8

The Mentalist Theory of Ethics and Law

When a man denominates another his enemy, his rival, his antagonist, his adversary, he is understood to speak the language of self-love, and to express sentiments, peculiar to himself, and arising from his particular circumstances and situation. But when he bestows on any man the epithets of vicious or odious or depraved, he then speaks another language, and expresses sentiments, in which he expects all his audience are to concur with him. He must here, therefore, depart from his private and particular situation, and must choose a point of view, common to him with others; he must move some universal principle of the human frame, and touch a string to which all mankind have an accord and symphony.

David Hume, *An Enquiry Concerning the Principles of Morals*

No doubt wickedness hath its rewards to distribute; but whoso wins in this devil's game must needs be baser, more cruel, more brutal than the order of this planet will allow for the multitude born of woman, the most of these carrying a form of conscience – a fear which is the shadow of justice, a pity which is the shadow of love – that hindereth from the prize of serene wickedness, itself difficult of maintenance in our composite flesh.

George Eliot, *Daniel Deronda*

8.1 A FRESH LOOK AT FRAMEWORKS OF MORALITY

Our discussion thus far has shown that a solid analytical theory of morality is crucial as a starting point for further theory-building about the nature and origin of moral cognition. Once this analytical theory of morality has been achieved, we can attempt to reconstruct the psychological mechanisms underpinning human moral judgment and investigate how they are acquired. There are no a priori constraints on the ways that human morality *could* be structured, and in particular no constraints that can be derived from the theory of evolution. What plausibly can be taken as constitutive of human morality is simply a matter of the evidence given. Therefore, any account of morality's constitutive elements is well-advised to take

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

seriously the rich debates of moral philosophy, whose insights must inform plausible theories of morality.

In order to develop an analytical theory of morality, one needs to identify the building blocks of the human moral world. Such a theory must be based on an analysis of the practice and phenomenology of moral judgment. It needs to tell us what humans actually do when they exercise their moral understanding. One complication of such a study – and a major one at that – is the fact that moral judgments are intrinsically contested. Which moral judgments form the basis of theory-building? The views of a misogynist racist or those of a female Black Lives Matter activist?

One plausible way of proceeding is to look at qualified moral judgments. These judgments have to be "considered judgments," to borrow a useful term, in the sense that they are reflective, dispassionate judgments that are not skewed by interest, passion, errors of fact and so forth.[1] This methodological move has as its background the distinction between competence and performance, the faculty to perform a certain cognitive task and the actual performance of this task.[2] Humans have the competence to construct an image of the external world using the specific structures of their visual cognition. This does not mean that the effects of imbibing a certain amount of alcohol may not affect the functioning of this competence and make what is perceived appear strangely blurred. As this example shows, only indirect conclusions about an agent's competence can be drawn from their performance, because the performance is influenced by many other factors than just the structure of their competence. There is no doubt that the loud sound of techno music will influence the mathematical problem-solving capacity of people exposed to it. However, nobody would ever entertain the idea that techno music is of great relevance to studying the cognitive apparatus enabling humans to do math. The capacity for moral evaluation is another such competence usually possessed by human beings. Nevertheless, the performance of this capacity, the final evaluation of an action can be biased – for example, by the interests of the evaluating person. Consequently, such influences need to be factored out of the analysis if we are to properly study the cognitive competence in question, which is not an easy thing, particularly in empirical work.

[1] Cf. on this matter John Rawls, *A Theory of Justice*, revised edition (Cambridge, MA: Harvard University Press, 1999), 42: "Thus in deciding which of our judgments to take into account we may reasonably select some and exclude others. For example, we can discard those judgments made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside. All these judgments are likely to be erroneous or to be influenced by an excessive attention to our own interests. Considered judgments are simply those rendered under conditions favorable to the exercise of the sense of justice, and therefore in circumstances where the more common excuses and explanations for making a mistake do not obtain"; Mikhail, *Elements*, 51 ff.

[2] Cf. for the (crucial) competence/performance distinction Noam Chomsky, *Aspects of the Theory of Syntax* (Cambridge, MA: MIT Press, 1965), 3 ff.; Mahlmann, *Rationalismus*, 73 f.; Mikhail, *Elements*, 51 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

This crucial issue is sometimes overlooked in recent moral psychology studies, which claim to be studying human moral competence, but in fact to a surprisingly large degree are concerned with performance problems, such as the skewing of moral judgment by nonmoral factors, from smells[3] to the feeling of being controlled.[4] Another methodological approach to dealing with the contested nature of moral judgment is to look at highly idealized and often artificial cases that appear to be as little politically and culturally loaded as possible.[5] By contrast, studying the human moral faculty by looking at opinions about issues as contested and ideologically charged as abortion, for example, is a methodological nonstarter.

One preliminary result of our discussions so far has been the observation that some concepts of morality in contemporary debates are too narrow, both as to the substantive material principles of morality and as to the subjects of moral concern. In particular, morality is neither an ultimately selfish enterprise seeking to reap the profits of in-group cooperation, nor is it simply a set of preferences or aversions. The analysis below will flesh this out in more detail.

One theoretical approach that explains how these findings may fit into a theory of moral cognition is the so-called mentalist approach to ethics and law. A mentalist model of moral cognition investigates the question of whether it is possible to identify generative principles of moral judgment specific to human moral cognition that are universal and uniform across the species – a universal moral grammar, if you will, to use a metaphor sometimes employed to capture the basic intuition of this approach.[6] The mentalist model has been a very influential research paradigm in

[3] Simone Schnall et al., "Disgust as Embodied Moral Judgement," *Personality and Social Psychology Bulletin* 34, no. 8 (2008): 1096 ff. Cf. Haidt for further examples, Haidt, *Righteous Mind*, 35 ff.

[4] Cf. Kevin J. Haley and Daniel M. T. Fessler, "Nobody's Watching? Subtle Cues Affect Generosity in an Anonymous Economic Game," *Evolution and Human Behaviour* 26 (2005): 245 ff.

[5] Cf. Mikhail, *Elements*, 56 ff.; Mahlmann, *Rationalismus*, 107.

[6] Cf. e.g. Noam Chomsky, *Language and Problems of Knowledge* (Cambridge, MA: MIT Press, 1988), 152; Matthias Mahlmann and John Mikhail, "Cognitive Science, Ethics and Law," in *Epistemology and Ontology*, ed. Zenon Bankowski (Stuttgart: Franz Steiner Verlag, 2005), 95 ff.; John Mikhail, "Rawls' Linguistic Analogy: A Study of the 'Generative Grammar' Model of Moral Theory Described by John Rawls in *A Theory of Justice*," PhD dissertation, Cornell University, 2000; Mikhail, *Elements*; John Mikhail, "Chomsky and Moral Philosophy," in *The Cambridge Companion to Chomsky*, 2nd edition, ed. James McGilvray (Cambridge: Cambridge University Press, 2017); Mahlmann, *Rationalismus*; Mahlmann, "Ethics," 577 ff.; Gilbert Harman, "Using a Linguistic Analogy to Study Morality," in *Moral Psychology, Vol. 1: The Evolution of Morality*, ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2008), 345 ff.; Erica Roedder and Gilbert Harman, "Linguistics and Moral Theory," in *The Moral Psychology Handbook*, ed. John M. Doris (Oxford: Oxford University Press, 2010), 273 ff.; Ray Jackendoff, *Language, Consciousness, Culture: Essays on Mental Structure* (Cambridge, MA: MIT Press, 2007), 277 ff.; Susan Dwyer, "Moral Competence," in *Philosophy and Linguistics*, eds. Kumiko Murusagi and Robert Stainton (New York: Routledge, 1999), 169 ff.; Marc Hauser, *Moral Minds* (New York: Harper Collins, 2006). For a critique, cf. Pardo and Patterson, *Minds, Brains, and the Law*, 12 ff., 63 ff., especially because of the (externalist)

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

other domains of the theory of mind – for example, in the study of language – stirring many intense controversies and debates.[7] The thrust of the mentalist argument echoes a long tradition of thought on moral understanding – after all, the belief that human beings are endowed with a particular faculty of moral evaluation has been one of the thoughts guiding moral philosophy ever since antiquity.

8.2 SOME PROPERTIES OF MORAL COGNITION

8.2.1 *The Cognitive Space of Morality*

A close and careful look at the phenomenology of morality gives rise to some important observations.[8] One is that something like a moral experience exists at all. Humans operate naturally within a mental space that has a normative dimension. There is a specific mental domain of morality, a qualitatively distinguished element of conscious thought, an introspectively accessible, distinctive, intuited, subjectively experienced aspect of our mental life, a *qualia*, as it is often said, or – to use standard understandings of this term – a certain phenomenal character of certain forms of experience. The availability of such a cognitive domain is not self-evident; rather, this domain represents an empirical property of the human mind that not all organisms share. Human beings do not perceive ultrasound, but bats do. Human beings see the world in the distinct colors of morality, but bats do not. This cognitive domain does not concern a side issue but defines nothing less than an element central to the identity of the human species: the moral dimension of human lives. Its existence consequently merits close attention.

A further interesting observation concerns the fact that there is a highly and intricately qualified limited set of possible objects of moral evaluation. This set already restricts the kinds of morality that are possible. The dropping of an apple from a tree into the hands of a hungry person is not a virtuous action on the tree's part. Or, as Hume observed: "A young tree, which over-tops and destroys its parent, stands in all the same relations with Nero, when he murdered Agrippina," but does not commit matricide.[9] This is because agency is a precondition for the moral evaluation of certain events in the world. If these events cannot be attributed to agents, questions of moral evaluation do not arise.[10]

thesis that there can be (on conceptual grounds) no unconscious rule-following. On Wittgenstein's concept of the rules underlying this argument and its critique, Mahlmann, *Rationalismus*, 121 ff.

[7] Cf. Chomsky, *Aspects of a Theory of Syntax*; Noam Chomsky, *The Minimalist Program* (Cambridge, MA: MIT Press, 1995); Noam Chomsky, *New Horizons in the Study of Language and Mind* (Cambridge: Cambridge University Press, 2000).

[8] Cf. Mahlmann, *Rechtsphilosophie und Rechtstheorie*, 374 ff.

[9] Hume, "Enquiry Concerning the Principles of Morals," 293.

[10] This does not mean that inanimate things have not been taken as agents – cf. Xerxes' whipping of the sea to punish the sea for destroying his bridges, Herodotus, *Histories*, 7.35.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

Moreover, placing your pen gracefully on your desk cannot be the possible object of moral evaluation either (except under very particular circumstances), even though an agent performs this act. However, this act is a possible object of aesthetic evaluation – another distinctive element of human experience.[11] Kicking a ball so as to feel like Dzsenifer Marozsán for a few precious seconds is morally very different from kicking a defender or a dog that is in your way. A precondition of moral evaluation is thus – to put it very roughly – something like volitionally controlled or controllable bodily actions or omissions of agents with consequences for the well-being of sentient beings and other qualified objects of moral concern; intentions concerning such actions or omissions; states of affairs resulting from such intentions, actions and omissions; or qualified emotions directed at the well-being of others.[12]

8.2.2 *Principles of Morality*

As we already have seen, in analyses of morality it is particularly important to distinguish *behavior* and motivational *inclinations* to behavior from the *moral evaluation* of this behavior and the intentions underlying it. Concrete behavior, including what is called prosocial behavior, and the inclination or preference for such behavior are one thing, the reflexive appraisal of this behavior, inclination or preference with deontic dimensions quite another. Only the latter falls within the proper realm of morality and ethics.[13]

If we turn to the content of morality and carefully analyze some qualified (only seemingly simply structured) considered judgments of the kind described above, we see that these judgments seem to be guided by principles of egalitarian justice and altruism across a wide range of cases. That it is just to distribute party favors equally to the young guests of a child's birthday party seems as uncontroversial as that it is a morally good deed to help prevent starvation in Yemen.[14] Respect for others is a further important principle. This state of affairs is not particularly surprising, as all of these principles run through the history of ideas, too, as the core of morality, across cultures and millennia, albeit accompanied throughout by influential skeptical voices from Thrasymachus to Nietzsche and beyond, who have argued (with greater

[11] On the classical distinction between the distinct and potentially contradictory moral and aesthetical evaluation, Kant, *Kritik der praktischen Vernunft*, 204 f.

[12] Intricate problems arise in this area. Actions that have an effect for objects of art or the environment represent a matter of complex debate, for example. The latter in particular are of great practical concern. In both areas, ethical principles matter. Another problem is posed by virtues that are not other-regarding. A fuller statement of the possible objects of moral evaluation would need to take account of these special cases, refining the basic principles stated here.

[13] This point is relevant, for example, to the question of "animal morality," cf. Chapter 7. Prosocial behavior of nonhuman animals does not in itself constitute morality in the sense understood here. On the question of a possible continuum and differences between humans and (nonhuman) animals, e.g. John Mikhail, "Any Animal Whatever? Harmful Battery and Its Elements as Building Blocks of Moral Cognition," *Ethics* 124 (2014): 750 ff.

[14] Cf. for some remarks Mahlmann, *Rechtsphilosophie und Rechtstheorie*, 374 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

or lesser philosophical sophistication) that morality has no content that lends itself to rational reconstruction and that apparently core normative concepts do not mean anything at all.[15]

Substantial empirical work carried out in recent years points in the same direction, supporting the observation that certain identifiable elements – more precisely, egalitarian and altruistic principles – seem to play an important role in the evaluation of the justness and goodness of actions, despite some theoretical limits of parts of this research that already have been reviewed above.[16] Other frontiers of research include the prohibition of the instrumentalization of human beings that plausibly lies at the heart of a proper analysis of the extensively empirically researched trolley problems and distinctions between different subjective cognitive and volitional attitudes towards potential actions, including direct and oblique intentions and their relevance for human moral evaluation.[17]

If we analyze the empirical work on justice and basic uncontroversial judgments about the justice and injustice of intentions, actions and states of affairs and do not forget what the struggles of social history seem to suggest about ideas of justice, then differentiated principles of equality as already discussed above[18] seem to have considerable explanatory power to account for many patterns of moral evaluation. We have said that a just distribution demands proportional equality between the value of the specific criterion of distribution reasonably related to a particular sphere of distribution on the one hand and the amount of the good distributed on the other. Other normative demands concern the equal treatment of persons, restitution and respect for the equal worth of each human being – as indicated, the ultimate yardstick of just treatment and a just state of affairs. Such egalitarian principles match the empirically identified patterns of moral evaluation recalled in this study, including standard examples such as the ultimatum or dictator game, the many debates about the proper interpretation of these findings notwithstanding.[19] The same also holds for other clues to the moral world in which human beings live beyond experimental data, not least social history and its many egalitarian struggles, some of which we have recalled. It should be noted that a main bone of contention in these political struggles was not the

[15] Cf. for just two more recent examples Kelsen's attempt to show the emptiness of concepts of justice, Hans Kelsen, "Das Problem der Gerechtigkeit," in Hans Kelsen, *Reine Rechtslehre* (Vienna: Deuticke, 1960), 357 ff., or Luhmann's idea that justice is a "contingency formula" (*Kontingenzformel*) whose function is to hide that law is not based on a notion of material legitimacy because no such legitimacy exists, Luhmann, *Recht der Gesellschaft*, 235.

[16] Cf. Chapter 6.

[17] Cf. for instance Edmond Awad et al., "Universals and Variations in Moral Decisions Made in 42 Countries by 70 000 Participants," Proceedings of the National Academy of Sciences 117, no. 5 (2020): 2332–7, concluding that the observed patterns (bystander: permissible; footbridge: impermissible) are "best explained by basic cognitive processes rather than cultural norms," 2332. Cf. Chapter 6.

[18] Cf. Chapter 5. Note again that the principles of justice also are foundational for notions of substantial equality.

[19] These results indicate, uncontroversially, egalitarian intuitions and an interest in maintaining principles of just distribution for their own sake. On this and other examples, cf. Chapter 6.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

principle that equals ought to be treated equally, but the criteria of distribution and who and what actually fulfill these criteria. Regarding rights, for instance, an important question in history was whether fundamental rights are distributed in a society on the basis of the bearer's humanity or some other criterion (say, aristocratic birth) and who fulfills this criterion (for instance, whether women are fully human or some kind of deficient being and thus not entitled to a full set of fundamental rights). If we pay due attention to such factors, the core of the human quest for justice becomes considerably more transparent.

To achieve sufficient explanatory depth, however, the analysis of these empirical observations needs to remain aware of the distinction between a moral competence and its actual use, the plurality of motivational factors influencing human action, including nonmoral interests and the complex structure of moral judgments, with their cognitive, volitional and emotional components (to be discussed in more detail below), which are not reducible, for example, to preferences or aversions.

Whether there is such a thing as genuinely other-regarding altruistic behavior or whether any action beneficial to others is ultimately motivated by some (albeit perhaps refined and hidden) self-interest of the agent is one of the traditional questions of practical philosophy.[20] As in the case of justice, this is a huge debate that today is enriched by interesting empirical work.[21] In this context, it is important, too, to rely on a sufficiently complex theory of morality, in particular to distinguish between actual behavior and considered evaluation and thus distinguish the question of whether people are *in fact acting* because of a genuinely altruistic motivation from the question of whether genuine altruism is the precondition for *evaluating* something as morally good. There is not much reason to believe that people generally excel in altruistic behavior. However, this observation tells us nothing about the principles that guide moral judgment – for example, when evaluating the selfish behavior prevalent around us (including our own).[22] Concerning these

[20] Hume, "Enquiry Concerning the Principles of Morals," 212 ff., went to considerable length to refute the selfishness hypothesis, clearly influenced by the arguments of Joseph Butler, "Fifteen Sermons," in *The Works of Joseph Butler*, Vol. II, ed. William Ewart Gladstone (Oxford: Clarendon Press, 1896), 35 ff., 185 ff., or Francis Hutchinson, *An Inquiry into the Original of Our Ideas of Beauty and Virtue*, ed. Wolfgang Leidhold (New York: Garland Publishing, 1971), 125 ff.

[21] Cf. the empirical research on genuinely altruistic motivation and action, summarized in C. Daniel Batson, *Altruism in Humans* (Oxford: Oxford University Press: 2011); C. Daniel Batson, *A Scientific Search for Altruism: Do We Care Only About Ourselves?* (Oxford: Oxford University Press, 2019).

[22] The incongruence of justified moral principles and behavior is not a new observation, cf. Thomas Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970), 146: "To say that altruism and morality are possible in virtue of something basic to human nature is not to say that men are basically good. Men are basically complicated; how good they are depends on whether certain conceptions and ways of thinking have achieved dominance, a dominance which is precarious in any case. The manner in which human beings have conducted themselves so far does not encourage optimism about the moral future of the species."

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

principles of evaluation, there are reasons to think that such a genuinely altruistic motivation is in fact a core element of moral evaluation. More precisely, it seems plausible to assume that an action is morally good if it is performed with the direct intention (or purpose), not only the oblique intention (or knowledge), to foster the well-being of the patient. If this is so, it is irrelevant for the moral evaluation whether or not the fostering of the interests of the agent is – at the same time – a directly intended or foreseen (obliquely intended) consequence of the action and forms a second reason for action in a bundle of motives. The direct intention to foster the well-being of the patient of the action appears to be a necessary condition of morally good action.[23]

To illustrate the meaning of this principle, it is useful to look at one of the most refined versions of ethical egoism. This form of egoism holds that altruistic behavior is ultimately motivated by the desire to experience the satisfaction of having acted in a morally appropriate manner. This argument makes an important point, namely that moral action does indeed provide some particular form of satisfaction for the agent and that agents are certainly often aware of this. In addition, acting immorally can have unpleasant effects, too, such as feelings of shame. These observations do not settle the issue, however. Consider the following case: Pawel helps Mio, thinking: "I do not care for this person and her well-being at all (what a silly person she is, in fact!), it just happens (unfortunately) that I have to do something for her in order to reap the sweet fruit I really desire, namely to feel the satisfaction of being a truly nice person!" Is this really a morally laudable deed? If doubts arise about the moral praiseworthiness of an action with such an intention, it seems to confirm the analysis above. This is because the Pawel has only an oblique intention to help the other person and not the direct intention to be beneficial to her: His direct intention is to satisfy one of his own personal desires, and helping the other person is only a (perhaps even unwelcome) means to achieve that end.[24]

[23] Cf. Mahlmann, *Rechtsphilosophie und Rechtstheorie*, 375 ff. It is assumed that the evaluation of the action is dependent on the nature of the underlying intention. Cf. for the same criteria to provide a *definition of altruism*, Batson, *A Scientific Search for Altruism*, 22 ff. This definition provides an important clarification. The thesis pursued here is that one can take one step further: These criteria are understood as key to evaluating an intention or action as morally good, and thus to ascribe a deontic status to them, not only to identify them as altruistic.

[24] This analysis can be buttressed by the observable asymmetry between responsibility for a foreseen bad side effect and the praiseworthiness of a foreseen good side effect: Only action with the purpose of bringing about a good side effect is morally praiseworthy, not an action with a merely foreseen but not intended good side effect. This asymmetry is traditionally framed in terms of intentions: Bad side effects are taken as intentional, good side effects as unintentional, cf. Joshua Knobe, "Intentional Action and Side Effects in Ordinary Language," *Analysis* 63, no. 3 (2003): 190 ff. Cf. Batson, *A Scientific Search for Altruism*, 41 ff., on sets of experiments that exclude other intentions than empathy-based altruism as motivation for certain other-regarding behaviors. These alternative motivations include the egoistic desire to remove emphatic concern, avoiding guilt, the desire for esteem-enhancing reward, sadness relief, the pleasure of emphatic joy and self–other merging. This is an important result. However, these experiments concern motivation not evaluation and do not include motivations to act because

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

Another point is perhaps worth noting: Justice seems to be something like a limiting condition of morally good action. There are reasons to believe that there is no morally good intention that violates principles of justice. If Pawel, for instance, helps three out of four people in need, not because he cannot help them all, but just because he feels like excluding one person on a whim, this is not a morally good action, despite his direct intention to help the other three, because it violates principles of equal treatment.

A third principle is respect for human beings. Respect in a moral sense must be distinguished from admiration for achievements of other kinds – for instance, for a skillful free kick right into the corner of the goal. Unlike admiration of this kind, respect in a moral sense has prescriptive consequences: One ought to behave towards other human beings in certain ways. One important element is to treat them as ends-in-themselves and not only as means to achieve other purposes, to once again use the Kantian version of an ancient idea. Evidently, reducing somebody to a means of some ulterior purpose degrades this person. As we saw when discussing the trolley problem, substantial empirical data suggest that people do in fact apply this principle when evaluating certain morally salient situations: The recorded moral evaluations of the footbridge scenario and its most promising analysis show that these evaluations are best explained by an operative principle demanding that people not be instrumentalized based on a structural means/ends distinction.[25]

Furthermore, there are other forms of disrespect for human beings that do not instrumentalize them: Regularly flooding a prison cell with feces, not to torture the inmate but simply due to negligence concerning the sanitary conditions of inmates, is an illuminating example of this from constitutional case law.[26]

The importance of recognition as a being of equal worth has rightly been flagged as a key element of social struggles:[27] These struggles are not only about bread-and-butter issues, but also about exploited and degraded groups of people's demand to be respected as human beings. This is a central ethical and political dimension in the fight against slavery, old and new, and in the fight against the subjugation and exploitation of women. The same holds for struggles for the emancipation of the working classes, which likewise were not only about material concerns like decent wages, but also about respect for the humanity of workers, who demanded (and indeed still demand) that they not be reduced to beasts of burden.

of a perceived moral obligation. A full analysis of morality has to account for these elements of human moral judgment, too.

[25] Cf. Chapter 7.

[26] BVerfG: Verfassungsgebot menschenwürdiger Haftbedingungen (March 16, 1993), *Neue Juristische Wochenschrift* 1993, 3190.

[27] Bloch, *Naturrecht und menschliche Würde*; Axel Honneth, *Der Kampf um Anerkennung* (Frankfurt am Main: Suhrkamp, 1992).

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

These findings help us to tackle some of the problems that the supervenience of moral judgments over facts poses.[28] They help us to identify the principles that determine, first, the moral evaluation and normative consequences of moral judgments that are triggered by certain facts (e.g. by the fact of a direct intention to harm somebody) and thus to determine *how* moral judgments supervene over facts (not just that they do) and, second, which facts are morally relevant in the first place (e.g. that the agents' intentions to act matter morally but not their haircuts). These tentatively outlined principles, which are obviously in need of much refinement, are abstract but not without meaningful content, as can be exemplified by the justification of human rights, as we have seen.[29] To recall: The principles underlying the attribution of rights to persons have to be equal for all potential rights-bearers. It would, for example, be unjust to let some people enjoy fundamental rights because of their personhood and deny these rights to others because for them the color of their skin (and not their personhood) is taken to be relevant. In addition, the reasonable – more precisely, the only reasonable – criterion for the attribution of rights is a person's humanity. As a just system of rights has to preserve a relation of equality between the value of this criterion and the distribution of rights, and as all humans are equal in their humanity, only a system of equal rights is consequently a just system of rights. The theories of human rights that imply a diminishing humanity of older people and, correspondingly, a diminishing set of protected rights, which may seem to be at odds with this principle, actually confirm it: These theories simply entertain the (implausible) idea that the humanity of elderly people diminishes, even though the elderly, of course, enjoy the same humanity and consequently the same rights as every other human being – no minor point, as the Covid-19 pandemic has reminded us.

Moreover, fostering the enjoyment of rights is morally good. Accordingly, as has been said before, given the importance of the goods that rights protect and the significance of the rights themselves, the promotion of rights is a prima facie obligation of human solidarity. The theory of the justification of human rights has shown, too, that one of its building blocks is respect for human beings, today widely understood as grounded on their dignity. In sum, justice, the altruism of solidarity and respect for the worth of others take us a long way when trying to identify the normative principles that are key to the justification of human rights.

[28] Cf. e.g. Henry Sidgwick, *The Methods of Ethics*, reprint of the 7th edition, 1907 (Indianapolis, IN: Hackett, 1981), 208; Richard Mervyn Hare, *The Language of Morals* (Oxford: Clarendon Press, 1952), 80 ff.

[29] Another example showing that these principles are not meaningless is that, for example, Rawls' principles of justice can be derived from them: The first principle of universal freedom and the principle of equal access to office are principles of equally distributed goods, namely freedom and offices. The difference principle is a prudential modification of an egalitarian distribution of material goods in a society.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

8.2.3 *Basic Harms, Human Rights and the "Seeds of a Collective Conscience"*

Human rights play an important role in John Mikhail's pioneering work on the mentalist framework of ethics and law.[30] His reconstruction of a set of principles of a "universal jurisprudence" based on common moral precepts that are "the seeds of a collective conscience"[31] includes reflections on human rights that constitute some of the most advanced thought on the topic of our inquiry. His main thesis is that the prohibition of basic wrongs that is an element of the inborn structures of human moral cognition is mirrored in fundamental human rights norms.[32] The universal moral grammar "implies that human beings possess tacit or implicit knowledge of specific, human rights-related norms."[33] The small set of principles that guide human moral judgments include the prohibition of intentional battery, the prohibition of intentional homicide, the rescue principle and the principle of double effect.[34] Of particular importance for human rights are the prohibitions of intentional battery and homicide, which are "lesser included offenses of a wide range of human rights abuses, including murder, extermination, deportation, torture, rape, genocide, and other crimes against humanity."[35] Human rights protect against a set of "core human wrongs" like battery and other torts, "although they by no means exhaust these violations."[36] A "clear conceptual and empirical bridge between moral grammar and human rights can be built" because "[t]hese basic wrongs and, in particular, the tort of harmful battery likewise supply the basic perceptual and cognitive tasks of many influential research programs in the cognitive science of morality."[37] Mikhail's work on the trolley problems is itself an outstanding example of this.[38]

These remarks lead to three questions that can be addressed fruitfully in the light of the results of our analysis.

The first is the question of foundational principles. Our analysis suggests that the principles of justice, altruism and respect that are the normative core of our normative theory of human rights are consistent with Mikhail's analysis but seem to have additional explanatory power. As indicated, the obligation to foster the well-being of others is the flipside of the coin of the prohibition to inflict harm – the minimum you can do for others is not to harm them, we said. It is the normative core of demands for mutual support and solidarity as well. Moreover, human rights are not just about the prohibition of specific wrongs, but also about the allocation of

[30] Mikhail, *Moral Grammar and Human Rights*, 161 ff.

[31] Mikhail, *Moral Grammar and Human Rights*, 170.

[32] Mikhail, *Moral Grammar and Human Rights*, 173 ff.

[33] Mikhail, *Moral Grammar and Human Rights*, 173.

[34] Mikhail, *Moral Grammar and Human Rights*, 180.

[35] Mikhail, *Moral Grammar and Human Rights*, 184.

[36] Mikhail, *Moral Grammar and Human Rights*, 196.

[37] Mikhail, *Moral Grammar and Human Rights*, 196.

[38] Mikhail, *Elements*.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

goods, such as normatively protected spheres of liberty. Such an order of freedom presupposes yardsticks for the justified allocation of such goods – in our analysis, for instance, principles of justice. This is important not least for ascertaining the justified limitations of human rights – for instance, when rights of different persons collide. Moreover, human rights only make sense, as we have seen, if there are reasons to believe that the beings who enjoy them have some kind of worth that justifies protecting these beings' goods. Principles of human worth, today spelled out in human rights ethics and law as human dignity, are therefore likewise key to the understanding of human rights. These principles and their normative consequences, such as the principle of noninstrumentalization, are arguably also key to the solution of the trolley problems.

The second question asks why goods are guaranteed by rights and not some other means in the first place. If normative means to protect human goods are employed (and not just sheer force), why are duties and prohibitions not enough? After all, influential theories argue for the importance of systems of duty and their superiority at least in some respects to systems of rights.[39] Our analysis has suggested (apart from the analytical connection of duties and rights) that normative principles like justice and altruism provide the answer, as these principles give rise to the normative phenomenon of rights that have the important function of normatively empowering people, turning them from patients of others' obligations to subjects of justified claims.

The third issue is the question of construction. Our analysis has suggested that a process of complex construction is necessary to transform concrete, principled moral judgments about moral claims into explicit, critically reflected human rights in ethics and law, a process that spanned thousands of years. This process includes only seemingly straightforward tasks like the generalization, universalization and objectification of the abstract core content of particular moral judgments. Moreover, human rights systems are highly selective. To account for this selectivity and its justification, among other things, a theory of human goods is essential. The importance of a political theory of human rights has been underlined in our inquiry as well. It is essential to include these complex issues in a theoretical account of the link between human moral cognition and the idea of explicit human rights in ethics and law.[40]

8.2.4 *Volitional Consequences of Moral Judgment*

A further important element of an analysis of morality is that human moral judgment has volitional consequences: A moral evaluation does not yield information

[39] Cf. Chapter 1.

[40] A mentalist theory of ethics of law is, thus, not about the mechanical application of unchangeable rules of fully explicit moralities, as Hanno Sauer, *Moral Judgements as Educated Intuitions* (Cambridge, MA: MIT Press, 2017), 41 ff. assumes. It is about clarifying the cognitive resources necessary to form concrete sets of moral principles.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

about a fact of the world like a descriptive proposition, but has prescriptive content.[41] Saying "X is just" is different from saying "X is a table," and an important element of this difference is that a moral judgment tells us how we ought to act. This can have direct volitional consequences: An ought creates a motive for action for agents who are evaluating their options for acting.[42] Their will is *bound* to form certain intentions to act, to use a metaphor for a common introspectively accessible internal state of an agent. Everybody knows how it feels when you come to the conclusion that you *ought* to hand in the smartphone you have found lying on the ground.

If an observer evaluates not his own options for action but the intention or action of another agent, moral judgments still provide information on how one ought to act. For example, they tell the observer how the observed agent ought to act if the observed agent is in a position to act as a person is obligated to act in the particular situation in question – for instance, to hand in the smartphone found, even though the observer himself is not in this position.

A moral judgment can also remain abstract in the sense that there is nobody currently in a position to act accordingly. "One should hand in the valuables of others one has found" is a meaningful statement even if all valuables on the planet are safely in their owners' pockets. Moral judgment then tells us what one ought to do if the conditions for this obligation obtain.

The fact that a normative statement is about an obligation, a permission or a prescription – in short, about a moral ought that provides a motivation to act – does not mean that the obligation necessarily forms the predominant, decisive, let alone only motivation to act. To be sure, human motivation encompasses many other inclinations that have great power and have nothing to do with moral considerations. Human history is to a large extent the history of greed and the pursuit of power, not the history of moral niceties. The claim is thus that only an element – and perhaps a precious element – of human moral motivation derives from moral insight.[43]

The prescriptive content of a moral judgment can – and this is important for our particular topic – constitute a *right*.[44] If an act is just, the agent has an obligation to

[41] Mahlmann, "Ethics," 599 ff.

[42] Cf. for a concise statement Richard Price, *A Review of the Principal Questions in Morals*, ed. David Daiches Raphael (Oxford: Clarendon Press, 1948), 186: "When we are conscious that an action is *fit* to be done, or that it *ought* to be done, it is not conceivable that we can remain *uninfluenced*, or want a *motive* to action" (emphasis in original). On the background debate of motivational externalists and internalists, e.g. Hare, *The Language of Morals*, 20, 30, 169, 197; Richard Mervyn Hare, *Moral Thinking: Its Levels, Methods and Point* (Oxford: Clarendon Press, 1982), 23; David Owen Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989), 39; Gilbert Harman, *Explaining Value: And Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 2000), 30; Philippa Foot, *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 1978), 148; John Leslie Mackie, *Ethics: Inventing Right and Wrong* (London: Penguin Books, 1977), 40.

[43] Cf. Mahlmann, *Rationalismus*, 158 ff.; Mikhail, *Moral Grammar and Human Rights*, 169 ff.

[44] Cf. above the analysis of rights and the connection of duties and (claim-)rights. On the relation of moral judgment and rights, cf. Mikhail, *Elements*, 295 ff.; Mikhail, *Moral Grammar and*

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

act justly and the patient has a right to that action: In the limited time available to comment and set things in intellectual order after an abysmal lecture on the foundation of human rights by a legal theoretician from Zürich, every discussant has the right to the same amount of time because this is a just distribution of this scarce good. The chair of the discussion has the obligation to ensure this fair distribution of time – for example, by restraining the loquacious Swiss theoretician's vacuous responses. The connection between obligation and right holds for an obligation stemming from a duty to benefit somebody, too, unless it is a supererogatory action: Not only is there an obligation to pick up your phone to call an ambulance if somebody in front of you collapses, but the person who has collapsed has a right that you do (at least) this. The obligation to respect others is the correlative of others' right to be respected: Respect is not an act of grace, but means acting in a way to which others have a claim.

8.2.5 *Questions of Metaethics and the World of Moral Emotions*

The principles of justice and altruism that guide reflexive evaluation have cognitive content. Whether or not there is (for example) a relation of equality between patients of actions or between a criterion of distribution and the good distributed in the sense explained above is not felt physically in the same way that cold or heat are, for example, but stems from a complex structural analysis[45] of the evaluated act that eventually predicates a relation of equality or its absence and thus constitutes a judgment with cognitive content. The same is true for other structural elements of the action that play a role for moral evaluation, such as agency, the properties of patients of the action (e.g. whether they are sentient or not) and intentions and their kind (direct or oblique) and object. The presence of these elements is not physically felt either, but is ascertained by a judgment with cognitive content – the obtaining of a direct intention to benefit somebody, for instance.

As we already have seen, this does not mean that the relevance of moral sentiments is diminished. They are evidently crucial to the impact of morality on a human life. There may even be emotions that are "geological upheavals" of moral thought.[46] Nor is the importance denied that certain emotions have for the design of law, not least for the skeptical project of constitutionalism.[47] However, insofar as they are moral emotions, these emotions are not simply emotions relevant for ethics

Human Rights, 160 ff.; Matthias Mahlmann, "The Cognitive Foundations of Law," in *Foundations of Law*, ed. Hubert Rottleuthner (Dordrecht: Springer, 2005), 75 ff.; Mahlmann, *Grundrechtstheorie*, 517 ff.

[45] Cf. the discussion of the trolley cases in Mikhail, *Elements*, 77 ff. for an example of how complex such analysis is.

[46] To use the Proust metaphor of Martha Nussbaum, *Upheavals of Thought* (Cambridge: Cambridge University Press, 2001), 1.

[47] Cf. András Sajó, *Constitutional Sentiments* (New Haven, CT: Yale University Press, 2011), not least on fear. Mortimer Sellers, "Law, Reason, Emotion," in *Law, Reason, Emotion*, ed.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

(like the fear of tyrants) but are the *consequence* of moral judgment with the cognitive content just identified, and thus do not *constitute* moral judgment. The moral evaluation based on the obtaining of the necessary elements of a good or just intention or action is the precondition for the experience of moral emotions – for instance, the unequal treatment of a person who is equal in the relevant respects in question is the precondition for the feeling of indignation about injustice. If there is no unequal treatment of equals, we will not feel moral indignation because of an injustice.

In light of this observation, traditional arguments about the constitutive role played by human moral sentiment in moral evaluation fail to convince because they are not precise enough. Hume wrote:

> If any material circumstance be yet unknown or doubtful, we must first employ our inquiry or intellectual faculties to assure us of it; and must suspend for a time all moral decision or sentiment. While we are ignorant whether a man were aggressor or not, how can we determine whether the person who killed him be criminal or innocent? But after every circumstance, every relation is known, the understanding has no further room to operate, nor any object on which it could employ itself. The approbation or blame, which then ensues, cannot be the work of the judgement, but of the heart; and is not a speculative proposition or affirmation, but an active feeling or sentiment.[48]

Our argument so far indicates that two elements of the structure of moral evaluation are missing from Hume's account. Hume rightly emphasizes the importance of ascertaining the facts of a case to be evaluated. What he does not address, however, is the structural analysis of the intention or action that determines the outcome of the evaluation – for example, the analysis of whether Norbert killed another person with the direct intention to do so for personal gain or whether he did so to defend himself against an aggressor. We have already clarified that this analysis is not a sentiment, but a judgment with cognitive content. The second missing element is the evaluation that ensues on the basis of this analysis – for example, that killing a person for personal gain is (deeply) immoral. This evaluation assigns a deontic status (good/bad, just/unjust) to an intention or action. This is the propositional content of statements such as "Murder is a heinous crime." The moral judgment causes and is accompanied by a sentiment – for instance, the feeling of abhorrence towards murder for gain. The feeling of abhorrence does not constitute the entirety of the evaluation of murder for gain, however.

Interestingly, Hume rightly draws our attention to agency as a structural precondition of evaluating something as morally right or wrong. In Hume's example of the young tree outstripping and eventually overwhelming and killing its parent, the

Mortimer Sellers (Cambridge: Cambridge University Press, 2007), 11 ff. for some general comments.

[48] Hume, "Enquiry Concerning the Principles of Morals," 290.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

tree's growth does not elicit any moral feeling because a precondition of moral evaluation, agency, is not fulfilled.[49] Once again, whether or not a tree is an agent is not felt in the same way as our skin feels a cold breeze. Rather, any such conclusion is a statement about a complex state of affairs constituting agency (responsible, spontaneous beginning of new chains of causation, authorship of intentions to act and of actions, etc.). This cannot be reconciled with simply identifying moral judgment with sentiment. There certainly are moral sentiments, but they depend on such a prior structural analysis of the eliciting situations, including of criteria such as agency, and a consequent moral evaluation.

A central concern for Hume and the many thinkers following in his footsteps is to explain moral motivation. How could a judgment of reason cause a motivation to act? A proposition stating a matter of fact has no motivational impact. "There is a table" is, even if true, motivationally neutral, as we clarified in the preceding section. Only sentiment, it thus seems, can cause people to act. There is, however, a third path beyond the traditional but not exhaustive dichotomy of moral judgment as an act of reason and as sentiment. This third path offers the key to the problem: The motivational effects of moral judgments we highlighted above are crucial in this respect. Moral evaluation has a direct motivational effect, we said, because of its *prescriptive* content. Moral evaluation is not limited to stating the deontic status of an intention or action. It does not merely inform us about this status like a proposition about a fact of the world. Agents do not react to moral judgments in the same way they react to factual propositions. They do not say after a moral judgment, "Oh, this is unjust, how interesting!" in the same way they may observe, following some scrutiny, "Oh, this stone is in fact blue, how interesting!" Moral evaluation gives rises to a moral *ought*, and this ought alone can motivate people to act: It binds the human will, to use this basic metaphor. This is very much an everyday experience. Take the lost object example: You see a wallet that somebody has lost lying in the street. You ask yourself whether you should pocket the money in it. You look around – nobody is watching you. Do you do it? Perhaps you do not, because you have come to the conclusion that taking the money would harm the owner and thus would constitute an immoral act. This moral evaluation does not leave you untouched, because it creates an obligation not to do what is immoral that affects your will and may ultimately make you hand in the wallet. Perhaps you overcome this moral impulse by thinking of something nice you want to buy with the money, but even in this case the moral impulse is not nothing. It was simply not strong enough to outweigh the attraction of the thing you want to buy. Unsurprisingly, the direct motivational force of the moral ought has not escaped the attention of important contributions to moral philosophy.

In discussions about the nature of morality, we sometimes find the idea that a cognitivist approach to moral judgment is wedded to conscious, explicit reasoning

[49] Hume, "Enquiry Concerning the Principles of Morals," 293.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

based on deontological moral rules. As there is ample evidence for spontaneous moral judgments, such intuitions are understood as showing that moral emotivism is correct: These intuitions are emotions, not acts of reason-based cognition. This assumption, too, fails to fathom the nature of morality. Evidently, a structural analysis of the kind discussed – for example, as regards agency – is a quick, largely unconscious process that is not necessarily transparent to the agents themselves. However, this does not mean that the structural analysis is based on a sentiment – it clearly is not, as we just have seen. One thus has to broaden the analysis of moral judgment and include these structural analysis mechanisms, which, while they are largely intuitive, are not simply emotive. Such mechanisms can be made explicit, as Hume did in the case of agency, a process very important for the understanding and practice of morality.

Thus, it turns out that the simple dichotomy between moral judgments as cognitive acts of reason or as expressions of sentiment does not adequately capture the intricacies and rich content of moral judgments, their cognitive, volitional and emotional dimensions that need to be differentiated and accounted for in a theory of human moral cognition.

To repeat: None of this doubts the importance of moral feeling. It is no less than one of the most important elements of human identity. It makes morality the powerful force that it sometimes is in human life. The rich colors of moral emotions are sources of both the beauty and the profound sorrows of human life. These remarks simply intend to clarify what the role of these emotions in moral thought actually is.

A further central function of emotion in moral judgment is to fathom what an action means for the patient of the action. It is a piece of moral heuristics: Without empathy for the experience of victims of racial discrimination, for instance, without the emotional understanding of how it feels to be degraded and humiliated, nobody will be able properly to evaluate the significance of this injustice. If one thinks that being relegated to the back of a bus is a minor issue, the issue that the *Freedom Riders* in the USA were fighting for will remain inexplicable.

8.3 EXPLANATORY LIMITS OF EMOTIVISM

8.3.1 *Ruled by Moral Taste Buds?*

Contemporary contributions to moral philosophy and psychology with a – broadly understood – emotivist background do not call these findings into question. We have already investigated the explanatory power of the mental gizmo thesis. A further influential perspective, the *social intuitionist model* or *moral foundations theory* by Jonathan Haidt, does not change this perception either. This theory argues along emotivist lines that emotional intuitions form the basis for moral judgments. Arguments that appear to be rational are used to justify such emotion-based moral

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

attitudes post hoc. Moral reasoning is not engaged in to critically improve one's own moral point of view.[50] Rather, moral argument is directed at manipulating others – in fact, reason is a "public relation firm" of emotional intuitions for "strategic purposes such as managing reputation, building alliances, and recruiting bystanders to support your side."[51] At the same time, however, rational argument is supposed to have an influence on these emotional reactions, although the theory does not clarify exactly how this influence is to be understood, in particular whether reasoning causes and changes moral judgments or overrides moral judgment.[52]

Jonathan Haidt claims that there are six "moral taste buds of the righteous mind."[53] Haidt intends to expand the ethics of autonomy of WEIRD (Western, educated, industrial, rich and democratic) people (who form the sample of many psychological studies) to include moral systems that are "hierarchical, punitive, and religious" because they include ideas of loyalty, authority and sanctity.[54] The six foundations of morality are the pairs care/harm, fairness/cheating, loyalty/betrayal, authority/subversion, sanctity/degradation and liberty/oppression.[55] Fairness is about the allocation of goods proportionally to merit; demands for equality are matters of nondomination and therefore are based on the liberty taste bud.[56] These pairs are moral modules that form the basis and common ground upon which all of the different human moralities develop. They define the possible contents of any given moral code. It is a misconception, Haidt argues, to focus on just some of these foundations as Western individualism does – for example, overlooking other moralities where loyalty and sanctity are of great importance.

The thrust of the argument is illustrated well by the following example:

> [W]ithin any given culture many moral controversies turn out to involve competing ways to link a behavior to a moral module. Should parents and teachers be allowed to spank children for disobedience? On the left side of the political spectrum, spanking typically triggers judgments of cruelty and oppression. On the right, it is sometimes linked to judgments about proper enforcement of rules, particularly rules about respect for parents and teachers. So even if we all share the same small set of cognitive modules, we can hook actions up to modules in so many ways that we can build conflicting moral matrices on the same small set of foundations.[57]

[50] Haidt, *Righteous Mind*, 50.

[51] Haidt, *Righteous Mind*, 46, 74.

[52] Haidt, *Righteous Mind*, 67 f., admitting that there are good arguments in moral disputes, although he does not explain their nature, content or origin.

[53] Haidt, *Righteous Mind*, 113.

[54] Haidt, *Righteous Mind*, 110, 166.

[55] Haidt, *Righteous Mind*, 125, 155 ff.

[56] Haidt, *Righteous Mind*, 170 ff., 176 ff.

[57] Haidt, *Righteous Mind*, 124.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

This, then, explains why "good people" can disagree about right and wrong (for instance, the right and wrong of spanking) – they are simply operating in different moral matrices.[58]

Haidt uses these moral matrices to explain US politics. If care, fairness and liberty are dominant, one becomes a liberal, if loyalty, authority and sanctity are guiding, a conservative.[59] The dominant influence of these taste buds is genetically fixed.[60] These moral taste buds are explained by their evolutionary functions. The moral foundations have evolved, it is argued, in response to the adaptive challenges of "caring for vulnerable children" (care/harm); of "reaping the rewards of cooperation without getting exploited" (fairness/cheating); of "forming and maintaining coalitions" (loyalty/betrayal); of "forging relationships that will benefit us within social hierarchies" (authority/subversion); and of "the omnivore's dilemma, and then to the broader challenge of living in a world of pathogens and parasites" (sanctity/degradation).[61] There is no clear account of why the liberty/oppression foundation evolved.

Group selection and gene–culture coevolution may have played a role in this evolutionary development, it is argued.[62] The point of morality is to enable cooperation, albeit limited to the groups of which the agent is a member. Parochial altruism is the ultimate frontier of ethics. Oxytocin is important for explaining this within the framework of the theory of group selection: "Oxytocin should bond us to our partners and our groups, so that we can more effectively compete with other groups. It should not bond us to humanity in general. Several studies have validated this prediction."[63] The benefits of cooperation also are why groups evolved that entertain religious beliefs: These beliefs make people better cooperators.[64]

Human rights are explained in this framework: They are post-hoc rationalizations of the intuitions of people with particularly receptive liberty taste buds.[65] The different moral matrices complement each other – liberals are, for instance, right about the necessity of some regulation, while conservatives are right that markets are "miraculous."[66]

[58] Haidt, *Righteous Mind*, 181 ff.

[59] Haidt, *Righteous Mind*, 181, 294 ff.

[60] Haidt, *Righteous Mind*, 312: "People whose genes gave them brains that get a special pleasure from novelty, variety, and diversity, while simultaneously being less sensitive to signs of threat, are predisposed (but not predestined) to become liberals. … People whose genes give them brains with the opposite settings are predisposed, for the same reasons, to resonate with the grand narrative of the right (such as the Reagan narrative)."

[61] Haidt, *Righteous Mind*, 153 f., assuming also an ongoing evolution of the cognitive faculties of modern humans, for which there is no evidence at all.

[62] Haidt, *Righteous Mind*, 189 ff.

[63] Haidt, *Righteous Mind*, 234.

[64] Haidt, *Righteous Mind*, 246 ff.

[65] Haidt, *Righteous Mind*, 175.

[66] Haidt, *Righteous Mind*, 294 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

This psychological theory is descriptive, it is underlined. Haidt notes that if taken as a normative definition of morality, the simple acceptance of loyalty and authority could lead to a defense of political orders such as fascism, which would be given "high marks" as long as they produced high levels of cooperation.[67] For questions of normative theory, therefore, other arguments are necessary – more precisely, what is needed is rule utilitarianism, which should be decisive and guide public policy decisions that take into account the importance of loyalty, authority and sanctity.[68] For the individual, virtue ethics is tentatively endorsed.[69]

The overall vision of morality presented is a narrow one: "Parochial love – love within groups – amplified by similarity, a sense of shared fate, and the suppression of free riders, may be the most we can accomplish."[70]

8.3.2 *A Testing Case: Corporal Punishment – A Question of Taste?*

Spanking is an interesting example for assessing moral foundations theory in the context of human rights, because the corporal punishment of children is a standard human rights issue and in a leading decision of the ECtHR, for instance, was declared to violate Art. 3 ECHR on inhuman or degrading treatment or punishment.[71]

Was this decision just an emotional gut reaction on the part of the majority of judges of the ECtHR, who were led by their particular moral taste buds in their interpretation of Art. 3 ECHR, while the dissenting judge arguing for the permissibility of this kind of sanction simply used other taste buds?

This example reveals a major shortcoming of moral foundations theory quite clearly: It does not distinguish between factual, traditional morality and critically reflected ethics. One major point of moral philosophy is to critically investigate certain moral issues that may turn out not to stand the test of such critical scrutiny, such as the morality of loyalty to fellow slaveholders or respect for the sanctity of moral rules subjugating women – or the fairness of the physical punishment of children. The effects of this critical scrutiny can make history – as the abolition of slavery and the women's liberation movement illustrate. Only critical reflection on morality, based on considered judgments, will tell us what human morality ultimately is about, while factual, traditional morality, by contrast, will tell us about precisely those influences that skew moral judgment – say, racist ideas, misogyny or an authoritarian tradition of child-rearing. Studying a traditional morality that sanctions violence against children as a good starting point for understanding moral cognition is comparable (with a dose of exaggeration) to studying mathematical cognition by looking at the belief that 79 + 86 = 164. Such studies may tell us a lot

[67] Haidt, *Righteous Mind*, 271 f.
[68] Haidt, *Righteous Mind*, 272.
[69] Haidt, *Righteous Mind*, 272 n. 68.
[70] Haidt, *Righteous Mind*, 245.
[71] ECtHR, Tyrer v UK, Judgement of March 15, 1978, appl. no. 5856/72.

about the factors influencing counting, such as a lack of attention, but they are not key to the psychological foundations of mathematics.

Let us take a closer look at the example of corporal punishment to better understand what the further problems of the moral foundations theory may be. There is much to be said about this example, particularly in relation to children. Thinking about the effects of such punishment on children, its educational merit, brutalizing consequences and so on, is an important exercise if one wants to do justice to children. None of this means simply to rationalize emotional intuitive reactions. Rather, this exercise clarifies the issues – in particular, whether corporal punishment does any good or simply harms children, as very many people (rightly) assume today, and whether it should come with corresponding legal consequences. Beyond these important arguments, the ultimate moral evaluation relies on the elements of the action identified above – for instance, the kind of intentions of the agent of corporal punishment. If the intention is to gratify the cruel instincts of the punishing adult, the evaluation will differ from the evaluation of the ill-advised but well-meaning intention of a loving parent who uses corporal punishment as means of education, even though in the latter case, too, it is ultimately an unjustified harm done to children.

This search for moral understanding is the daily business of any responsible human being. It is thus analytically unconvincing to regard morality simply as a manipulation device to strategically influence others, not as a guide to one's own actions. One distinguishing feature of moral judgment is that is aspires to correctness based on reasons – and there are certainly good reasons for certain forms of action. This is what motivates the painstaking efforts of many human beings who honestly want to get it right. Interestingly, the social intuitionist model underlines this itself because it includes reflection in its reconstruction of morality, albeit ambivalently and without clarifying its precise role. The fact that it refers to rule utilitarianism as the ultimate yardstick for the evaluation of social policy and as a bar to "high marks" being awarded to fascism bears witness to this. By trying to limit the harmful conclusions that may be drawn from its arguments by reference to the principles of an ethical maxim such as rule utilitarianism, moral foundations theory denies intuitions generated by moral taste buds the ultimate rule that it appears to defend. In this context, we can note once again that utilitarianism itself is based on notions of equality that are not the consequence, but the normative precondition of utilitarian evaluation. Utilitarianism takes us right back to moral principles of justice and offers no escape route from them, as we have seen. Moral foundations theory's inconsistent impression is only deepened by making virtue ethics a yardstick for individual action. Virtue ethics is, like all ethical systems, contested. However, it does not endorse the rule of emotional intuitions as the core of ethics.[72]

[72] Cf. for a concise statement, Rosalind Hursthouse, *On Virtue Ethics* (Oxford: Oxford University Press, 1999).

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

These observations are important for assessing the merits of the six foundations of morality. They point to further deficiencies in the analysis of morality: Care and harm, fairness and betrayal point to notions of justice and altruism, albeit not very precisely, as betrayal, for instance, is simply one form of unfairness. The relationship of proportionality and equality is analytically misunderstood, too: If the criterion of distribution is equally fulfilled in two agents, equal treatment is the consequence of principles of proportional equality, as we have seen. Therefore, for example, the equality of rights is a consequence of proportional equality: The humanity of humans as a criterion for the enjoyment of rights is simply the same for each person.

Liberty is a central good for human beings, to be sure. Moral foundations theory, however, does not discuss the normative principles that are relevant in balancing the liberty of one with the liberty of all and that may even create rights. The reference to a "liberty taste bud" can be seen as a metaphor that can be reconciled with (but is no substitute for) the theory of liberty as a human good sketched above. However, this reference does not answer the problem faced by any normative theory of liberty that has been at the heart of many debates since antiquity, namely the question of who justifiably enjoys which liberty and to which degree in relation to others.

Another example: Loyalty, authority and sanctity are only secondary virtues. Loyalty and respect for authority are morally justifiable if they serve the good of people; loyalty to dictators and reference of their authority is not a virtue but a vice, as Haidt himself admits, realizing that its content may cause "high marks" to be awarded to fascism if it manages to achieve social cooperation. Sanctity is a difficult concept, too. Respect for the sanctity of a certain conception of marriage may be bad news for gay couples. Here, too, other moral principles, such as respect for the autonomy of other human beings and their dignity, need to be considered.

Moreover, the theory does not engage in any detail with what basic notions of justice or altruism actually entail – for example, as to the intentions of the agents, the effects of actions, means/ends distinctions and so forth, as roughly outlined above.[73]

[73] Cf. for experiments providing empirical evidence that such distinctions and relations are part of the mental representations that underlie moral judgments, Sydney Levine, Alan M. Leslie and John Mikhail, "The Mental Representation of Human Action," *Cognitive Science* 42, no. 4 (2018): 1229 ff. The tools chosen to clarify the structure of these representations are action trees, Mikhail, *Elements*, 125 ff. The authors rightly underline that these findings pose a "challenge to those researchers who either ignore the problem of how moral intuitions arise from eliciting situations … or who uncritically assume that the mental representations of human action underlying moral judgement are exceedingly simple and can be adequately described in terms of heuristics and biases," ibid. 1259, referring to Jonathan Haidt, "The Emotional Dog and Its Rational Tail: A Socialist Intuitionist Approach to Moral Judgement," *Psychological Review* 108, no. 4 (2001): 814 ff. and Sunstein, "Moral Heuristics," 531–41. Hugo Mercier and Dan Sperber, *The Enigma of Reason* (Cambridge, MA: Harvard University Press, 2017), 299 ff., endorse Haidt's model but highlight the effects of deliberation in the case of abolitionism and other cases of moral evaluation without, however, analyzing the structure of moral argument and without specifying the normative principles that have the power to convince.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

Finally, as a last point: The examples of political differences explained by the different operation of the moral taste buds are loaded with prudential arguments, such as arguments about the question of when regulation actually works or whether and under which conditions markets deliver good results. Regulatory choices or the design of markets evidently have ethical implications. But discussing such prudential arguments as part of the core normative machinery of morality does not help to clarify what morality is about.

This notwithstanding, these analytical shortcomings teach us a constructive lesson: They show that the six categories of moral foundations theory do not adequately capture the normative principles that are central for ethical systems. In light of its critical discussion, the principles of justice, altruism and respect continue to seem valuable candidates for the foundational elements of ethics that a psychological theory needs to account for.

The evolutionary framework of moral foundations theory suffers from a failure to present any evidence whatsoever that the evolutionary story it recounts is actually true. It remains one of those "just-so" stories that are falsely taken as valid evolutionary theory-building.[74] Moreover, as indicated above, the first step for any convincing evolutionary theory of morality is to construct a theory of morality that possesses sufficient analytical precision. As this is lacking, Haidt's evolutionary theory is unable to get off the ground.

Moreover, as we have seen, the idea of human rights has highly complex roots in history and human normative reflection. To reduce it to the effects of a liberty taste bud and the post-hoc rationalization of its operations seems not to do full justice to these findings.

The social intuitionist model illustrates a danger already identified previously: When the idea that corporal punishment is admissible, for instance, is interpreted as the expression of certain moral taste buds that simply are different from the taste buds used to criticize this form of punishment, the criticism of corporal punishment loses its ground because there is no reason to prefer the one taste over the other. The theory tries to deal with this problem by introducing rule utilitarianism as a normative yardstick for public policy and virtue ethics for individual acting, contradicting its emotivist message in doing so. Nevertheless, ideas of loyalty to in-groups, respect for authority without questioning its legitimacy and reverence for sanctity without inquiring into its origin and content are understood as expressions of moral judgment on par with any other moral code, in particular one based on human autonomy. In this way, psychological theory may serve to shield unjustified moral precepts against criticism by presenting them as products of a genetically fixed and thus unchangeable natural morality.

In line with various theorists of morality, moral foundations theory argues that only a morality of "love within groups," a "parochial altruism" is the horizon of

[74] Cf. above on evolutionary theory. Interestingly, Haidt criticizes just-so stories – only to then develop one himself, cf. Haidt, *Righteous Mind*, 122 f.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

human moral possibilities. Concern for all humanity is simply too much for human beings. The human rights project, which is a real practice, not an ephemeral dream, indicates that these theses do not fathom what morality is really about.

8.3.3 *Sentimental Rules*

In another very interesting approach, Shaun Nichols argues for a sentimentalist picture of morality. It takes as its starting point the body of data suggesting that children from an early age distinguish between conventional and nonconventional moral rules.[75] The former can be changed by authorities, the latter not. The former deal with prudential arrangements, the latter are centered on the avoidance of harm. The latter are more serious than the former. The latter can be generalized over situations and contexts, the former less so.[76] Moreover, moral objectivism is the psychological default position, Nichols argues.[77]

Nichols claims that moral rules are based on moral judgment, which he argues has two components: information about normative violations and a noncognitive response. The former is provided by rules proscribing certain behavior, while the latter consists of specific emotions. Only those rules relating to behavior that triggers reactive concern and distress cues are nonconventional moral rules.[78] It is the particular emotional engagement of the evaluating person that turns them into moral rules. These emotions are also the reason why these rules are able to survive historical change and development.[79]

As we saw above, the mental gizmo theory already claimed that emotions explain moral judgment such as that reached in the footbridge case, namely that it is impermissible to kill the bystander to save five lives. The sentimental rules approach goes further than the mental gizmo theory, however, in claiming that this is not only because of the emotionally salient features of the act. Rules on the prohibition to perform this act are also important. However, these rules need to be backed by emotions in order to yield the impermissibility judgments observed in the footbridge case.[80] A third factor influencing moral judgment consists of cost–benefit

[75] Shaun Nichols, *Sentimental Rules: On the Natural Foundations of Moral Judgement* (Oxford: Oxford University Press, 2004), 6.

[76] Larry P. Nucci, Elliot Turiel and Gloria Encarnacion-Gawrych, "Children's Social Interaction and Social Concepts," *Journal of Cross-Cultural Psychology* 14, no. 4 (1983): 469 ff.; Judith G. Smetana and Judith L. Braeges, "The Development of Toddlers' Moral and Conventional Judgement," *Merril-Palmer Quarterly* 36, no. 3 (1990): 329 ff.

[77] Nichols, *Sentimental Rules*, 166 ff.

[78] Nichols, *Sentimental Rules*, 187. Norms on what is disgusting (e.g. spitting in a glass one drinks from) are also nonconventional because of emotional reactions to the disgusting action.

[79] Shaun Nichols, "On the Genealogy of Norms: A Case for the Rule of Emotion in Cultural Evolution," *Philosophy of Science* 69, no. 2 (2002): 234 ff.; Nichols, *Sentimental Rules*, 16 ff.

[80] Shaun Nichols and Ron Mallon, "Moral Dilemmas and Moral Rules," *Cognition* 100, no. 3 (2006): 530 ff., 540.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

assessments along utilitarian lines.[81] These factors interact in a complex way, without there being a unified normative theory that captures all moral intuitions.

Metaethically, morality therefore is not objective, but bound to feelings relative to particular groups of people. There is no argument why one should prefer one set of emotions to another. It is not possible to evaluate the emotions determining the content of morality with these very emotions.[82] The illusion of objectivity hides the respondent-dependent relativity of emotionally based sentimental rules. This notwithstanding, the objectivist intuition persists and guides human lives.[83]

This theory faces the same problems already identified as decisive for emotivist accounts of moral judgment: The emotional reaction to some harm is not all there is to moral judgment. A harm may cause considerable unease without it being immoral – to use Nichols' own example, a medically justified operation, for instance. It is only immoral if the intention and the action have certain properties – for instance, the intention to torture the patient and not to heal them. These properties need to be determined based on an analysis of the intention and action, as outlined above in some detail. This analysis then elicits the moral evaluation. This, in turn, gives rise to moral feelings, which are different and need to be distinguished from feelings such as unease at witnessing an operation. Feeling nauseous at the sight of blood during such an operation is not already a moral feeling, while feeling abhorrence because of the wrongness of an attempt to torture a person is. The evaluation of an action as morally wrong is, however, the precondition and cause of these moral feelings. Like an Escher cube, the same action can change its moral nature, depending on what evaluation the analysis of the action's intentions and effects yields.

Nichols developed a theory of moral learning that adds a further qualification, underlining another influence that reason has on moral judgment. It is best considered in the framework of the discussion in Section 8.5 on the acquisition of moral knowledge.

8.4 EXPLAINING MORAL DISAGREEMENT

Given the great variety of moral opinions visible both today and in history, any theory of moral cognition will need to formulate a theory of moral disagreement as part of its explanatory enterprise. Moral disagreement is yet another traditional and vast topic of practical philosophy, and empirical work has been carried out in this field, too.[84] It is sometimes argued that the mere existence of moral disagreement

[81] Nichols and Mallon, "Moral Dilemmas," 530 ff., 540.

[82] Nichols, *Sentimental Rules*, 188.

[83] Nichols, *Sentimental Rules*, 197 f.

[84] Cf. the attempts to explain different reactions, for example, to insults and other issues through different "cultures of honor" in the north and south of the USA, Richard E. Nisbett and Dov Cohen, *Culture of Honor: The Psychology of Violence in the South* (Boulder, CO: Westview

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

already proves moral relativism.[85] In this respect, however, conclusions should not be drawn too rashly: "Establishing the best explanation of stubborn ethical disagreement requires understanding all the possible origins of these conflicting beliefs and all the possible resources that might resolve the conflict – no quick or easy job."[86] There may be ways to account for moral disagreement, even radical disagreement, with substantial explanatory power without any implications about irreconcilable foundational moral principles.

A first task, thus, is to determine precisely what kinds of moral disagreement actually exist. This is far from obvious. Some studies on the supposedly different moral orientations of different cultures, for instance, under critical scrutiny turn out to have overlooked important communalities.[87] The second task is to see what the causal factors for apparently different moral beliefs are and what they teach us about the structure of moral judgment. Not just law, but morality, too, is often based on express moral rules. These rules can embody conceptions of what is good and just that do not withstand critical scrutiny. It is therefore necessary to distinguish between traditional, prereflective social moralities and a critical, reflective ethics in the context of analytical theories of moral disagreement, too. The conditions for the successful criticism of traditional moral practices are key to understanding the principles underlying human moral deliberation and judgment that may in the end lead to the abandonment of such practices.

One factor of considerable importance in explaining the existence of moral disagreement is disagreement about the nonmoral preconditions of moral judgment. This includes understanding what an action, practice or institution means for other persons. These assumptions may even lead to denying certain persons moral status: As we have seen, a central question during the conquest of the Americas in the sixteenth century was whether the indigenous people of the Americas were actually fully human or some other kind of creature. The identification of the proper objects of evaluation likewise shows the relevance that these factual questions hold for moral judgment: For instance, the institutions of the state have to be regarded as products of human volition and action and not as unchangeable elements of the makeup of the world to be evaluated on the basis of principles of justice.

Other factors that rightly play a prominent role in this debate are the influence of interests and the impact of ideological constructions, both the source of extremely

Press, 1996); Richard E. Nisbett, *The Geography of Thought: How Asians and Westerns Think Differently … and Why* (New York: Free Press, 2003); Haidt, *Righteous Mind*, 11 ff. The disagreement can encompass the domain of morality as such, ibid. 14 ff.

[85] Mackie, *Ethics: Inventing Right and Wrong*, 36.

[86] Griffin, *On Human Rights*, 129 ff.

[87] Cf. for instance, for some comparative research, John Mikhail, "Is the Prohibition of Homicide Universal? Evidence from Comparative Criminal Law," *Brooklyn Law Review* 75, no. 2 (2009): 497 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

powerful emotions and motives of action. An entirely fantastic ideology like National Socialism, for instance, was capable to motivate people to commit mass murder before leading them to their own deaths and the destruction of their country. Taking these factors into account may already reduce the cases of real moral disagreement about basic principles of morality considerably. Take the (important) example of the rights of women: The denial of equal rights of women was (and is) partly based on wrong factual anthropological assumptions, such as the idea that women lack the capacity for autonomous self-determination and rationality and therefore have to be guided by men.[88] The suffering of women who were denied an equal part in social life because of these assumptions about their interests and capabilities needed to be understood. This required massive cultural and political efforts, sustained over generations and still ongoing today. The interests of men in comfortable structures of domination were an evident further factor: “When, however, we ask why the existence of one-half the species should be merely ancillary to that of the other – why each woman should be a mere appendage to a man, allowed to have no interest of her own, that there may be nothing to compete in her mind with his interests and his pleasure; the only reason which can be given is, that men like it.”[89] Ideological constructs, partly in religious garb, about the place of women in the world buttressed these social structures, too. Such factors continue to play an important political role today, despite the progress made. If these false assumptions about the nature of women, their experience of repressive patriarchic structures and the power of interests and ideologies lose their influence, apparently irreconcilable moral disagreements can disappear quickly and the equal rights of women appear as an evident truism even across cultural boundaries (as indeed they should).

Another important issue is the process of clarifying ethical concepts – for instance, that responsibility for actions depends on the internal state of the agent, an intention to act, an insight of major importance for the development of criminal law. Some forms of human behavior may be wrongly moralized or wrongly demoralized, often as a consequence of false factual assumptions, interests and ideologies. Take the example of LGBTIQ* rights: If one understands that the only consequence of preventing consensual intimate relations between same-sex partners is to cause suffering while not promoting any discernable good enjoyed by human beings, traditional moralities about the wickedness of same-sex partnerships quickly crumble, as we have witnessed in the last thirty years or so. If one understands that

[88] For another example, Griffin, *On Human Rights*, 244, on the exclusion of some people from democracy: “These exclusions were supported by largely factual beliefs: that certain races were of lower intelligence, that they were child-like, that women were not interested in politics, that they were already adequately represented by their husbands, and so on. Once the falsity or irrelevance of these beliefs was recognized, these excluded groups had to be admitted into the class of ‘people’ referred to in the defining formula ‘the people rule’.”

[89] Taylor Mill, “Enfranchisement of Women,” 62.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

harassing people because of their sexual orientation means doing an unjustified harm to them, then refraining from mistreating them in this way becomes a moral imperative and may even lead to a prohibition by law of certain qualified forms of this kind of behavior. Reflective ethics helps us to achieve consistency – if autonomy, for instance, is a key value, autonomous decisions about the persons one loves should be protected robustly. Ethical reflection of this kind can expand the circle of persons included in moral deliberation: If all persons count equally, the interests of future generations of human beings and thus questions of intergenerational justice should be factored into our debates about the appropriate measures to combat climate change, as courts around the world have now started to do explicitly.[90] If sentient beings are objects of moral concern, then ethical and legal protections for nonhuman animals are required. Moral values and standards of virtuous acting may not stand the test of critical thought – that it is imperative to fight a duel with somebody who has insulted one has ceased to be a convincing interpretation of proper, self-respecting behavior. A further example illustrating the importance of critical normative thinking is the scrutiny of theories of morality themselves: The critique, for instance, of the (as we have seen) false idea that human rights are products of the post-hoc rationalization of emotional reactions to certain stimuli will prevent the conclusion that human rights should be made irrelevant.

Finally, there are real moral dilemmas to which no easy solution is available – for instance, in the case of conflicting duties, say, the duty to save lives in a pandemic and the duty to prevent the domestic violence increased by lockdowns. In sum, there are a plethora of factors helping to explain different moral points of view without taking recourse to substantially and unchangeably different underlying moral conceptions.[91] There is thus reason to believe that under the surface of apparently insurmountable moral disagreement, there may be a deep, shared structure of common moral principles.[92]

[90] Cf. for example BVerfG, Order of the First Senate of March 24, 2021, 1 BvR 2656/18. The challenges for human rights caused by climate change include, first, the consequences that are related to climate change for the protection of classical human rights positions – for instance, state repression of climate activists or climate migration. Moreover, the question arises as to how to "climatize" human rights, by developing new doctrinal tools as to the bearer of rights (rights of nature?), the addressee or causality and responsibility, cf. for more details César Rodríguez-Garavito, "Human Rights 2030: Existential Challenges and a New Paradigm for the Human Rights Field," in *The Struggle for Human Rights: Essays in Honour of Philip Alston*, eds. Nehal Bhuta et al. (Oxford: Oxford University Press, 2021), 328 ff., 342 ff.; César Rodríguez-Garavito (ed.), *Litigating the Climate Emergency. How Human Rights, Courts, and Legal Mobilization Can Bolster Climate Action* (Cambridge: Cambridge University Press, 2022).

[91] Buchanan and Powell, *The Evolution of Moral Progress*, 54 ff., present a very helpful theory of moral progress that considers comparable factors (better compliance with moral norms; better moral concepts; better understanding of virtues; better moral motivation; better moral reasoning; proper demoralization and moralization; better understanding of moral standing and moral statuses; improvements in the understanding of the nature of morality; better understanding of justice).

[92] Cf. Mahlmann, "Ethics," 593 ff.; Mikhail, *Moral Grammar and Human Rights*, 170 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

8.5 THE DEVELOPMENT OF MORAL COGNITION

8.5.1 *How Do We Learn to Be Moral?*

One question that remains to be explored is the ontogenetic origin of moral cognition. The development of moral cognition forms the subject of landmark debates in moral psychology.[93] Our analysis thus far suggests that central issues here include – among other factors – the acquisition of the cognitive domain of morality, the restrictive principles that determine the possible objects of evaluation, the material principles of morality discussed, the prescriptive, volitional effects of moral judgments, the moral ought, including necessary connections between duties and rights, and the emotional consequences of moral experience. Such phenomena could be constructed and acquired by secondary learning processes (instruction, repetition, role-taking, peer pressure, sanctioning, etc.), as many normative principles are (e.g. the intricacies of Swiss law on unjust enrichment), or alternatively they could be at least in part the product of the unfolding of innate cognitive structures triggered by experience. The latter is the way a mentalist approach to ethics would approach the issue.

It is important to emphasize that the importance of social and cultural influences on ethics and law is not denied in the latter case: As noted, the freedom of the press, for instance, as a legal norm and the underlying principles of political morality presuppose the cultural achievement (a late achievement with plenty of preconditions) of the press and, in addition, the experience of its suppression even by democratic governments, among many other things. "The freedom of the press is protected" is certainly not an inborn principle of human moral cognition. As we saw in our survey of the history and the theory of justification of human rights, fundamental moral judgments about the wrongness and justice of certain intentions and actions are the seeds of the idea of human rights. The real question is therefore: What kind of mind do you need to possess in order to develop such an idea? What is so special about the human mind that only humans and no other organism developed a concept like human rights? What are the cognitive preconditions for starting the long cultural process leading, after many thousands of years of human cultural and social development, to the idea that one can not only wish for or have an interest in, but in fact enjoy a *right* to a free press? None? Is simply being smart enough? Or does one need to have certain specific conceptual tools with which to build a system of rights?

[93] E.g. of the work of Jean Piaget, *Le jugement moral chez l'enfant: perspectives piagétiennes* (Paris: F. Alcan, 1932) or Lawrence Kohlberg, *Essays on Moral Development, Vol. I: The Philosophy of Moral Development: Moral Stages and the Idea of Justice* (San Francisco, CA: Harper & Row, 1981) and Lawrence Kohlberg, *Essays on Moral Development, Vol. II: The Psychology of Moral Development: The Nature and Validity of Moral Stages* (San Francisco, CA: Harper & Row, 1984).

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

The simple observation already recalled above that any child acquires a differentiated set of moral concepts and categories whereas no nonhuman animal acquires any of the same, even if it is raised in the same environment, seems to indicate strongly that very different cognitive abilities are in place. Some of the empirical work on this difference between humans and nonhuman animals has been discussed in Chapter 7. To repeat: The question therefore is not whether there are species-specific inborn cognitive structures enabling the formation of moral precepts, but what exactly these cognitive structures are.

8.5.2 *The Moral World of Infants and Toddlers*

Jean Piaget pioneered the idea that children apply complex moral ideas and concepts, not least when they play, as illustrated by his fascinating analysis of children playing with marbles.[94] Recent years have produced many more highly creative studies on the moral psychology of children, increasingly focusing on young children. We have already encountered some important examples in our inquiry. As we have seen, these studies operate with different theoretical background assumptions and sometimes yield contradictory results. Some research is constrained by theoretical and conceptual problems, limits of the design of experiments and contestable interpretations of their results. One important problem is the lack of a sufficiently fine-grained analytical theory of moral judgment that does not commit, for instance, the category error of conflating moral judgment with simple preferences and aversions. The thrust of this research is, however, that children operate in a richly textured moral world from very early on. This fits well with research into other cognitive domains that also indicates that human beings are not born blank slates.[95]

A classical set of experiments concerns the moral–conventional distinction, for instance: As mentioned above, children distinguish between conventional norms and moral, nonconventional norms.[96] Other experiments suggest that toddlers by the age of eighteen months and even earlier exhibit spontaneous helping behavior towards others.[97] They engage in so-called paternalistic helping: Children want to accommodate the well-being of others, not just their wishes.[98]

Children act with an intrinsic moral motivation: They also help when nobody is watching or when others do not know that they are being helped.[99] External rewards

[94] Piaget, *Jugement moral*.

[95] Cf. Stephen Pinker, *The Blank Slate* (London: Allen Lane, 2002).

[96] Nucci, Turiel and Encarnacion-Gawrych, "Children's Social Interaction," 469 ff.; Smetana and Braeges, "Development of Toddlers'," 329 ff.

[97] Felix Warneken and Michael Tomasello, "Altruistic Helping in Human Infants and Young Chimpanzees," *Science* 311, no. 5765 (2006): 1301–3; Felix Warneken and Michael Tomasello, "Helping and Cooperation at 14 Months of Age," *Infancy* 11, no. 3 (2007): 271 ff.

[98] Tomasello, *Becoming Human*, 226 ff.

[99] Tomasello, *Becoming Human*, 226.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

even seem to undermine their intrinsic motivation.[100] Helping somebody themselves and seeing somebody being helped render them equally content.[101] There are studies on extended sympathy – for instance, for the victims of some harmful action after the deed.[102]

Studies indicate that – unlike great apes – small children act according to principles of distributive egalitarian justice, particularly in collaborative contexts.[103] While there are studies suggesting that children limit fairness to in-group members,[104] there is also substantial evidence that three-year-olds apply egalitarian principles universally.[105] Such intuitions extend to procedures of distribution.[106] Three- to five-year-olds take need or merit as criteria of distributive justice. They protect the entitlements of others, create social norms, employ normative categories and engage in third-party punishment.[107] Intentions are a central factor in this punishment.[108] There are studies on the specific moral feelings of guilt and shame – for instance, arguing that three-year-old children feel guilt over the harm they inflict, not just sympathy for the victims of this harm.[109]

According to various studies, it is plausible to assume that infants already operate with at least some kind of proto-morality. Twelve-month-olds categorize actions on the basis of their social valency.[110] Studies on social evaluation indicate that six- to ten-month-old infants base their evaluation of another individual as appealing or aversive on this individual's actions towards others: For instance, they like individuals better who help others and who act more cooperatively to facilitate the achievement of these others' aims than those who hinder the achievement of others' goals. They also like helpers better than neutral individuals, and the latter better than hinderers.[111] These experiments concern actors who are unknown to the infants

[100] Tomasello, *Becoming Human*, 226.

[101] Tomasello, *Becoming Human*, 226.

[102] Tomasello, *Becoming Human*, 227.

[103] Cf. Tomasello, *Becoming Human*, 229 ff., 245: Children have an aversion to disadvantageous and advantageous inequity and an equality bias.

[104] Ernst Fehr, Helen Bernhard and Bettina Rockenbach, "Egalitarianism in Young Children," *Nature* 454 (2008): 1079 ff.

[105] Tomasello, *Becoming Human*, 252, 258.

[106] Tomasello, *Becoming Human*, 240.

[107] Tomasello, *Becoming Human*, 257, 259, 264.

[108] Tomasello, *Becoming Human*, 264.

[109] Tomasello, *Becoming Human*, 282.

[110] David Premack and Ann James Premack, "Infants Attribute Value to the Goal Directed Actions of Self-Propelled Objects," *Journal of Cognitive Neuroscience* 9, no. 6 (1997): 848 ff.

[111] J. Kiley Hamlin, Karen Wynn and Paul Bloom, "Social Evaluation by Preverbal Infants," *Nature* 450 (2007): 557 ff.; J. Kiley Hamlin and Karen Wynn, "Young Infants Prefer Prosocial to Antisocial Others," *Cognitive Development* 26, no. 1 (2011): 30 ff.; Julia W. Van de Vondervoort and J. Kiley Hamlin, "Evidence for Intuitive Morality: Preverbal Infants Make Sociomoral Evaluations," *Child Development Perspectives* 10, no. 3 (2016): 143 ff.; Valerie Kuhlmeier, Karen Wynn and Paul Bloom, "Attribution of Dispositional States by 12-Month-Olds," *Psychological Science* 14, no. 5 (2003): 402 ff. The experiments rule out any influence of superficial perceptual factors on judgment.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

and actions that have no effect on them as observers. This rightly has been identified as an important element of moral judgment: It is not personal experience with the agent that causes the agent to be evaluated a certain way, but the latter's action affecting unrelated others. Eight-month-olds showed a preference for individuals with good intentions, basing their evaluation on mental states of the agent, not outcome.[112] Other research indicates that actions towards inanimate entities do not enter into infants' evaluation of intentional agents.[113] Knowledge of an agent's goals is relevant for ten-month-olds' evaluations: Only an agent who knowingly helps another is preferred to other agents. It is not far-fetched to draw the conclusion from this research on infants that the capacity to evaluate others based on their behavior towards third parties is universal and unlearned, given their early age. It should be noted once again, however, that a social preference is not the same as a moral evaluation of an intention or action. The former is at best indirect evidence for the latter.

Many of these studies understand these mental capacities as biological adaptations for cooperation and interpret them as tools to identify free riders, cooperators and reciprocators. As we have seen already, this biological interpretation is too narrow and overlooks the many forms of cooperation that are both possible and entirely reconcilable with the constraints of a plausible evolutionary trajectory.

8.6 POVERTY OF STIMULUS AND THE DEVELOPMENT OF A MORAL POINT OF VIEW

These studies on child psychology tell us something interesting about the early cognitive stages of human mental development. The younger the children are for whom plausible evidence shows that they operate within a moral cognitive domain, the less plausible it becomes that these children were born as blank slates in moral terms. There are some studies on the lack of cultural variation in crucial areas of the development of moral cognition that point in the same direction.[114] This kind of evidence supports the hypothesis that foundational elements of morality are part of the natural cognitive endowment of human beings that matures during childhood, just like other cognitive faculties. But even if clear traces of a mature system of morality were not in place at an early age, the possibility of such a natural endowment could not be excluded, for it is entirely conceivable that a specific mental faculty matures only later in childhood. The fact that puberty (including its cognitive components) occurs in the second decade of human beings' lives does not speak against it being based on inborn properties of the human species.

[112] J. Kiley Hamlin, "Failed Attempt to Help and Harm: Intention versus Outcome in Preverbal Infants' Social Evaluations," *Cognition* 128, no. 3 (2013): 451 ff.

[113] Amanda L. Woodward, "Infants Selectively Encode the Goal of an Actor's Reach," *Cognition* 69, no. 1 (1998): 1 ff.

[114] Cf. for a review Tomasello, *Becoming Human*, 232 (helping), 241 (sharing), 268 (justice).

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

The decisive argument is therefore the *poverty of stimulus argument*: If the input of an organism's experience is not sufficient to generate a certain cognitive ability, at least some of the cognitive structures underlying this ability must be inborn.[115] This argument holds for any cognitive ability of any organism. To take a fascinating example that illustrates the structure of this argument:[116] Honeybees communicate the location of food through their dance in the hive. They determine the position of food by the position of the sun relative to a known terrain. They can do this even if the sky is overcast, because their circadian clock and their orientation systems enable them to determine the solar ephemeris – that is, the position of the sun relative to the time of the day. Their dance can thus refer to this position even if they were unable to see the sun itself. How do bees acquire this striking faculty? The following experiment gives the answer: Bees were raised in an incubator and foraged only in the late afternoon and had thus no direct experience of the sun's morning position. One day, they were let out to forage in the morning under an overcast sky. They still were able to determine and communicate the position of food relative to the position of the sun they had never observed: They had determined the solar ephemeris and with it the position of the food. As they had no experience of the sun's morning position but were still able to determine it, this ability cannot be learned. The only explanation for bees' acquisition of this ability is an innate cognitive mechanism of orientation operating in conjunction with the circadian clock.

This example illustrates the point of the poverty of stimulus argument and shows what intricate cognitive mechanisms are in place in organisms like bees. It may encourage us to ask seriously what structures the vastly more complex apparatus of human cognition may contain.

If the poverty of stimulus argument is considered in detail in the context of morality, many questions arise. Given the sheer scope of the issue, some hints at the gist of the argument must suffice here. Hume's implicit use of it *avant la lettre* provides a good illustration of its structure:

> This principle, indeed, of precept and education, must so far be owned to have a powerful influence, that it may frequently increase or diminish, beyond their natural standard, the sentiments of approbation or dislike; and may even, in particular instances, create, without any natural principle, a new sentiment of this kind; as is evident in all superstitious practices and observance: But that *all* moral affection or dislike arises from this origin, will never surely be allowed by any judicious enquirer. Had nature made no such distinction, founded on the original constitution of the mind, the words, *honourable* and *shameful*, *lovely* and *odious*,

[115] Cf. Stephen Laurence and Eric Margolis, "The Poverty of the Stimulus Argument," *The British Journal for the Philosophy of Science* 52, no. 2 (2001): 217 ff.; Mahlmann, *Rationalismus*, 74 ff.; Mikhail, *Elements*, 70 ff.

[116] Cf. Charles R. Gallistel, "Learning Organs," in *Chomsky Notebook*, eds. Jean Bricmont and Julie Franck (New York: Columbia University Press, 2007), 193, 197 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

> *noble* and *despicable*, had never had any place in any language; nor could politicians, had they invented these terms, ever have been able to render them intelligible, or make them convey any idea to the audience.[117]

Hume's point is that basic moral concepts are not learned from scratch, they are the *precondition* for moral learning. If one takes as examples a constitutive element of human morality like the moral cognitive space, some aspects of the principles governing moral judgment, the concept of ought and moral emotions like guilt or shame, the argument can be fleshed out somewhat more: As to the moral cognitive space, the poverty of stimulus argument asks whether a child can acquire this particular dimension of its perception of the world *de novo*, through the example of peers, instruction, imitation and so forth. In order to answer this question, one has to imagine a child who has no idea whatsoever of the qualitative content of the moral perspective: It sees the world only as a world of facts and events, including actions of other persons, but not in the very different colors of moral evaluation. How could a child possibly step from this perception of the world into the very different kind of world possessing a moral dimension? Quite aside from the fact that there is no such thing as instruction about this matter in real educational settings, even direct instruction would not help: How could a child understand what is meant by explanations of the moral dimension of human actions if it did not have access to this category of thinking? It would be like explaining the scent of chocolate to somebody who cannot smell.

The same is true for the concept of ought or moral emotions. What is a child to do with the explanation that an ought binds the will? How can the child be enlightened about what is meant by this if it has no access to this experience? Moreover, how is such a verbal definition to be turned into the actual subjective experience of being under an obligation?

Other mechanisms often used to explain the development of a moral orientation do not help either. Sanctions as such do not create the experience of an inner ought, only the experience of outward compulsion. A child with no access to the experience of moral obligation will feel a sanction as a harm inflicted upon it and a prudential reason not to show certain behaviors in order to avoid this harm, but will not feel any inner obligation not to do certain things. Sanctions enforce obligations but cannot trigger the subjective experience of a moral obligation.

Another example is role-taking. Sometimes it is argued that learning to take the perspective of another person leads to the understanding of the idea of a moral ought.[118] Role-taking, however, informs us about the perspective of another person only in a qualified sense: We understand how *we with our capacities* would perceive the world if we were in the other person's place. A person who cannot smell cannot

[117] Hume, "Enquiry Concerning the Principles of Morals," 214 (emphasis in original).

[118] Cf. for instance Tomasello, *Becoming Human*, 281, who argues that role reversal is the origin of conscience.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

understand what scent is just by taking the role of somebody who can smell. The same is true for moral concepts: If, as imagined, a child is a moral blank slate, nothing about this state of affairs will change by taking the role of another person. The child will only imagine what a world without access to the idea of moral obligation looks like from the point of view of the other person.[119]

The case of moral emotions seems even more obvious. How are we to teach a child what the feeling of guilt or shame is like if this child is assumed to have no access to this emotion? Note that this is not about teaching the reasons for feeling guilty or ashamed, but about the emotion itself and its quality. Invoking internalization does not help either. One cannot internalize something to which one has no cognitive access. What the hypothesis of internalization tries to capture in these contexts is in fact the process of some experience triggering the maturation of cognitive capacities that make some cognitive or emotional phenomena accessible to the person in question.[120]

Finally, it is not particularly plausible that a child is ever instructed about the differentiation between direct and oblique intentions for the moral evaluations of seemingly altruistic acts. The same seems true for the prohibition of instrumentalization or the intricacies of proportional justice.

The same kinds of question need to be asked about other elements of an analytical theory of morality, in particular the dependency of moral judgment on agency, the limited class of possible objects of moral evaluation and the foundational relation between rights and the principles of justice and altruism. The concept and normative category of "right" is itself of great interest in this respect, as is what this category entails, namely the intricate web of normative positions sketched above, the necessary connections between claims and duties, privileges and negations of duties, the intentional content of these deontic categories, the

[119] The point is related to the discussion of irreducible subjective experience in the theory of consciousness, cf. Thomas Nagel, "What Is It Like to Be a Bat?" in Thomas Nagel, *Mortal Questions* (Cambridge: Cambridge University Press, 2020), 165–81. The problem illustrated by the thought experiment is how to teach a "moral bat" the subjective experience of morality – a problem that is comparable to the question of how to teach a child the experience of orienting oneself via a sonar system.

[120] Tomasello, *Becoming Human*, 214 posits: "[T]he sense of obligation is basically the internalization of an interpersonal commitment (given an agent who already has a sense of instrumental pressure to do what is needed to attain goals), and guilt is likewise the internalization of an interpersonal process of second-personal protest (given an agent who already engages in executive regulation)." This passage is useful to illustrate the problem: An interpersonal commitment is a normative phenomenon and presupposes that an obligation is a cognitively accessible phenomenon. There is no bridge from "instrumental pressure" to a normative obligation because the two are categorically different. Second-person protest can elicit all kinds of reactions – for instance, sarcasm, contempt, boredom, counterprotest, etc. Second-person protest does not necessarily give birth to feelings of guilt in others. Guilt is simply another primordial category of the moral life of human beings. However, it can be and often is triggered by the recriminations of others, because the agents realize that they have done something morally wrong.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

semantics of obligation, permission and prohibition and the necessary volitional and emotional consequences of moral judgment's implications for rights. In all these cases, one needs to consider carefully what the poverty of stimulus argument may teach us about the acquisition of this complex set of normative positions.

8.7 SENTIMENTAL RULES, RATIONAL RULES?

8.7.1 *The Power of Statistical Learning*

It is instructive to look at an account of moral learning developed by Shaun Nichols that takes seriously the argument made thus far, namely that there is no direct input from parents and peers that explains the development of moral cognition.[121] He underlines that there is a consensus that for some capacities an empiricist account and for other capacities a nativist account is more plausible.[122] His account, already referred to briefly above, complements his moral sentimentalism with a rationalist element[123] and presents an alternative to assuming that concrete moral rules are based on certain inborn structures, claiming that statistical Bayesian learning abilities suffice to infer the content of the existing rules of a community from the given evidence of identified violations of rules.[124] This is how deontological rules arise: "Our hypothesis is that non-utilitarian judgement derives from learning narrow-scope rules, i.e. rules that prohibit intentionally producing an outcome, in a way that approximates Bayesian learning."[125] Nevertheless, there is a nativist element in his theory, or his theory is at least consistent with a nativist approach: Learners have an aptitude for concepts like agent, intention and cause, and the capacity for acquiring rules.[126]

More concretely, Nichols' argument takes its start from the observation that a distinct evaluation of doing on the one hand and allowing on the other is a reality of human moral psychology. This distinction presupposes that humans apply act-based rules, not consequence-based rules. Moral rules are about prohibitions of or prescriptions for action and not about minimizing unwelcome outcomes or maximizing welcome consequences.

Nichols goes on to ask how children can learn that the moral rules that they apply are act-based, not consequence-based.[127] His argument is that they are exposed to act-based input by adults, such as "Don't do X!" or "Do Y!" Through mechanisms of

[121] Shaun Nichols et al., "Rational Learners and Moral Rules," *Mind and Language* 31, no. 5 (2016): 530 ff.; Nichols, *Rational Rules*, 49 f.

[122] Nichols, *Rational Rules*, 20.

[123] Nichols, *Rational Rules*, 10, on rationalism as evidentialism.

[124] Nichols, *Rational Rules*, 57 ff.; for background cf. Fei Xu and Joshua B. Tenenbaum, "Word Learning as Bayesian Inference," *Psychological Review*, 114, no. 2 (2007): 245–72.

[125] Nichols et al., "Rational Learners," 549.

[126] Nichols, *Rational Rules*, 22.

[127] Nichols, *Rational Rules*, 50 ff.

statistical learning, the children acquire act-based rules, rather than consequence-based rules, as experimental evidence confirms in his view. The core statistical mechanism is the size principle: When a learner has to choose between two hypotheses, one of which is a nested subset in the other, it is rational on probabilistic grounds to choose the hypothesis with the smaller scope if all of the evidence available is consistent with this smaller hypothesis. This is because there is a higher likelihood that the smaller hypothesis is in fact correct. It would form a "suspicious coincidence" if all of the evidence falls in the smaller hypothesis while the larger hypothesis is true.[128] Humans are able to perform this statistical operation because they possess substantial statistical learning abilities, Nichols holds.[129]

As the input that learners are exposed to is based on the application of act-based rules (for instance, the command "Don't do X!"), this evidence is consistent with both the hypothesis that act-based rules are applied and the hypothesis that consequence-based rules are applied, the latter hypothesis also including act-based rules. As the first hypothesis is a nested subset of the latter and the evidence that the learners is exposed to is act-based, it is rational to conclude that act-based rules are applied. Act-based rules are thus acquired by the learners. Nichols concludes, albeit somewhat hesitatingly, that this learning process could explain why people apply the principle of double effect – for instance, in the trolley cases.[130] It can also account for the acquisition of parochial moral codes.[131]

Furthermore, there is evidence that humans expect new rules to be act-based, showing a "pronounced prior" in this respect.[132] The reasons for this are overhypotheses in the sense defined by Nelson Goodman – because the experience of rules consists of act-based rules, an overhypothesis is formed that rules tend to be act-based.[133]

Normative systems need a default principle of the normative status of those intentions and acts that are not explicitly prohibited or permitted. *Everything that is not prohibited is allowed* or *everything that is not permitted is prohibited* are such closure principles, for instance.

Nichols argues that learners acquire the one or the other depending on the input received: If they encounter permissions, they conclude that what is not permitted is prohibited. If, however, they encounter prohibitions, they conclude that what is not prohibited is permitted. This conclusion is based on pedagogical sampling: Learners assume that their teacher is using methods of rational and efficient instruction. Based on this principle, it is rational for learners to conclude that if they are only

[128] Nichols, *Rational Rules*, 57 ff., 134.
[129] Nichols, *Rational Rules*, 16 ff.
[130] Nichols, *Rational Rules*, 64 ff., 73.
[131] Nichols, *Rational Rules*, 74 ff.
[132] Nichols, *Rational Rules*, 82 ff., 135.
[133] Nichols, *Rational Rules*, 84 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

presented with prohibitions of certain acts by the instructor, then other acts are permitted, and vice versa.[134]

Nichols also explains perceptions about the universal or relative validity of norms along these lines. If learners encounter widespread consensus, they acquire the idea that the respective norms are universally valid, while if they encounter disagreement, they conclude that these rules are of relative validity. The statistical principle at play is the trade-off between fit and flexibility: The hypothesis has to fit the data but must remain flexible so as to accommodate other factors producing consensus or variety of opinion (for instance, unreliability of the persons evaluating the topic), without becoming too flexible and thus empty, accommodating any data.[135] These mechanisms also explain the distinction between moral and conventional rules. If people (on statistical grounds) judge an action as universally right or wrong, they should regard it as a wrong in itself, independent of authority, too.[136]

8.7.2 *The Limits of Statistical Learning*

Importantly, this approach rightly underlines the significance of a descriptively adequate account of morality. It acknowledges the necessity for theory-building to determine properly the "acquirendum," the mental structure that human beings actually acquire in the moral domain.[137] It convincingly refutes various theories that identify morality as the expression of a small set of primitive emotions, seeking to deny that the operations of moral cognition are based on a rich systems of structures, rules, principles and representations.[138] It applies the poverty of stimulus test but arrives at the result that there is in fact no such poverty: The input available to children and statistical principles suffice to acquire the basic elements of the moral world of human beings it investigates. The theory is, therefore, an important constructive contribution to the understanding of human moral ontogeny.

However, several problems arise with regard to the reach of statistical learning. As just discussed, cognitive access to basic elements of the moral world is the precondition for understanding what moral judgment and explicit norms are about in the first place. You need to know what *ought* means, for instance, before you can grasp what you ought to do, as Hume already argued. Statistical learning is unable to bridge this gap. No quantity of references to ought by others will tell you what ought means from a first-person perspective, just as no quantity of references by others to the pleasant smell of coffee in the morning can teach you the nature of this odor

[134] Nichols, *Rational Rules*, 95 ff., 135 f.

[135] Nichols, *Rational Rules*, 109 ff., 132 f.

[136] Nichols, *Rational Rules*, 199 ff., argues that "default universalism" is functional – for instance, because it facilitates cooperation.

[137] Nichols, *Rational Rules*, 22.

[138] Nichols, *Rational Rules*, 8 ff., 150 f.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

from a first-person perspective if you cannot smell. The same holds for the other elements of human beings' moral world, including moral emotions like shame.

Another problem stemming from an insufficient determination of the acquirendum consists in the following: The moral principles human beings acquire are not captured with sufficient precision if one looks only at the doing/allowing distinction. The distinction between (intentionally) *doing something and allowing* X *to happen when you have a duty to act to prevent* X *from happening* on the one hand and (intentionally) *allowing* X *to happen when there is no such duty* on the other comes closer to a central element of human moral cognition.[139] The prohibition against a girl hitting her brother as punishment for taking her ball is morally equivalent to the prohibition against the child not preventing her younger brother from falling from a swing as punishment for taking her ball when she is playing with her sibling at a playground and has a duty to see that her younger brother does not hurt himself. In contrast, there is no equivalent moral duty to prevent all possible harm to all other children on the playground. Unsurprisingly, the relevance of this distinction is mirrored in basic provisions on criminal omissions in penal law around the world.

Moreover, children do not just learn that rules are act-based. They acquire the principle that internal states like direct intention (purpose) or oblique intention (knowing) matter when evaluating an action. Moreover, these internal states are relevant in complex ways for moral assessment: If one commits armed robbery of a bank with the purpose of getting money and knowingly creates a risk that a guard will be killed, and this indeed happens, in many legal systems this will count as intentional (knowing) homicide. The foreseen but not intended negative side effect of a purposive morally wrong act only makes it worse. In other constellations, for instance, when one intentionally acts to prevent harm but foresees harm to third parties as a side effect of the benevolent action, the act may be justified, as in the standard trolley bystander case. Thus, the acquirendum seems to be of a different and much more complex nature than is allowed for in the doing/allowing distinction that Nichols addresses. That any of these acquired complex structures can be learned through the statistical operations Nichols employs is far from clear, given our findings thus far. Take his example of intentional act-based rules. The evidence he considers ("Do not do X!" etc.) allows for the conclusion that somebody wants the agent to refrain from doing X, perhaps backing the prohibition to do X with the threat of sanctions. It does not allow for the conclusion that one is to refrain from acting with a specific internal state – for instance, an intention – or that one is to see intentions and actions causing harmful foreseen consequences as sometimes prohibited, sometimes justified, let alone that one is to acquire the capabilities to

[139] The question of whether the agent has a duty to act is crucial to evaluating such cases as "Footbridge-Allow," Nichols, *Rational Rules*, 4. It is also relevant for possible constraints on possible moral allow-based rules, Nichols, *Rational Rules*, 159 ff. Cf. n. 147.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

experience in oneself a moral ought and the other cognitive, volitional and emotional elements of the subjective world of morality already discussed above.[140]

The discussion of closure principles is also of interest. It tells us something relevant about the reactions to explicit instruction by prohibitions or permissions concerning the specific type of action (in Nichols' study, for instance, mice entering a barn).[141] For moral theory-building, however, the problem that closure principles deal with is a different one. It concerns the question of how people assess the permissibility of action even *without* explicit permissions or prohibitions concerning a specific type of action: Do you need an explicit prohibition to take an action as prohibited or do you need an explicit permission to regard it as permitted? What is the default principle (if there is indeed one) if there are no such explicit prohibitions or permissions? Nichols does not investigate this question because he is concerned with reactions after explicit instruction, though he indicates that there may indeed be a default principle of liberty. Only if one addresses this question, however, can one reasonably discuss the problem of how such a principle – for instance, of residual liberty – becomes part of the moral and (in liberal orders) legal world of human beings.[142]

As to the discussion of universalism and relativism, it is important that the validity of a norm is not dependent on the quantity of assent it finds, but on material validity conditions. Thus, the norm that Jews are just as entitled to life as other human beings was valid even in German extermination camps. Statistical learning thus allows for the hypothesis that many people agree about the validity of a certain norm, not that it is (in fact) valid. So-called authority independence is not related to consensus either. The point is rather that such universally valid, authority-independent norms are the basis to challenge even majority opinions. The phenomenon of critical reflection creates the possibility of changing a consensus using reasons – a process in which the universality of some norms is a possible argument. In Nichols' account, any norm could be understood as universal and authority-independent, as its universality depends solely on factually existing consent patterns. If the input that a child is exposed to consists of a consensus that it is justified to kill Jews, this norm enjoys universal validity, for instance. Nichols is not very clear on when and under which conditions the acquired rules change, but it seems that this change is dependent merely on new experiences of consensus or varieties of opinion. This account appears to miss the point that

[140] Nichols argues that the moral relevance of internal states like intentions can be explained by our general interest as humans in intentions, Nichols, *Rational Rules*, 159. The phenomenon to be explained is, however, not a general interest in intentions, but the origin of the constitutive and complex role of internal states for moral evaluation, which is already found at an early age, as we have seen.

[141] Nichols, *Rational Rules*, 101 ff.

[142] Cf. for such a discussion Mikhail, *Elements*, 132 ff. Nichols, *Rational Rules*, 107, reports the interesting result that the residual permission principle seems to be limited by prohibitions of harm.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

content matters for the question of universality, and that only certain content is a serious candidate for universally valid norms – for instance, norms prohibiting harm, but not norms prescribing gratuitous cruelty. This is also observed in the child psychology studies on the moral/conventional divide that Nichols accepts – not just any norms are at issue, but in particular prohibitions of harm.[143] That any of these norms constantly change with the shifting of opinions is hard to reconcile both with experience and with evidence.

As to some concrete rules that we discussed, it is not clear what kind of statistical evidence there is, for instance, for the prohibition of the instrumentalization of people. Do children observe a lot of this? How come human beings are able to critique and transcend norms that were robustly enforced in their environment – for example, on the permissible instrumentalization of women in patriarchal societies?

Nichols rightly underlines the difference between possible learning processes and actual learning and emphasizes that his account is only concerned with the former, not the latter.[144] The reviewed literature on the ontogeny of moral cognition seems to underline the relevance of the identified problems of his account. Moral development in fact seems to take not the course that statistical learning foresees, but a rather different one, including differentiated moral concepts in young children, cross-cultural equal age trends of developments of moral cognition, intention-based moral evaluation and so forth.

An interesting point that Nichols raises is the origin of the limited hypothesis space with which child learners operate. The theory offers no explanation (as it underlines itself) of how learners arrive at the hypothesis space. This hypothesis space is thus understood as possibly innate,[145] though there is some counterevidence against even such constraints, it is argued.[146]

The problem of constraints on hypotheses of possible morality is very important: What if children were exposed to the punishment of unintended actions? Some kind of strict liability morality? Would they then acquire a morality in which intention does not count in the moral evaluation of actions? Is there any evidence of this happening? Or is rather the opposite indicated not only by experimental work, but also by the passionate protests of children around the world when they are punished for something that they did not intend to do? The answer must be based on the poverty of stimulus argument: If the learning child does not receive sufficient

[143] The experimental evidence Nichols adduces uses nonmoral norms with unknown content. It seems to be not about the statistical learning of norms regarded as universal or relative, but rather about something different, namely how subjects use information in vignettes to assess the universality of norms whose content is unknown. They seem to regard consensus as a proxy.

[144] Nichols, *Rational Rules*, 152 ff.

[145] Nichols et al., "Rational Learners," 549; Nichols, *Rational Rules*, 154 ff.

[146] Tyler Millhouse, Alisabeth Ayars and Shaun Nichols, "Learnability and Moral Nativism: Exploring Wilde Rules," in *Methodology and Moral Philosophy*, eds. Jussi Suikkane and Antti Kauppinen (New York: Routledge, 2019), 64 ff.

input to form the hypothesis space – including the constraint that intentions count in a morality, for instance – then the constraints must be innate.[147]

In sum, the reference to statistical learning thus does not seem sufficient to account for the acquisition of the basic elements of morality.[148] Exploring the problem of the origin of the initial conception of morality that frames the acquisition of moral knowledge, the hypothesis space and its possible nativist explanation is an important point of the theory and highlights the need for open-minded research in this area. In any case, nothing in this research rules out that the Bayesian learning mechanism would lead to the acquisition of a morality incompatible with the content of human rights – on the contrary, this theory argues that human beings can learn any rule.[149] They are in no way naturally limited to narrow parochial tribalism. Statistical learning is open to acquiring inclusive norms,[150] such as – one may add – human rights.

8.8 THEORIES OF MIND AND HUMAN MORAL PROGRESS

There is thus a prima facie case for seriously considering the possibility that the moral space of human cognition, the concept of ought, moral emotions or certain elements of the principles of morals such as altruism, justice and respect for others form part of human beings' natural cognitive endowment that matures throughout childhood.

As illustrated by the mental gizmo thesis, the moral foundations theory, various approaches of evolutionary psychology, behavioral economics and the joint intentionality theory of cognitive development among others, the time has passed when studying the structures of the mind was considered unimportant because only one kind of theory of mind was believed to be plausible, namely a theory that assumed that the only inborn property of the human mind is that it is an unspecified learning

[147] The counterevidence adduced against this is not conclusive. The studies in Millhouse, Ayars and Nichols, "Learnability," 64 ff., investigate the doing/allowing distinction in moral rules to check whether a rule that only prohibits allowing something but not doing it violates possible innate constraints. This overlooks the fact that the crucial distinction is (as explained) not between doing and allowing, but between (for instance) intention (and its various forms) and negligence – one can intentionally do or allow something or negligently do or allow something. If there is a duty to act, allowing something to happen is morally relevant. The test case, therefore, is whether there is a norm not violating such constraints that prohibits the negligent killing of human beings but not their intentional killing, or that prohibits negligently allowing the death of human beings but not intentionally allowing them to die when there is a duty to act – a duty that will often exist in the case of danger to the life of others if one is in a position to help.

[148] Cf. for similar criticisms, with additional examples of the problems of underdetermination by statistical learning of children's moral knowledge, John Mikhail, "Review of Shaun Nichols, Rational Rules: Towards a Theory of Moral Learning," *The Philosophical Review* (2022) 131 (3): 399–403.

[149] Millhouse, Ayars and Nichols, "Learnability," 77.

[150] Nichols, *Rational Rules*, 189.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

device.[151] The mental gizmo thesis and moral foundations theory, for instance, are substantial empirical theses about the structure of the human mind, as is any assumption about the importance of shared intentionality, heuristics, framing effects and biases. Some of these theories have triggered a vast amount of research across the globe. Their claims may be right or wrong, but they certainly are serious scientific efforts that need to be evaluated based on their explanatory merits, and the same is true of the mentalist approach to ethics and law.

These remarks show that it is possible to frame an empirically minded theory of moral psychology that understands deontological principles not as cognitive illusions, but as part of the makeup of the human mind that may be the precondition enabling the cultural development of moral systems and the law.

Substantial empirical evidence suggests that there is a faculty of language with highly restrictive principles in which natural languages unfold, and there are many theoretical reasons for assuming its existence. This is a highly contested area of research, but even if this hypothesis is true, the theory of a faculty of language is still a long way from explaining the origin of the verses of *King Lear*. However, the language faculty is a precondition for humans' ability to produce and enjoy something like *King Lear*, whatever the hidden secrets of human creativity ultimately may turn out to be that put the language faculty to such thrilling use. Similarly, what has been said about the history, justification and psychological theory of human rights perhaps renders plausible the idea that a theoretical account of the human moral faculty is a long way from an understanding of the sources and justification of the *Universal Declaration* or other concrete, historically shaped catalogues of rights in constitutions and international bills of rights. But this moral faculty could – in the same way as the faculty of language for the verses of *King Lear* – turn out to be the cognitive precondition for the possibility of ultimately producing something like the *Universal Declaration* and the aspirations it implies.

8.9 CRITIQUE AND CONSTRUCTION: EXPLANATORY THEORY AND NORMATIVE ARGUMENTS

Let us assume that a mentalist account of morality and law has some merits and is preferable, for example, to the mental gizmo thesis or the moral foundations theory. This would be a very substantial insight for an explanatory theory of human moral cognition and for the theory of mind in general. But what normative significance

[151] Cf. the remarks in Chomsky, *Aspects*, 47 ff. For an overview of research on language acquisition, cf. Charles Yang et al., "The Growth of Language: Universal Grammar, Experience, and Principles of Computation," *Neuroscience and Biobehavioral Reviews* 81, no. B (2017): 103 ff. Note that these findings have epistemological consequences (e.g. for the problem of induction, Nelson Goodman, *Facts, Fiction, and Forecast* [Cambridge, MA: Harvard University Press, 1983], 64 ff.), as they help us to clarify the origin of the core conceptual space of human beings.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

would this have if one wanted to avoid a naturalistic fallacy? This is the next question our inquiry will address.

The discussion thus far has already made clear a first function of the analysis of the relation between mind and rights: *identifying unwarranted human rights criticism based on unconvincing theories of the mind* – for instance, on unsatisfactory neuroscientific studies or insufficient evolutionary theory. This obviously already is a very important function.

Our findings so far refute, for instance, the idea that deontology is no more than a cognitive or moral illusion from a hard-headed, non-armchair, scientific point of view. Deontological arguments are not discredited in any way by the theory of mind, psychology or neuroscience. Of course, critiquing implausible theories of moral cognition does not in itself justify the normative principles that enter into a theory of justification of human rights. But defending cognitive principles outlined as reasonable is quite a different matter from defending the normative value of principles that are products of the post-hoc rationalization of hardwired emotional gut reactions. Moral psychology cannot substitute normative theory-building in ethics and law. But it is indispensable to show that normative theory is not just the illusionary offspring of hidden mechanisms of the mind and thus can be reconciled with what is known about the mind's structure and workings.

The same holds for forms of evolutionary tribalism: Our discussion showed that a kind of universalist ethics is at least not an evolutionary impossibility. While this does not make the normative case for such an ethics, it is a different justificatory task to argue the legitimacy of rights that are contrary to human beings' cognitive nature, assuming, for instance, that "parochial love – love within groups . . . may be the most we can accomplish,"[152] and where it remains a complete riddle as to how the cognitive machinery that produced them could have evolved, than to argue for a set of rights that is entirely reconcilable with cognitive mechanisms that easily fit into a plausible theory of the evolution of the human mind.

This critique, is, if you will, preparatory work that lays the ground for normative arguments by showing what kind of counterarguments are insufficient to discredit a justificatory theory of human rights.

This critical function of an inquiry into mind and rights already justifies the efforts made thus far. It is quite clear from our review of current thinking about moral psychology that the is/ought distinction does not prevent the theories scrutinized from influencing how human rights are conceptualized. They are among the influential sources that feed into human rights skepticism today.

But can such a theory of human moral cognition provide more than this crucial criticism? Can it perhaps even provide some kind of additional normative justification of human rights?

[152] Haidt, *Righteous Mind*, 295.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

Of particular interest in this respect is the question of whether a universalist justification of human rights exists or whether human rights are culturally relative. In our survey of human rights history and normative theories of human rights, we encountered no compelling evidence for the latter.

The question of universalism is of substantial interest because it concerns a central claim of human rights – to be valid for everybody, everywhere. This question is important in both practical and political terms: The answer decides, for instance, whether Uighurs have a claim to religious freedom and nondiscrimination against China, or whether the cultural difference between Zürich and Beijing acts as a bar to such claims.

Our discussion of the justification of human rights has helped us to clarify the issue. As we have seen, a theory of the justification of human rights contains three elements: first, a theory of goods and, as part of it, anthropological assumptions; second, a political theory of the role of rights in society; and third, normative principles of justice, solidarity and respect. Anthropological assumptions are a matter of empirical knowledge about human beings and have no direct connection to moral cognition. The political theory of rights has normative elements but concerns other questions, too, such as the factual effects of a political order of rights on people's well-being. The question of universalism thus mainly concerns the normative element of the justificatory theory of human rights. Are the guiding principles and the normative tenets of such a theory universally valid? This is the core question at issue in debates about the universalism of human rights.

What do our findings so far tell us about the justification of normative universalism? Is the idea of a universal and uniform human moral faculty, a universal moral grammar, perhaps the high road to universalism? Is this the claim made and the ultimate point of the argument? If one takes the is/ought distinction seriously, however, it seems that this cannot be true. Whatever the factual makeup of the human mind may be, it has no bearing on the question of normative justification because no ought follows from it.

On the other hand, there are theories that doubt the relevance of the is/ought distinction.[153] Does the discussion thus far offer new support for these theories?

Two steps are necessary to answer this question. First, we will clarify in greater detail what normative universalism is about and what its justification might be. In light of this, we will then ask what kinds of consequences a plausible account of human moral cognition has for the understanding of the universalism of human rights.

8.10 THE EPISTEMOLOGY OF HUMAN RIGHTS UNIVERSALISM

Are human rights only legitimate for some groups of human beings – say, Europeans, North Americans, Christians and whites? Is the global appeal of human

[153] Cf. for example John R. Searle, "How to Derive 'Ought' from 'Is'," *Philosophical Review* 73, no. 1 (1964): 43 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

rights based on an epistemological error, namely on the flawed idea that the core normative ideas of human rights can be justified across the borders of different communities? Or are these principles potentially of universal validity, and thus the project of human rights enjoys universal validity, too?

These are standard questions of any discourse on human rights, and ones that we hope to answer, if we are able, to clarify the content of universalism and determine at least the rough contours of its epistemological merits.

Universalism in the sense relevant for ethics and law is an epistemological stance.[154] It holds that the truth conditions of normative claims are the same for all human beings. These truth conditions are thus not relative to certain contingent properties of human beings, such as the groups that they belong to or their social, cultural, religious or other background. This epistemological doctrine applies to any normative propositions, including those concerning the moral rightness and justness of intentions, actions and states of affairs brought about by such intentions and actions. It also applies to the obligations people have, to what they are forbidden and permitted to do and to what rights they enjoy. Such universal standards are valid even if a given subject thinks otherwise. From such a point of view, a man is obligated to respect a girl's right to education, even if he thinks that this is contrary to important norms of his patriarchal customary morality.

A defense of universalism may be based on two lines of argument: The first is an indirect argument for the plausibility of universalism derived from the implausibility of the opposite of universalism, which is relativism.[155] One argument commonly cited in support of relativism that we have already encountered is moral disagreement. The great diversity of moral opinions is taken to be an argument for the relativity of moral propositions. This is a fallacious argument, however, and for the following reasons.

To start with, universalism is a theory about the justification of normative propositions. It does not imply that universally valid principles are in fact universally accepted. The existence of even deep moral disagreement does not contradict universalist perspectives. Parallel observations are possible in the realm of science. For example, there are quite good reasons to assume that the Earth is not flat. These reasons are in no discernible sense relative to the contingent properties of a thinking subject making this proposition about the shape of the Earth. This does not mean,

[154] For some other uses of the term "universalism," cf. e.g. Seyla Benhabib, "Another Universalism: On the Unity and Diversity of Human Rights," *Proceedings and Addresses of the American Philosophical Association* 81, no. 2 (2007): 7, 11.

[155] Cf. e.g. Gilbert Harman and Judith Thompson, *Moral Relativism and Moral Objectivity* (Oxford and Malden, MA: Blackwell Publishing, 1996), 3–64; Philippa Foot, "Moral Relativism," in Philippa Foot, *Moral Dilemmas and Other Topics in Moral Philosophy* (Oxford: Clarendon Press 2002), 20–36; Richard Rorty, *Objectivity, Relativism, and Truth: Philosophical Papers* (Cambridge: Cambridge University Press, 1991); Bernhard Williams, "The Truth in Relativism," in Bernhard Williams, *Moral Luck* (Cambridge: Cambridge University Press, 1982), 132–43; Dworkin, *Justice for Hedgehogs*, 23 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

however, that there were not times in history when the very opposite of this insight was the common wisdom of the age. And even today, flat-earthers courageously try to make their case. However, their disagreement does not imply that there is no justification for assuming that the Earth is round.

Moreover, as we have already seen, there are factors entirely reconcilable with universalism that account for this – evident – diversity, such as nonmoral preconditions of moral evaluation, competing interests and passions or problems of ethical reflection – for instance, concerning the proper conceptualization of morality and its content or the consistency and coherence of normative reasoning.

Another important point is that relativism suffers from an insufficient determination of the factors to which moral principles are supposed to be relative. The reference to cultures or religions is a good example of this. Cultures and religions are not monolithic wholes but encompass a whole variety of ideas. Western culture (whatever the exact boundaries of this entity may be) was characterized for centuries by religious intolerance leading to bloody wars, authoritarian regimes and dogmatic systems of thought, contempt for human beings, various forms of racism, slavery, colonialism and imperialism, to name just some features that may spring to mind. At the same time, great ideas of human benevolence, freedom, autonomy and justice were outlined in forms of which some made history. Who is the true European – Kant or Friedrich Wilhelm III? Or – more precisely – the Kant of the principle of humanity and dignity,[156] the Kant of the disenfranchisement of women[157] or his Prussian king?

Another problem is the following: How do background factors such as culture or religion determine moral perceptions *exactly*? Cultural determinism is a highly implausible theory given the fact of constant change in the normative sphere. How can normative principles be determined by the cultures that these principles ultimately transform?

This points to the central problem of relativism that we already encountered in our survey of human rights history: It overlooks the importance of *human subjectivity and autonomous reasoning* for the development of ethics and the legitimation of law. Any cultural influence is mediated by human reflective subjectivity and reasoning. There are many cultural influences on human beings, but no person's ethical identity is necessarily merely the product of the lullabies sung by the hand that rocked the cradle. On the contrary: Human reasoning is the ultimate source of ethical beliefs. Critical reflection can transcend the given parameters of culture and history and ultimately change their course. This does not mean that such autonomous thinking is what always or even predominantly determines the actual moral and political path of the human species. The reality is that ethics and law often fall prey to outlived customs, the self-righteous perpetuation of principles without

[156] Kant, *Grundlegung zur Metaphysik der Sitten*, 429.
[157] Kant, *Grundlegung zur Metaphysik der Sitten*, 313 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

thought, reverence to social authorities, powerful and harmful political emotions such as the hatred of minorities, fear and other such influences. But these need not be the last word. People are not just the malleable victims of such factors. As thinking subjects, they are in a position where they are responsible for reducing the importance of such driving forces for the course of history, and sometimes do so successfully – as the many steps taken towards a more humane world indicate, from the partial vindications of the rights of women to the defenses of democracy and human rights.

The second way of arguing for the plausibility of universalism is to defend certain normative propositions as universally justifiable. But is this possible? Is this not epistemologically naive? To answer this question, some remarks on the epistemology and ontology of morals are necessary.

8.11 THE EPISTEMOLOGY AND ONTOLOGY OF MORALS

The justification of moral judgments is crucial.[158] There is no a priori metacriterion that does not require scrutiny and defense. While ethical reflection is not about the revelation of higher truths, it is not about whimsical skepticism either. One needs reasons to justify any normative proposition, but one also needs reasons to doubt it. Simply maintaining that a certain normative position still could be doubted is not good enough. One can doubt any proposition, including the idea that the Earth is not flat, as flat-earthers happily do. No bolt of lightning will punish such a doubt, no voice from heaven will confirm that the Earth is in fact not flat. Nothing renders doubt about anything impossible. Nothing in the best argument imaginable compels a thinking subject to get its point. Demanding that a good argument be literally indubitable is a flawed demand because it is too exacting. Not even the best argument about the shape of the Earth irresistibly commands assent. This is a necessary consequence of the fact that the assessment of the truth value of any proposition is an act of human mental cognition. There is no epistemic authority over and beyond such acts of cognition. There is no more stable ground of human insight. Such acts of human cognition are the stuff of which the perception of truth is made.

This analysis is confirmed by a plausible ontological theory of morality: The problem of whether moral judgments refer to objective, mind-independent moral facts in the world (as moral realists assert) or not is a classic epistemological question. In the former case, a truth condition of moral propositions consists in correspondence with moral facts, in the latter case other truth conditions are key. As far as the ontology of morality is concerned, it is plausible to take moral cognition as being

[158] As convincingly emphasized by Forst, *Recht auf Rechtfertigung*, irrespective of whether one is convinced by this approach's discourse on ethical foundations.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

nonreferential: There are no objective moral facts in the world, the correspondence with which is the truth condition of moral predicates, as moral realists maintain.[159] Nevertheless, moral judgments are not merely subjective in the sense of them being idiosyncratic and relative to the outlook of a specific agent.[160] This is because their truth is authenticated by internal mental yardsticks for justified propositions, as in other areas of thought.[161]

It is useful to note that not only from the point of view of a nonreferential moral ontology, but also from a moral realist point of view, there is no escaping from acts of cognition that rely on internal standards of human understanding for their truth. This is because even a moral realist ontology ultimately is based on the assumed truth of such nonreferential statements: The correspondence of a moral proposition with a moral fact can only be ascertained by such an act of cognition. Moreover, the moral realist thesis that a truth condition of moral judgments is that moral predicates correspond to objective moral facts in the world does not itself correspond to an

[159] On this cf. Mikhail, *Elements*, 317; Mahlmann, "Ethics," 580 ff. For a defense of the view that there are, to the contrary, objective, irreducibly normative facts, cf. e.g. Russ Shafer-Landau, *Moral Realism: A Defence* (Oxford: Oxford University Press, 2003); David Enoch, *Taking Morality Seriously: A Defense of Robust Realism* (Oxford: Oxford University Press, 2011). A nonreferential theory of ethics does not commit one to noncognitivism, desire-based ethics, expressivism and the like, as explained in the text. The debate between moral realists and antirealists – as it stands today – does not exhaust the theoretical possibilities.

[160] On a standard view on this and its critique, Griffin, *On Human Rights*, 111 ff.: Factual judgments are objective, value judgments are "subjective – subjective in both the most common senses. They are, first of all, merely expressions of taste or attitude. And, second, values are not part of the furniture of the world; the world contains physical objects, properties, events, minds, but it does not also contain values."

[161] That there are genuine normative reasons whose truth does not depend on correspondence with entities that are part of the nonmental fabric of the world is defended from different points of view. Cf. Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996), 108, 122 ff., 165 arguing for a "reflective endorsement theory" that bases normativity on the self-endorsement of the humanity of the autonomous self; Dworkin, *Justice for Hedgehogs*, outlining an interpretative theory "all the way down"; Derek Parfit, *On What Matters*, Vol. 2 (Oxford: Oxford University Press, 2011), arguing that there are "some irreducibly normative reason-involving truths," which are "not about entities or properties that exist in some ontological sense," ibid. 618; Thomas Michael Scanlon, *Being Realistic about Reasons* (Oxford: Oxford University Press, 2014), developing a realistic "reasons fundamentalism." On the question of the authentication of truth ultimately through foundational intuitions of truth, Ray Jackendoff, *A User's Guide to Thought and Meaning* (Oxford: Oxford University Press, 2012), 213 ff., taking this as evidence that (in the terminology of the dual-process model of the mind) System 2 (slow thinking) rides on top of System 1 (fast thinking) with the means of language, without, however, making thinking irrational or emotional, because "it behoves us to show intuitive thinking more respect," ibid. 215. On a view that bases arguments on a specific language game and lifeworld, Griffin, *On Human Rights*, 113: "Certain values are part of the necessary conditions for our language, which sets for us the bounds of intelligibility" (on Wittgenstein and Davidson). Bernhard Williams has defended a related view on thin and thick ethical concepts, ultimately basing moral judgment on particular "languages" in the specific sense of particular comprehensive systems of belief, cf. Bernhard Williams, "Truth in Ethics," *Ratio* 8, no. 3 (1995): 227–42. In his view, an alternative is to try to identify a deep structure of moral judgment – this is what is attempted in the argument of this book.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

objective epistemic fact in the world. Its truth thus depends on other sources of epistemic justification.[162]

Given this epistemic state of affairs, it is always possible to ask: Is this right? The possibility of doubt is the necessary consequence of the human capacity for unfettered, free thought. It is a fallacy, however, to mistake this possibility of doubting any proposition for the proof that no proposition is better justified than any other. It is necessary to have arguments for doubt that are more convincing than the arguments that (for the time being) confirm a certain normative position, at least if the argument is supposed to be more than a superfluous game.

Let us take the example of the principles important for the project of human rights, the idea that every human being has the right to protected dignity, autonomy and equality. There are many thoughtful and rich theories of human rights, as we have seen. And there are rather good reasons to assume that these rights are legitimate not only in Zürich, but also in Laos, Bogotá and Beijing, because the goods they protect, if ever enjoyed, turn out to be of extremely high value. This is shown, for example, by the longing for equal dignity and freedom of the victims of the Tiananmen Square protests, a powerful example of important "Asian values," as is the struggle for freedom in Hong Kong, the fight against the military dictatorship in Myanmar or the rebellion driven by Iranian women against the regime, although in all these cases expressed from below, not through government announcements.[163]

[162] Cf. on this argument recently, Dworkin, *Justice for Hedgehogs*, 76; Scanlon, *Being Realistic*, 16 n. 1. For a critique of such theories, arguing in particular with the impossibility of distinguishing on internal grounds true reasons and false moral propositions, rendering all argument indistinguishable from fictions, Enoch, *Taking Morality Seriously*, 121 ff. The central point to counter this concern is that internal truth conditions of moral propositions are not matters of the agents' whim. Moreover, the argument is self-defeating: the moral-realist claim about objective moral facts, the correspondence with which is the truth condition for moral propositions, is necessarily itself based on a nonreferential theory of human epistemology, as just explained in the main text.

[163] There is an interesting debate about such "Asian values," which is paradigmatic for some important features of the debate about relativism, not the least its political side, more precisely on the political instrumentalization of relativism for authoritarian purposes. Cf. e.g. Zakaria, "Culture is Destiny," 109 and the rejoinder: Kim Dae Jung, "Is Culture Destiny? The Myth of Asia's Anti-democratic Values," *Foreign Affairs* 73, no. 6 (1994): 189. These discussions sometimes revolve around whether there is a particular tradition that predominantly endorses human rights or other such values, or rather an authoritarian tradition. This is an interesting question, but it misses the most important problem, however: The crucial point is not whether or not there is a given tradition (e.g. of authoritarianism), but whether such a practice is justified. An authoritarian tradition certainly is manifest in much of European history. One central achievement of dawning constitutionalism, for instance, was to break with this tradition to vindicate some of the most important rights of human beings, step by step. The same holds for any other tradition as well (if it is more than an ideological construct): "The so-called Asian values that are invoked to justify authoritarianism are not especially Asian in any significant sense. . . . The case for liberty and political rights turns ultimately on their basic importance and on their instrumental role. This case is as strong in Asia as it is elsewhere." Amartya Sen, "Human Rights and Asian Values," in *Ethics & International Affairs*, ed. Joel H. Rosenthal (Washington, DC: Georgetown University Press, 1999), 170 ff., 190.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

That there is a political case for human rights to protect such goods has been argued extensively above. That this basic good should be distributed equally among human beings, that it is not justifiable for some groups – say, party functionaries and their partners in the economy – to be entitled to such freedoms while peasants are not, seems equally plausible in light of the basic principles of equal treatment that are constitutive of justice. If these arguments and their foundations are doubted, reasons need to be given for this – for example, that serfdom is in fact good for some people (say, people of color) and that justice can as plausibly mean unequal treatment as it does equal treatment, at least for some (e.g. in the Global South). While the preceding remarks on the justification of rights surely have many flaws, they at least indicate that shouldering this burden is not an easy task.

8.12 EPISTEMOLOGICAL RESILIENCE

There has been a long debate about foundational theories in ethics and law.[164] What are the ultimate principles of justification of normative precepts and what is their status? How can one argue for these principles without becoming entangled in infinite regress, dogmatism or tautology?[165] The line of the argument seems to point in the following direction: The normative justification of human rights is ultimately based on a fallible account of certain foundational moral principles that have proven to be resilient to systematic, conscious theoretical doubt. This account is informed by critical thinking that is committed to the decisiveness of reasons and open to the possibility that sometimes such reasons are good enough to rest a case. This stance implies a certain amount of trust that human understanding does not lead us entirely astray, and that consequently what seems well justified to us may, all things considered, in fact be true – an epistemological stance that can be defended against skeptical challenges.

This reflexive resilience of fallible normative principles when scrutinized by critical, in principle reliable human thinking that is respectful of reasons is the epistemological alternative to the trilemma of infinite regress, the dogmatic termination of the justificatory argument and tautology. The principles of justice, altruism and respect play an important role because there is good reason to believe that they (or some variant of them) are foundational not only for human rights, but for morality as such. They are not self-authenticating truths, but improvable, preliminary approximations to those principles that constitute morality.[166] Other principles may play an important role, too, such as the principle of the noninstrumentalization

[164] Cf. for a survey Mahlmann, *Rechtsphilosophie und Rechtstheorie*.

[165] Cf. on this trilemma Hans Albert, *Traktat über kritische Vernunft* (Tübingen: Mohr Siebeck, 1991), 14.

[166] Cf. on these matters Mahlmann, "Ethics," 593 ff.; Mahlmann, *Rechtsphilosophie und Rechtstheorie*, 510 ff. For some comments on why such judgments should be taken as foundational, cf. Mahlmann, "Cognitive Foundations," 75 ff.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

of human persons as a concretization of obligatory respect for others or those specifying the permissibility of otherwise-prohibited acts.[167]

On this basis, further, more concrete questions can be asked – for example, whether the ethical thought formulated in the principle of humanity and the idea of human dignity can be put to ethical and legal use to guide the evaluation of specific problems, from assisted suicide to abortion to "dwarf-tossing".[168] In this way, we may be able to approach a more comprehensive theory of human rights at a much-needed level of detail.

8.13 UNIVERSALISM WITHOUT DOGMATISM AND HUMAN RIGHTS PLURALISM

There is thus a case for the universal justification of human rights. There is, however, another problem: The ethical ideas about human rights and the legal systems enshrining them differ, and sometimes more than just in detail. How are we to deal with this fact of ethical and legal human rights pluralism?

Can human rights pluralism in the real world be reconciled with normative universalism? This may seem implausible if the assumption is that universalism must demand normative uniformity around the world. But this is by no means a necessary conclusion, and for various reasons.

The normative questions human rights are concerned with are difficult and complex. Even if we assume that normative universalism makes good sense as an epistemological theory, this does not imply that it is reasonable to expect everybody to agree on all normative matters after a little deliberation. Moral and legal problems are more complicated than that. There is no reason to believe that the tortuous path of cognition is easier to walk in practical reflection than in other fields of human knowledge, including natural science. Modern theory of science is very varied, and the historicization of scientific paradigms is an important element of this reflection on the nature of science.[169] However, these qualifications do not change the epistemic claim of science that whatever seems to be the best theory available at a given moment enjoys this status universally. Higgs particles are supposed to make as much sense at CERN in Geneva as in Saigon. Science's history of getting where it has got so far has not been a smooth ride to insight. On the contrary, its path is so fascinating and rich precisely because of the many difficulties encountered, dead-end roads pursued and breakthroughs achieved. The expectation that it might be any easier to gain normative insight through practical, ethical and legal thought seems somewhat outlandish.

[167] The latter is the argument advanced in Mikhail, *Elements*. On the common law concept of "battery" and its possible role in a mentalist ethics, Mikhail, "Any Animal Whatever?" 750 ff.

[168] On this question, cf. Mahlmann, "Good Sense of Dignity," 593 ff.

[169] Cf. e.g. Kuhn, *Structure of Scientific Revolutions*; Paul Feyerabend, *Against Method* (London and New York: Verso, 2010).

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

Universalism as an epistemological stance gives no reason whatsoever for anybody to be sure that one has in fact reached normative insights that hold universal validity. This is because human thought, not least practical thought, is irredeemably fallible. It is important both for individuals and their sense of epistemic modesty and for collectives to remain conscious of this fact, so as to avoid unwarranted self-righteousness about the justification of the respective realizations of certain basic values that they (at the given historical moment) deem to be right. Any normative claim is at best a better or worse approximation of principles that are perhaps really universally justified.

The internal pluralism of interpretations of given norms within ethical reflection or legal systems therefore comes as no surprise. The origins of such diverse opinions are manifold. Legal norms, in particular, pose many problems and are often ambiguous and vague. The difficulty of identifying the content of such norms and the underlying normative principles is one reason for the diversity of legal opinions, even about the content of very fundamental norms. In addition, there are political agendas, interest-based legal interpretations and the like at play.[170]

This consciousness of the fallibility of human practical thought has a pragmatic side: It implies the necessity of freedom for experimental ways of living,[171] of attempts to explore different ways of realizing certain important normative principles.

This is also true for fundamental rights. How to best protect rights is not always obvious, and it makes good sense to see how certain approaches work in practice. There will be limits to such experiments,[172] but universalism certainly does not categorically exclude the possibility of such searches for normatively justified solutions – to the contrary, the awareness that any normative conclusion will be tentative in fact demands scope for such searches.[173]

The plurality of attempts to approximate a set of well-justified norms is therefore one of the practical consequences of this uneven path of practical cognition.

[170] "There is nothing that interpretation just is," Cass R. Sunstein, *A Constitution of Many Minds: Why the Founding Document Doesn't Mean What It Meant Before* (Princeton, NJ: Princeton University Press 2009), 19 ff.

[171] Mill, "On Liberty," 281.

[172] From this perspective, the importance of the "Western" systems of rights protection can be relativized. Cf. Heiner Bielefeld, "'Western' versus 'Islamic' Human Rights Conceptions?: A Critique of Cultural Essentialism in the Discussion on Human Rights," *Political Theory* 28, no. 1 (2000): 90, 101 f., on the "Western model" being an example, not a normative paradigm for the struggle for human rights.

[173] J. Donnelly argues, for example, for a "relative universalism," a variation of universal concepts derived from alternative conceptions and implementations of rights on the basis of an overlapping consensus on universalism, taken as a Rawlsian political approach. In his view, freedom of religion can be reconciled with the prohibition of apostasy, albeit with the crucial qualification that sanctions of the prohibition should not violate human rights. With this qualification, the argument becomes tautological. Whether human rights allow for such qualification is the question at stake. Cf. Jack Donnelly, "The Relative Universality of Human Rights," *Human Rights Quarterly* 29, no. 2 (2007): 281, 298 and more generally Jack Donnelly, *Human Rights in Theory and Practice* (Ithaca, NY: Cornell University Press, 2003).

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

Human rights pluralism is a necessary tribute to the difficulties of the problems posed for practical reflection on human affairs and to the fallibility of human thought.

Another reason why normative universalism is reconcilable with human rights pluralism is very straightforward, namely *respect for personal autonomy and democratic choices*, another source of the legitimacy of experiments in different ways of living. If autonomy means anything, it means the ability to make such choices, without any control of their content as long as they do not violate basic normative principles, most importantly the rights of others. Any legitimate order must leave scope for such acts of self-determination. This includes acts that may (and with good reason) appear less than wise, acts that are the product of partisan political interests or passing political passions. Human rights orders are not just products deduced from the textbook of normative insights. They are embedded in societies' political processes, with all of the political upheavals and untidy decision-making that make democracies a living, admirable and demanding form of political life. Such political processes have effects on human rights law, sometimes becoming law, sometimes coloring the interpretation of norms, adding to and modifying the meaning of the pluralism of human rights.

Defending the kind of universalism outlined here therefore does not fall prey to dogmatism, Eurocentrism or moral imperialism. The epistemic status of a proposition is one thing, the practical consequences drawn from the status of such propositions quite another. One may think that certain normative principles have universal validity, such as the equality of women, without implying that this provides sufficient reason to invade other countries to spread this gospel by force in communities that think otherwise. Universalism with humility is a very helpful starting point.[174] The parallel to science may once again be illuminating here: Scientists may have compelling reasons to assume the existence of black holes. This does not mean, however, that they are entitled to muster an army to subdue all those astronomers who think otherwise. The reasons for this are the very rights that universalism defends, prominently including the autonomy of free human thought and decision-making. As this is so, the pluralism of normative orientations is the welcome companion of universalist normative theory. Pluralism, including human rights pluralism, thus not only is reconcilable with universalism, but is in fact its necessary consequence.[175]

[174] Cf. András Sajó, "Introduction: Universalism with Humility," in *Human Rights with Modesty: The Problem of Universalism*, ed. András Sajó (Leiden: Martinus Nijhoff, 2004), 26: "Overreaching application of norms should really only be grounds for criticism of what they in fact are: the erroneously extended application of universalism, and not for criticism of the underlying concept of universalism itself."

[175] This does not imply that all factually held positions are equally justified. It just means that there are reasons derived from human autonomy to respect normative positions despite the fact that they may not be fully justified. For a different view, arguing for the idea of the possibility of diverging positions that are equally justified, Rawls, *Political Liberalism*, 144. Rawls, however, assumes the shared norm that human beings are free and equal, which seems not to be open for reasonable disagreement from his point of view: The idea of reasonable disagreement does not reach all the way down to the core principles.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

A further point needs mentioning: Values are not just intellectual positions. A certain view on freedom of expression or religion, equality or human dignity is not the same in nature as a position on the question of whether or not Higgs particles exist. In contrast to descriptive propositions, value statements have emotional dimensions in at least two senses. First, certain emotions are the consequences of moral judgments – say, moral outrage at the sight of gross injustices. As we have seen, the moral judgment is the precondition of such emotions; the emotions do not constitute the moral judgment, as emotivists maintain.[176] Second, emotions have a heuristic function, as we underlined above: They are necessary to fully understand what certain acts mean for human beings. One needs to have fathomed the importance of the goods protected by human rights to give them their proper weight: One will not fully understand what freedom of expression means if one has not experienced the preciousness of the possibility of free speech. Freedom of religion will only be appreciated appropriately if one has a conception of the existential importance that religion holds for many people, irrespective of one's own religious, atheist or agnostic views. The meaning of prohibitions of discrimination can only be understood if one has a sense of the emotional harm inflicted through degradation by the many forms of discrimination – for example, by racism. Human equality will only acquire its full importance if one has a sense of what respect and disrespect for humans as equals feel like.

This consciousness of the existential meaning of human rights, of how they and their violations are spelled out in concrete human lives, is of central importance not only for ethics, but also for law. It is a key to the weight of rights. The appreciation of the importance of certain rights guides decisions about the inclusion or exclusion of rights in a given human rights catalogue. This appreciation is of great importance in the specifying of rights and rendering them concrete, in particular with regard to weighing and balancing rights in the framework of a proportionality analysis. One's answer to the question of whether a limitation on religious expression – for instance, a ban on headscarves, kippahs, monks' or nuns' habits or turbans – is legitimate or not depends very much on how important one considers freedom of religion to be.

Such sensitivities need time to develop, and they will do so unequally in different societies. They may fade away because certain experiences become remote – for example, the experience of what religious strife may mean for a society or what life in societies without freedom feels like. In addition, powerful political passions influence the course of history, including constitutional history, which in the best case are channeled and put to good use to foster the core ideas of human rights, which is not an easy task: "We are not passion's slave, but we have to apply cunning for reason's success."[177] However, the taming of such passions may not succeed, with potentially severe consequences for certain elements or even for the general

[176] Cf. Mahlmann, "Ethics," 585 ff.

[177] Sajó, *Constitutional Sentiments*, 2.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

architecture of law. Accordingly, normative views will shift, adding, reinforcing, altering the variety of normative positions held on certain issues. All of this can be expressed in human rights norms and interpretations. Given today's internationally interwoven legal systems, such a variety of approaches will include a transnational dimension of norms with a human rights function. Human rights in legal reality are thus inevitably of a protean nature – heterogeneity and variety will continue to be their mode of existence as law.

These observations lead us to a final point: There can be no teleology in any universalist theory worth considering. Universalism does not imply that universally valid norms will necessarily become the law of humankind. There is nothing in history or theory to support this thought. Human beings have moral insights, but they are driven by many motives, and some of the most powerful such motives are the most destructive. The possibility of human societies at least approximately ruled by universally justified norms, although these norms are always only tentatively identified, is a hope, not a certainty, and is entertained to a large degree despite, not because of human history.

The pluralism of human rights orders may therefore violate universally valid norms of equality, liberty and autonomy, and the equal worth of human beings. It did so in the past, and there is no guarantee that it will not do so in the future. But this may not necessarily be the case. Pluralist orders can be something else, namely the fertile attempts of human beings to come a few steps closer to the ideas of justice, solidarity and respect that are the core attractions of the human rights project.

8.14 A NEW CASE FOR UNIVERSALISM?

In sum: Universalism as an epistemological position does not demand a necessary uniformity of the moral judgment of all human beings in the real world with its profound complexities and many influences on moral opinions. Nor does it rule out human rights pluralism. What is, then – to return to the important question asked above – the normative importance of a theory of moral cognition? Does it add something to the justification of the validity of those core moral principles that seem to be sufficiently resilient against doubt to form the heart of the normative argument for human rights? In particular, would a theory of a universal and uniform human moral faculty strengthen the case for a universalism of the kind outlined?

The answer is: A theory of the structure of moral cognition has no normative consequences as such, because normative arguments are required to justify any normative point. This also is true for universalist claims with the qualified content outlined. Even if there is indeed something like a universal moral grammar with justice, altruism and respect as its core principles that specify how moral judgements supervene over facts, this would not add anything to the justificatory argument. One would still need to make the case that these

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

principles ought to guide ethical reasoning and legal norms and institutions and need not be corrected, say, by utility calculations or adherence to the principle of "might is right." One way to make this case involves the arguments developed above, including the absence of any discernible reasons to doubt the validity of these central principles. There is no good case for assuming that it is morally right to intend to harm others, that it is just to treat equals unequally or that one should disrespect other human beings.[178]

John Mikhail has explored the "weak normativity" of empirically given structures of moral thought. He argues that properly describing these structures of moral thought already is a meaningful step towards justifying their validity. This argument draws on Goodman's theory of induction and its use in Rawls's theory of justice. Goodman argues concerning the problem of induction:

> We no longer demand an explanation for guarantees that we do not have, or seek keys to knowledge that we cannot obtain. It dawns upon us that the traditional smug insistence upon a hard-and-fast line between justifying induction and describing ordinary inductive practice distorts the problem. And we owe belated apologies to Hume. For in dealing with the question how normally accepted inductive judgements are made, he was in fact dealing with the question of inductive validity. The validity of a prediction consisted for him in its arising from habit, and thus in its exemplifying some past regularity. His answer was incomplete and perhaps not entirely correct; but it was not beside the point. The problem of induction is not a problem of demonstration but a problem of defining the difference between valid and invalid predictions.[179]

Rawls' use of the reflective equilibrium can be interpreted along these lines. From this perspective, Rawls provides a descriptively adequate account of the sense of justice, as shown by considered judgments. A descriptively adequate account of the sense of justice is an account that captures correctly the main properties of the sense of justice. Considered judgments are judgments that are not skewed by interest, bias and so on, as explained above. The argument advanced on the basis of this interpretation of Rawls is not to follow directly in Goodman's footsteps and hold that a descriptively adequate account of the sense of justice already serves as the key to the justification of the principles of justice – the practice of justice in itself is not enough to justify the principles of justice.[180] This is because the principles of justice have to pass the test of rationality. The argument is, rather, that a descriptively adequate account of moral judgment provides presumptive, defeasible reasons for

[178] These findings are important for the project of "computational ethics," Awad et al., "Computational Ethics." Ultimately, not descriptive accounts of moral intuitions but normative arguments are decisive for answering normative questions, including the evaluation of the use and functions of AI designed to assist or substitute for human decision-making.

[179] Nelson Goodman, *Facts, Fiction, and Forecast* (Cambridge, MA: Harvard University Press, 1983), 64 f.

[180] Mikhail, *Elements*, 208.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

the justifiedness of the principles of justice identified as underlying considered judgments of human beings.[181]

For the problem of the normative relevance of the findings of moral psychology, the most important point regarding Mikhail's profound idea is that it also seems to point to the view that, ultimately, it is normative arguments that count for the validity of a moral principle: In the end, such normative arguments are the decisive resource for refuting the defeasible reason for the justifiedness of moral principles that descriptively adequately account for considered judgments. In this sense, it confirms the argument presented here that there is no getting away from normative reasoning.

A very interesting problem looms in the background here: If it is true that human beings have certain mental faculties, the properties of which enable their thought but necessarily at the same time also limit the scope of their thinking, then any normative argument ultimately draws on these given mental resources. In this case, the principles that in fact determine moral judgment ultimately are the principles used to evaluate these very principles that in fact determine moral judgment: The justness of the principles of justice that in fact direct human moral judgment is then based on principles of justice that in fact direct human moral judgment – an obviously circular argument.[182]

This can be formulated more generally as a structural problem of human thought: If it is true that human beings have certain mental faculties, the properties of which enable their thought but necessarily at the same time limit the scope of their thinking, any argument for the truth of a proposition ultimately draws on these very same mental resources. The principles that in fact determine human thought ultimately are used to determine the truth of these very principles that in fact determine human thought – a circular argument, as in the case of morality.

How to escape from this circle?

Referring to a hierarchy of principles does not help. The metaprinciples themselves would need to be evaluated with the resources of human thought whose validity is to be justified, begging the question just asked.

The conclusion to be drawn is a familiar one from human epistemology: There is no way of escaping the limits of human thought. But this does not mean that we have to embrace skepticism and abandon any hope of human insight. One important lesson from the history of skepticism is, after all, that there is no valid skeptical argument proving that what seems to be true to human beings is *not* really true. Any such assertion would lead skepticism into self-contradiction – the familiar self-contradiction of asserting that the proposition that there is no true proposition is

[181] Mikhail, *Elements*, 221 ff. Cf. also Mikhail, *Moral Grammar and Human Rights*, 164, 173, 197, on the relation between a universally shared structure of moral cognition and arguments for the universalism of human rights.

[182] For a sentimentalist version of this problem, Nichols, *Sentimental Rules*, 188.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

itself a true proposition. Given this situation, there is no alternative but to assume that human thought, in theory and morals, does not lead us astray, but reveals something meaningful about the world. This argument is based on trust in human beings' faculties of thought and understanding, which form the ultimate foundation of any serious effort to wrest some kind of insight from the vast swathes of what has not and perhaps never will be understood. This trust seems to be warranted by its results: Both science and practical life show that human thought does have a productive relation to the facts of the world – airplanes fly; vaccines offer protection; computers help us in our tasks. In the moral sphere, a world of justice, solidarity and respect is a rather attractive vision. There is thus no particular reason to mistrust human thought in principle. The specter of Descartes' evil demon having become flesh in the cognitive structures of human beings and frustrating our search for insight should now find its final resting place.[183] The task is, then, to engage in constructive scientific work, while remaining aware of the limits of human understanding, of the fallibility of any theory and of any theoretical stance's permanent need for critical revision.

For the normative force of psychological facts, these arguments lead to the following conclusion: Even if the properties of human cognitive abilities, including ethical cognitive abilities, ultimately set limits to human understanding, no normative proposition is valid or invalid simply because it is based on empirically operative principles rooted in the nature of human cognition. It is not such psychological facts that are decisive for the validity of normative propositions: Rather, a normative proposition convinces as a piece of normative argumentation against which no valid normative doubts can be mustered. Normative arguments ultimately count only insofar as they formulate compelling normative reasons, not because they are rooted in the psychological makeup of human beings, even if these normative arguments themselves ultimately are the offspring of certain empirical properties of human cognition.

If one accepts this conclusion, the need for normative theory-building in the defense of human rights needs to be reasserted. Psychological theory is no substitute for it, nor can psychological theories ultimately challenge the normative case for (or against) human rights.

This result does not render a theory of moral cognition superfluous: As we have seen, important points of a theory of human moral cognition are, first, to provide a rich explanatory theory of human moral judgment and, second, to perform the critical function of dispelling doubts stemming from theories of the mind and its evolution about whether an ethics of human rights is possible. Moreover, this theory

[183] This argument is the secular version of Descartes' argument that there is no *genius malignus*, no evil demon that betrays us, which we discussed in the introduction, cf. introduction Fn 59, 60 and René Descartes, *Meditationes de prima philosophia*, in *Œuvres de Descartes*, Vol. VII, eds. Charles Adam and Paul Tannery (Paris: Léopold Cerf, 1904), I, 16; IV, 6 – a methodological stance to which there is no alternative in science.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press

serves a third and rather intriguing function, one that can be identified as the ultimate perspective of this inquiry: A theory of human moral cognition can show that there may, perhaps surprisingly, be at least a partial convergence between what normative theory tells us is right and just and what humans in actual fact empirically regard to be morally right and just. This is not self-evident: Human beings could, for instance, be the selfish utility maximizers that some theories both past and present imagine them to be, creating a mismatch between the world of moral principle and the facts of human moral cognition. *In a profound sense, then, human beings in the realm of human ethical understanding are what they morally ought to be: beings committed to justice, respect and concern for others.* Given humanity's track record of self-inflicted suffering, this is perhaps a precious source of hope, although certainly limited in scope.

https://doi.org/10.1017/9781316875520.012 Published online by Cambridge University Press