of fairness is an attitude of impartiality where partners in dyadic cooperation recognize each other "as equally deserving individuals, equally worthy of respect" (Tomasello 2020: 6). This sense of fairness is based on a recognition of *self-other equivalence*, which arises when cooperating partners both understand that they each have a role that they should perform as a means of achieving a jointly intended goal, and that there are impersonal criteria for the proper performance of every role (Tomasello 2018a: 665). Second-personal morality also includes a capacity for *joint commitment*—a communicative act in which cooperating partners both pledge to fulfill their respective roles and adhere to a fair division of the gains. Furthermore, joint commitment includes an implicit or explicit avowal that whoever reneges on their pledge to cooperate deserves to be sanctioned. In addition, joint commitment involves a

capacity for deviants to feel guilt as a result of violating the terms of the cooperative partnership (Ibid.).[9]

The second transition in Tomasello's account led to the sense of moral obligation that we modern humans or *Homo sapiens* have. Tomasello dubs it "'*objective*' *morality*" (see Tomasello 2018a: 666 – 667; Tomasello 2016: loc. 163 – 179, 1712). According to Tomasello, two great demographic shifts gave rise to *Homo sapiens* about 150,000 years ago (Tomasello 2016: loc. 154; Tomasello 2018b).[10] First, intense competition between groups forced ancestral hominins to seek protection from marauders by coalescing into more tightly knit social groups. The groups created divisions of labor on which all group members depended for their survival in foraging and defense. Second, population growth led to tribal organization. Small foraging bands composed of a few families numbering in the dozens united into much larger tribes composed of thousands of individuals. Members of the same tribe cooperated among themselves, while they competed with other tribes. Fellow tribespeople included *unfamiliar non-kin*—individuals who neither had any genetic relation nor any history of face-to-face interaction with one another. However, it was essential for the early humans to differentiate unfamiliar members of their own tribe from outsiders. For only people in the same tribe could rely upon each other for cooperation and protection. Consequently, the tribes formed distinct *cultures* which served as markers of shared group allegiance, values, and skills. Those who shared the same tribal membership

[9] Tomasello stresses that chimpanzees and bonobos, our closest evolutionary cousins, do not have a second-personal morality because they do not form joint commitments. They do collaborate with others to acquire food, mates, and social dominance. And they do exhibit helping behavior which suggests that they feel sympathy for others in need. However, their sympathy is limited to those with whom they have collaborated in the past. And, crucially, they do not exhibit resentment elicited by a perception of unfair treatment (Tomasello 2016: loc. 431 - 692).

[10] Our species may well be older. Recent excavations of fossils from Jebel Irhoud in Morocco have been dated to 315,000 years ago (Boyd & Silk 2018: 325).

exhibited the same manner of speech, dress, food preparation, and the like (Tomasello 2018a: 666). So, the ancestral humans who survived and reproduced most successfully were the ones whose psychologies enabled them to learn the ways of their culture, conform to their culture's practices, teach their cultural practices to others, maintain a strong sense of cultural identity and allegiance, and generally care for the welfare of the cultural group (Tomasello 2018b; Tomasello 2020: 7 – 8).

With these demographic changes, the sympathies of individuals scaled up to a concern for all members of the cultural group, including unfamiliar non-kin in that group. The impartial sense of fairness also scaled up. Now it was understood that a complex division of labor, consisting of many interdependent roles performed by many individuals, had to be sustained in order to achieve *collectively* intended goals. Joint commitments gave way to social norms. Each member of the cultural group expected all members to comply with the group's norms; each was disposed to sanction norm-violators; and each felt accountable to social norms in such a way that one's own failure to comply would induce guilt and a troubling sense of identity-loss. Social norms were also internalized psychologically as an objective "view from nowhere." They were accepted by all group members as normative standards that everyone was obliged to live up to. At the same time, it appeared to group members that the social norms did not issue from any single individual (Tomasello 2016: loc. 2944 – 2961). This internalization of social norms extended impartial attitudes, so that all groupmates were thought to be equally deserving of others' compliance with the prevailing norms (Tomasello 2016: loc. 168, 2969).[11]

[11] The impartial perspective generated by social norms does not guarantee equal *status* in society. Of course, social norms can allow for gross inequalities in power, prestige, privilege, and wealth. Instead, the impartial attitude that arises from social norms is the attitude that everyone ought to comply with prevailing norms.

Tomasello's account explains how three facets of the moral sense may have been adaptations for early human cooperation—namely, (1) the capacity to make "ought" judgments, along with (2) sympathy and (3) the disposition to abide by norms. Hereafter, I will explain how these three facets operate as transmission biases favoring the cultural evolution of extended benevolence.

6 How Extended Benevolence Culturally Evolved

An explanatory challenge makes it difficult to see how extendedly benevolent behaviors and institutions could originally emerge. The challenge can be expressed as a question: why wouldn't the moral sense evolve to motivate parochial and xenophobic behaviors that exclusively serve the interests of a cultural in-group? Tomasello himself entertains the idea that what we modern humans consider to be our cultural in-group could potentially be extended to include all of humanity (see Tomasello 2016: loc. 182; Tomasello 2018b; Tomasello 2020: 7). But if, as Tomasello explicitly argues, our ancestors survived by making *distinctions* between insiders and outsiders, then wouldn't selection pressures eliminate any psychological tendency to perceive one's in-group as the whole human population? Moreover, even if we grant that evolutionary forces permitted a moral psychology that sees the entire human species as an all-inclusive in-group, the details of how this orientation would arise by cultural evolutionary processes are not clear. For this reason, Allen Buchanan and Russell Powell (2018) voice skepticism about the prospects for a Boyd and Richerson-style explanation of the cultural evolution of "inclusivist morality," which is Buchanan and Powell's term for extended benevolence (Buchanan & Powell 2018: 175). Buchanan and Powell even insist that cultural evolutionary transmission biases

"*cannot* explain why inclusivist norms rose to sufficiently high frequencies…or [were] found to be persuasive by large segments of the population" (Ibid., emphasis added).

My response to this explanatory challenge calls attention to an adaptive problem our ancestors faced. Indeed, they would have needed to distinguish unfamiliar non-kin of the same tribe, who were usually more reliable as cooperation partners, from outsiders who were usually less reliable. Because unfamiliar members of the same tribe needed to identify one another as trustworthy collaborators, symbols and rituals were used as markers of group identity.[12] *Symbols* are things to which meaning is ascribed by a social convention (Wurz 2012). A *ritual* is a pattern of behavior practiced by a social group. Rituals are often symbolic in that they carry meaning for the people who practice them. Anthropologist Joseph Henrich classifies rituals as a type of social norm (Henrich 2016: 36).

A team of anthropologists led by Kim Hill studied the social ties that bind collections of hunter-gatherer bands into a tribe (Henrich 2016: 162 – 164; Hill et al. 2014). They found that *ritual relationships* were more important than genetic and affinal relationships in facilitating crucial patterns of cooperation such as the sharing of meat and information, as well as receiving help when one is sick or injured (Henrich 2016: 163). Ritual relationships, such as participating in multi-band sparring clubs, were found to be strong predictors of inter-band interactions for two mobile hunter-gatherer groups—the Aché and Hadza (Hill et al. 2014: 7). Boyd and Richerson also highlight that symbolic markers of group identity

[12] There is accumulating evidence that early *Homo sapiens* were engaging in symbolic and ritual behavior by around 70,000 years ago (Boyd & Silk 2018: 327-330). Perforated shell beads were excavated from the Grotte de Pigeons cave in Morocco. This site is dated to 82,000 years ago. Some of the shells were painted with red ocher, and may have been worn on a cord or attached to clothing. Today, African peoples commonly use red ocher for symbolic purposes (Boyd & Silk 2018: 329). At Diepkloof Rock Shelter in South Africa, 60,000-year-old ostrich shell fragments were found. The shells were decorated with geometric patterns, and are believed to signify group identity in the same way that pottery decorations do for modern foragers today (Ibid.).

include shared language, dialect, styles of dress, and common adherence to rituals. Rituals include "gift exchanges, ceremonial activities, and rules of exogamy," and they are among the symbolic markers that can provide human groups with a kind of insurance against misfortune (Richerson & Boyd 2005: 221). For instance, the North American Blackfeet once hunted bison as their core subsistence activity. Since failed hunts were common, the Blackfeet developed a tribal-scale network of relationships among smaller bands. This allowed bands that had been unsuccessful in their hunts to seek the assistance of more successful bands within the same tribe (Richerson & Boyd 2005: 227).

Additionally, Boyd and Richerson have argued that symbolic markers were used by our ancestors to reap the benefits of cooperation *among* tribal societies (Richerson & Boyd 2004). Boyd and Richerson explain that the late Pleistocene hunter-gatherer ancestors who left Africa some 50,000 years ago maintained complex toolkits. These toolkits would have required a huge social network of people far larger than a tribe to correct accumulations of errors in reproducing the tools (Richerson & Boyd 2004: 69). In addition, there were other benefits that *Homo sapiens* attained through inter-tribal cooperation, including military alliances, long-distance trade, and intermarriage (Ibid.). For such inter-tribal cooperation and tool refinement to be possible, modern humans needed to find some means of signaling their reliability as collaborators to the people of other tribes. The solutions they came up with were of the same kind as strategies used to bring people together at the tribal level. They constructed symbols and rituals that turned out to have the power to unify countless masses under a single mega-group identity.

Boyd and Richerson reserve the term "*workaround*" for symbolic markers that can be used to establish mega-group identities (Richerson & Boyd 2004: 69 – 71). Like the

symbolic markers of intra-tribal membership, workarounds include symbols and rituals. Unlike the intra-tribal markers, workarounds could designate membership in nations comprising over a billion inhabitants who participate in a vast division of labor. Architectural monuments, for instance, provide symbols of national identity, and they can serve as sites of mass ritual performances.

Religions and political ideologies also perform the function of workarounds; indeed, they can bind people into mega-groups even larger than the nation. Boyd and Richerson attribute to "humanistic," "universalistic," and "liberal" ideologies the potential to establish an inclusive "global village" identity. A "global village" identity may form the basis for extendedly benevolent concern to all human beings, all sentient creatures, and even all the constituents of the biosphere (Richerson & Boyd 2004: 71, 73). The claim that ideologies can generate an ultra-inclusive group identity finds support in the research reviewed in section 3. That work suggested that normative attitudes favoring extended benevolence, including beliefs in animals rights and emancipative values, have an impact in bringing about extendedly benevolent institutions such as animal welfare protections, human rights, and democracy.

From Tomasello's research, we've seen evidence that the disposition to learn, follow, and enforce norms is likely to be an adapted element of the human moral sense. Following Joseph Henrich, let us call this disposition *norm psychology* (Henrich 2016: 188 - 189). Crucially, normative attitudes expressing a commitment to extended benevolence could spread through norms. In their comprehensive account of norms, Geoffrey Brennan and colleagues (2013) characterize norms as *clusters of normative attitudes*. On this view, a normative principle P is a norm within a group G if and only if (i) a significant proportion of

the members of G accept P and (ii) a significant proportion of the members of G believe that (i) is true (Brennan et al. 2013: 29). Under these conditions, people will be motivated to act as principle P prescribes. The fact that (i) and (ii) are conditions for the presence of a norm strongly suggests that a frequency-based bias is a key enabler of the emergence of norms. Indeed, one way in which new norms may emerge is through a "normative cascade" (Brennan et al. 2013: 98 - 99). Typically, there is variation among individuals in their respective *population thresholds* for what proportion of other people in their social group need to accept a normative principle P, before they are willing to accept P themselves. Some group members may have low population thresholds, in the sense that they accept a normative principle and are willing to follow it even when a very low proportion of others in their community share their attitude. Other group members may have somewhat higher thresholds, so that they become willing to accept and follow normative principle P only if they observe that a higher proportion of the community already accepts and follows P. Indeed, there may be a diffuse distribution of such thresholds in the relevant community. If there is, then a normative cascade may unfold: a few innovators who adopt a novel normative principle P may convince a few others with low population thresholds to accept P, and then this larger mass of individuals convinces still more people with slightly higher thresholds to accept P, and so on until virtually the entire community accepts and follows P. Normative cascades have been cited to explain the end of footbinding in China and the abandonment of female genital mutilation in hundreds of villages across Northwest Africa (Brennan et al. 2013: 99; Mackie and LeJeune 2009). Furthermore, Brian Greenhill (see section 4) found limited empirical support for the operation of a normative cascade in the establishment of human rights cultures within IGOs (Greenhill 2015: 98 – 101).

7 The Proliferation of Extended Benevolence

I've been arguing that normative attitudes favorable to extended benevolence can culturally evolve through norms and symbolic markers. Now I'll make a case for a final point: sympathy—mediated by contact and perspective-taking—is an adapted transmission bias that can explain how extended benevolence came to be as widespread as it is.

The definitions of "sympathy" and closely associated terms, like "empathy," have long been debated (for discussion, see Zaki 2019: 178 – 182). In the *Descent of Man*, Darwin uses "sympathy" to designate an emotion that motivates an individual to help others (see section 2 above; Darwin 1879: 121). Tomasello uses the term in the same way (e.g., see Tomasello 2016: loc. 80, 496-520). Other authors use different words—words such as "compassion" and "empathic concern"—to refer to this same emotion that motivates helping (see Zaki 2019: 180). I shall follow Darwin and Tomasello's usage of "sympathy."

I argue that sympathy works as a content-based, adapted transmission bias driving the adoption of the behaviors, institutions, and attitudes associated with extended benevolence.[13] To see how this happens, we need to appreciate how sympathy is mediated by two factors: contact and perspective-taking. The more people make contact with and take the perspective of others, the more they sympathize with others. Extended benevolence involves sympathy for all nations, races, and perhaps even all sentient beings. Such expansive sympathy is the product of social environments in which there is an extreme

[13] There is considerable evidence that sympathy is an adaptation. Chimpanzees and humans may share a common ancestor that possessed a capacity for sympathy. In experimental settings, chimpanzees have been observed helping conspecifics who they observe to be in need. For instance, chimpanzees help conspecifics trying to get food and tools (Tomasello 2016: loc. 596-618). In addition, human beings seem to be born with a capacity to sympathize. As Tomasello notes, infants as young as fourteen months help unfamiliar adults to fetch out-of-reach objects, and they comfort others who show signs of distress (Tomasello 2016: loc. 929-952). To explain why a sympathetic capacity might have enhanced the reproductive success of our ancestors, evolutionary theorists have cited the mechanisms of kin selection, mutualism, direct reciprocity, social selection, and cultural group selection (Tomasello 2016: loc. 225-414).

abundance of opportunities for people to make contact with each other and take one another's perspectives.

In social psychology, contact has consistently been found to reduce intergroup hostility, especially in the context of cooperative pursuits of common goals (Paluck et al. 2019). Granted, it is possible for contact to intensify intergroup antagonisms, because the groups are sometimes unable to reconcile their differences. This is known as *negative* intergroup contact. Nevertheless, there is evidence that *positive* intergroup contact, which results in *diminished* antagonisms, is more frequent than negative contact. Thus, the cumulative effect of many intergroup contacts can be reduced hostility and increased goodwill overall (Pettigrew 2008; Graf et al. 2014).

Contact is also an enabling condition for sympathy. When people are in contact with others, it presents them with the opportunity to take their perspective—to imagine what it would feel like to be in their situation. Another consistent finding from social psychology is that taking another person's perspective can generate sympathy, which in turn acts as a motivation to help the other (Stich, Doris, and Roedder 2010: 172-174). For instance, in a study by Dovidio et al. (1990), subjects who were instructed to take the perspective of a young woman in distress were more likely to help the woman. Vaish, Carpenter, and Tomasello (2009) found that 18-month-old children would look with concern at and subsequently help a person suffering from an injury, even when the victim did not display any overt emotions. This suggests that, despite the lack of overt emotional cues, the children could take the perspective of the victim, understand that they need help, and then actually offer help.

If sympathy produces extended benevolence, then the enabling conditions of sympathy—namely, contact and perspective-taking—should predict the presence of extended benevolence. And this relationship does indeed hold. I suggested earlier (in section 3) that extended benevolence manifests in the acceptance of emancipative values. It turns out that the extent to which people hold emancipative values is associated with a form of contact called "connective opportunities." A core proposition of Christian Welzel's research is that the popular acceptance of emancipative values can be predicted by three socioeconomic factors, which Welzel calls *action resources*: (1) material resources such as food, shelter and income; (2) intellectual resources such as information, skills, and education; and (3) connective opportunities such as modern transportation and mass communications (Welzel 2013: loc. 2979 – 3097). In one analysis, Welzel uses a country's per capita GDP as a measure of material resources, the average number of schooling years in a country as a measure of intellectual skills, and internet access per 1000 persons as a measure of connective opportunities. He again relies on the World Values Surveys to measure the acceptance of emancipative values within a society. He then runs regressions of emancipative values against these three socio-economic measures for samples of 60 to 80 societies, and finds that 57% of the variation in emancipative values is explained by GDP per capita, 64% is explained by schooling years, and 67% is explained by internet access (Welzel 2013: loc. 2979 – 2990). Other statistical models Welzel constructs with different measures and time-lagged data indicate the same strong dependency of emancipative values on the three action resources (Welzel 2013: ch. 4).

Welzel's finding that emancipative values depend on connective opportunities is unsurprising in light of the relationship between contact and sympathy. Access to transportation and communication technologies raises the likelihood that different

people—including people from very different walks of life—will come into contact. Through such enhanced contact, people have more opportunities to take the perspectives of others. Taking more perspectives could broaden people's sympathies for others, and as a result, people may be more inclined to adopt emancipative values—values that uphold equality of opportunity and freedom of choice for everyone.[14]

Furthermore, perspective-taking and sympathy may also explain the spread of extended benevolence toward nonhuman animals. Perspective-taking can be facilitated in many ways. One way is through texts and images that document the plights of others. Jasper's work (discussed in section 3) traced the way that texts and images recording the suffering of animals impelled people to join the animal rights movement. Additionally, Brian Lowe and Caryn Ginsberg (2002) conducted a survey of 100 animal rights activists from North America, Europe, and South Korea, and found that a strong majority of respondents rated pamphlets (75%) and books (76%) as a somewhat important or very important influence which had prompted them to get involved in the animal rights movement (Lowe & Ginsberg 2002: 207 – 208). In addition, pamphlets (87%) and books (83%) were overwhelmingly rated by the respondents as either "somewhat" or "very" important in their work to influence others. This finding lends credibility to the idea that texts and images provide opportunities to take the perspectives of others (see also Tamir et al. 2016). Successful perspective-taking can prime sympathy for humans and nonhumans alike.

[14] Welzel identifies internet connectivity as a form of connective opportunity. It may be doubted whether internet connectivity promotes extended benevolence rather than antipathy between different ideological groups who segregate themselves in digital "bubbles." While this question is certainly deserving of further study, a recent analysis by Jha and Kodila-Tedika (2020) found a strong positive correlation between the use of Facebook and democracy ratings in a cross-section of 125 countries. Evidently, there is no tension between social media and one form of an extendedly benevolent institution.

To summarize, the explanatory challenge was the task of explaining how the moral sense could evolve in such a way that it fosters extended benevolence beyond one's cultural in-group and even beyond one's species. My response to this challenge has been that ideological workarounds, norm psychology, contact, and perspective-taking can extend the range of beings with whom one sympathizes to include cultural outsiders and animals.

Here it may be objected that Tomasello's model is inconsistent with the above theory of the cultural evolution of extended benevolence. According to this objection, Tomasello's account predicts that there would be strong constraints on the scope of human benevolence. The reason is that there would have been no fitness-advantage for our hominin ancestors to sympathize with out-groups and animals. Instead, only cooperation with members of symbolically marked in-groups would have been fitness-enhancing, since on Tomasello's account, only other members of one's symbolically marked in-group would have been reliable and trustworthy partners in cooperation. An ancestral individual who was inclined to cooperate with outsiders would often be exploited by them; an ancestor who helped animals would get virtually no fitness-benefit from their helpful acts, since most animals can't cooperate in the ways that are most important for human survival.

My account of extended benevolence has claimed that some of the components of the moral sense—namely, norm psychology and sympathy—are adapted transmission biases. Adapted biases are *adaptations*—i.e., they exist because they helped our hominin ancestors to survive and reproduce in their environments. Although I've argued that extended benevolence is a product of the moral sense, I need not commit myself to the dubious idea that extended benevolence itself ever enhanced ancestral reproductive success. For some *products* of adaptations are not adaptations themselves. While adapted

transmission biases are adaptations, the cultural variants they select or generate may not be. As a case in point, Boyd and Richerson cite the trend of declining birth rates in developed countries known as the *demographic transition* (Richerson & Boyd 2005: 169 – 174). Modern economies typically include educated professionals—doctors, lawyers, managers, politicians—who tend to achieve high salaries and social status. Attaining that status normally requires investing considerable time in an education and career, which often limits the time people dedicate to raising children. The result is lower fertility rates in countries with highly professionalized workforces. Looking at the demographic transition from a cultural evolutionary perspective, prestige bias and success bias could explain why people prioritize education and careers over childrearing. If the high-status, successful people are well-educated professionals who have just a few children, their life choices will be imitated by others. So, although prestige and success biases are plausibly adaptations, some of the behaviors (the cultural variants) they motivate may be downright detrimental to our reproductive success. Similarly, it's possible that the moral sense is an assemblage of adaptations that enhanced our ancestors' reproductive success by facilitating cooperation, while its component capacities for sympathy and a sense of fairness are capable of selecting behaviors that do not at all advance reproductive success. Thus, even if extended benevolence is not itself adaptive, this is consistent with the claim that the moral sense is an adaptation which gave rise to extended benevolence.[15]

8 Conclusion: An Evolutionary Foundation for Extended Benevolence

[15] It is an open question whether extended benevolence will ever be outmoded by alternative cultural variants that do promote the reproductive success of individuals or groups. While this is possible, it is not inevitable. Whether or not it actually comes to pass depends on the relative strength of natural selection *against* extended benevolence compared to the cultural evolutionary forces that *favor* extended benevolence.

I conclude that the emergence and proliferation of extended benevolence can be explained to a significant extent by cultural evolutionary forces. The explanatory strategy of cultural evolutionary theory is recognizably Darwinian in style, since it characterizes some cultural evolutionary forces—namely, the adapted transmission biases—as adaptations. Moreover, the account defended above suggests that three of the five components of the moral sense identified by Darwin are sufficient to explain the emergence of extended benevolence: namely, the capacity to make normative judgments, the disposition to comply with community rules (norm psychology), and sympathy. Some commentators, including Buchanan and Powell, have doubted that evolutionary mechanisms could account for extended benevolence. However, I've argued that these observers underestimate the explanatory resources of cultural evolutionary theory. When we look to our deep past, we do find ample indication that our ancestors were parochial and xenophobic. But we can also find, in the historical process of our becoming cultural creatures, the better angels of our nature.

Acknowledgements

This research is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 1 (RG70/16).

References

Birch, Jonathan. 2017. *The Philosophy of Social Evolution*. Oxford: Oxford University Press.

Boyd, Robert and Joan B. Silk. 2018. *How Humans Evolved*, 8th edition. New York: W.W. Norton & Company.

Brennan, Geoffrey, Lina Eriksson, Robert E. Goodin, and Nicholas Southwood. 2013. *Explaining Norms*. Oxford: Oxford University Press.

Buchanan, Allen and Russell Powell. 2018. *The Evolution of Moral Progress: A Biocultural*

Theory. Oxford: Oxford University Press.

Cooney, Nick. 2014. *Veganomics: The Surprising Science on Vegetarians, from the Breakfast Table to the Bedroom*. CreateSpace Independent Publishing Platform. Kindle Edition.

Darwin, Charles. [1879] 2004. *The Descent of Man, and Selection in Relation to Sex*. London: Penguin Random House UK.

Dovidio, John F., Judith Allen, and David A. Schroeder. 1990. Specificity of Empathy-induced Helping: Evidence for Altruistic Motivation. *Journal of Personality and Social Psychology*, 59, no. 2): 249–60.

Fariss, Christopher J. 2014. Respect for Human Rights has Improved Over Time: Modeling the Changing Standard of Accountability. *The American Political Science Review* 108, no. 2 (May): 297 - 318.

Freedom House. 2020. Countries and Territories. From *Freedom in the World 2020*. URL: https://freedomhouse.org/countries/freedom-world/scores

Giménez, Emiliano. 2015. Argentine orangutan granted unprecedented legal rights. *CNN Espanol*, 4 January 2015. URL: https://edition.cnn.com/2014/12/23/world/americas/feat-orangutan-rights-ruling/index.html

Graf, Sylvie, Stefania Paolini, and Mark Rubin. 2014. Negative intergroup contact is more influential, but positive intergroup contact is more common: Assessing contact prominence and contact prevalence in five Central European countries. *European Journal of Social Psychology*, 44, 536–547.

Greenhill, Brian. 2015. *Transmitting Rights: International Organizations and the Diffusion of Human Rights Practices.* Oxford: Oxford University Press.

Hawkins, Roxanne D., Joanne M. Williams, and the Scottish Society for the Prevention of

Cruelty to Animals. 2017. Childhood Attachment to Pets: Associations between Pet Attachment, Attitudes to Animals, Compassion, and Humane Behaviour. *International Journal of Environmental Research and Public Health* 14, no. 5: 490.

Henrich, Joseph. 2016. *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton, N.J.: Princeton University Press.

Hill, Kim R., Brian M. Wood, Jacopo Baggio, A. Magdalena Hurtado, and Robert T. Boyd. 2014. Hunter-Gatherer Inter-Band Interaction Rates: Implications for Cumulative Culture. *PLoS ONE* 9, no. 7 (July): e102806.

Inglehart, Ronald F. 2018. *Cultural Evolution: People's Motivations Are Changing, and Reshaping the World*. Cambridge: Cambridge University Press.

Inglehart, Ronald and Pippa Norris. 2003. *Rising Tide: Gender Equality and Cultural Change Around the World*. Cambridge: Cambridge University Press.

Inglehart, Ronald and Christian Welzel. 2005. *Modernization, Cultural Change, and Democracy The Human Development Sequence*. Cambridge: Cambridge University Press.

Jasper, James M. 1997. *The Art of Moral Protest: Culture, Biography, and Creativity in Social Movements*. Chicago: University of Chicago Press.

Jasper, James M. and Dorothy Nelkin. 1992. *The Animal Rights Crusade: The Growth of a Moral Protest.* New York: Free Press.

Jha, Chandan Kumar and Oasis Kodila-Tedika. 2020. Does social media promote democracy? Some empirical evidence. *Journal of Policy Modeling* 42: 271–290.

Lewens, Tim. 2015. *Cultural Evolution*. Oxford: Oxford University Press.

Lowe, Brian M. and Caryn F. Ginsberg. 2002. Animal Rights as a Post-Citizenship Movement.

Society & Animals 10, no. 2: 203 – 215.

Lührmann, Anna, Seraphine F. Maerz, Sandra Grahn, Nazifa Alizada, Lisa Gastaldi, Sebastian Hellmeier, Garry Hindle and Staffan I. Lindberg. 2020. *Autocratization Surges – Resistance Grows*. Democracy Report 2020. Varieties of Democracy Institute (V-Dem).

Mackie, Gerry and John LeJeune. 2009. Social Dynamics of Abandonment of Harmful Practices: A New Look at the Theory. United Nations Children's Fund (UNICEF) Innocenti Research Centre.

Merry, Sally Engle. 2006. *Human Rights and Gender Violence: Translating International Law into Local Justice*. Chicago: University of Chicago Press.

Nichols, Shaun. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment.* Oxford: Oxford University Press. Kindle Edition.

Paluck, Elizabeth Levy, Seth A. Green, and Donald P. Green. 2019. The contact hypothesis re-evaluated. Be*havioural Public Policy* 3, no. 2: 129–158.

Pettigrew, Thomas F. 2008. Future directions for intergroup contact theory and research. *International Journal of Intercultural Relations* 32: 187–199.

Richerson, Peter J. and Robert Boyd. 2004. Darwinian Evolutionary Ethics: Between Patriotism and Sympathy. In *Evolution and Ethics: Human Morality in Biological and Religious Perspective*, ed. Philip Clayton and Jeffrey Schloss, 50 – 77. Cambridge, UK: Wm. B. Eerdmans Publishing Co.

Richerson, Peter J. and Robert Boyd. 2005. *Not By Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Roser, Max. 2020. Democracy. Published online at *OurWorldInData.org.* URL: https://ourworldindata.org/democracy

Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics*. Cambridge: Cambridge University Press.

Stich, Stephen, John M. Doris, and Erica Roedder. 2010. Altruism. In *The Moral Psychology Handbook*, ed. John M. Doris and the Moral Psychology Reading Group, 147 – 205. Oxford: Oxford University Press.

Tamir, Diana I., Andrew B. Bricker, David Dodell-Feder, and Jason P. Mitchell. 2016. Reading fiction and reading minds: the role of simulation in the default network. *Social Cognitive and Affective Neuroscience* 11, no. 2: 215–224.

Tomasello, Michael. 2016. *A Natural History of Human Morality*. Cambridge, MA: Harvard University Press. Kindle Edition.

Tomasello, Michael. 2018a. Precís of a natural history of human morality. *Philosophical Psychology* 31, no. 5: 661 – 668.

Tomasello, Michael. 2018b. How We Learned to Put Our Faith in One Another's Hands: The Origins of Morality. *Scientific American* 319, no. 3 (September): 70 – 75.

Tomasello, Michael. 2020. The Moral Psychology of Obligation. *Behavioral and Brain Sciences*, 43, e56: 1 – 58.

Treisman, Daniel. 2018. Is democracy really in danger? The picture is not as dire as you think. *The Washington Post*. 19 June 2018.

Vaish, Amrisha, Malinda Carpenter, and Michael Tomasello. 2009. Sympathy Through Affective Perspective Taking and Its Relation to Prosocial Behavior in Toddlers. *Developmental Psychology* 45, no. 2: 534–543.

Waldau, Paul. 2011. *Animal Rights: What Everyone Needs to Know*. Oxford: Oxford University Press.

Welzel, Christian. 2013. *Freedom Rising: Human Empowerment and the Quest for*

Emancipation. Cambridge: Cambridge University Press. Kindle Edition.

Welzel, Christian, Ronald Inglehart, Patrick Bernhangen, and Christian W. Haerpfer. 2019. Introduction. In *Democratization*, 2nd ed. Christian Haerpfer, Patrick Bernhagen, Christian Welzel, and Ronald F. Inglehart, 1 – 18. Oxford: Oxford University Press.

Wurz, Sarah. 2012. The Transition to Modern Behavior. *Nature Education Knowledge* 3, no. 10: 15.

Zaki, Jamil. 2019. *The War for Kindness: Building Empathy in a Fractured World*. New York: Broadway Books.