

6

Which Kind of Mind, Which Kind of Morals, Which Kind of Rights?

Was ist der Mensch? konnt' ich beginnen; wie kommt es, daß so etwas in der Welt ist, das, wie ein Chaos, gärt, oder modert, wie ein fauler Baum, und nie zu einer Reife gedeiht? Wie duldet diesen Heerling die Natur bei ihren süßen Trauben?¹

Friedrich Hölderlin, *Hyperion* oder der Eremit in Griechenland

6.1 ETHICS AND THE THEORY OF THE HUMAN MIND

The content of justice and the good forms one prime focus of human reflection about practical matters. Another question that has occupied practical philosophy ever since its beginnings concerns the mental means by which human beings achieve moral cognition. The answers to this question are manifold, mirroring the changing conceptions of what human beings are like. One constant of this debate is the notion that human beings possess a particular faculty through which they are able to gain insight into true normative propositions. One famous example of this is Socrates' daimonion, his inner voice that guided him, taking his decisions about what it was right and wrong to do.² The Stoics reflected on a human faculty of moral understanding – an idea that influenced the natural law tradition they so profoundly inspired in other regards, too.³ In Christianity, Paul's idea of a law written in the human heart already was very influential in Patristic thought: Moral precepts

¹ “What is man? I could begin; ‘how can it be that such a thing is in the world that ferments like a chaos or moulders like a rotting tree, and never grows to ripeness. How can nature suffer this sour grape amid her sweet clusters?’ Friedrich Hölderlin, *Hyperion, or the Hermit in Greece*, trans. H. Gaskill (Cambridge: Open Book Publishers, 2019), 38.

² Plato, *Apologia*, 31c; Vlastos, *Socrates*, 280 ff.

³ Cf. Chryssipus, “Chrysippi fragmenta moralia,” in *Stoicum Veterum Fragmenta*, Vol. III, ed. Hans von Arnim (Munich & Leipzig: K. G. Saur Verlag, 2004), 314, 323; Max Pohlenz, *Die Stoa* (Göttingen: Vandenhoeck & Ruprecht, 1992), 133.

seemed somehow ingrained in human nature.⁴ In later natural law theory, the question persisted. Thomas Aquinas, for instance, believed that synderesis was the means of understanding natural law.⁵ Others assumed a natural light or a *recta ratio*.⁶ In modern times, the thesis of innate ideas gained prominence, not only concerning the foundations of mathematics, scientific insight and logic, but also with regard to practical questions. Descartes did not say much about ethics or natural law, though there are some interesting remarks.⁷ Leibniz, however, developed quite a differentiated theory of moral judgment in which innate moral ideas played an important part.⁸

One influential idea advanced by the Scottish Enlightenment was that human beings have a moral sense that allows them to perceive the moral status of things.⁹ Hume argued that morality is part of humans' mode of thinking because of "some internal sense or feeling which nature made universal in the whole species."¹⁰ Kant argued with practical reason. In his view, the moral law is a "fact of reason," a given property of human understanding.¹¹ Other psychological assumptions supplement this view: For Kant, one central idea is "respect for the law," the ultimate reason why moral precepts matter.¹² According to this doctrine, the human psyche is structured in such a way that the cognition of the moral law gives rise to respect for this law. This respect for the moral law is the reason why morality matters in practical terms and influences both human motivation and ultimately – if other inclinations do not gain the upper hand – human action.¹³

The same cognitive interest motivated authors such as Locke who considered determining the nuts and bolts of the machinery of human thought a central task, even though they were critical of psychological theories that assumed the existence of innate ideas.

In contemporary thought, moral psychology continues to play a role in influential theories of morality. Habermas, for instance, took Piaget's and Kohlberg's theory of

⁴ Rom. 2, 15.

⁵ Aquinas, *Summa Theologica*, I–II, q. 94.

⁶ Grotius, *De Iure Belli ac Pacis*, I, X.

⁷ René Descartes, "Les Passions de l'Âme," in (*Œuvres de Descartes*, Vol. XI, eds. Charles Adam and Paul Tannery (Paris: Léopold Cerf, 1909), 445 ff.

⁸ Leibniz, "Nouveaux Essais," 91 ff. Matthias Mahlmann, "Die geistige Wurzel der Gerechtigkeit – Rationalismus und Epistemologie in Leibniz' praktischer Philosophie," in *Rechts- und Staatsphilosophie bei G. W. Leibniz*, eds. Tilmann Altwickler, Francis Cheneval and Matthias Mahlmann (Tübingen: Mohr Siebeck, 2020), 89 ff.

⁹ Cf. for instance Francis Hutcheson, *An Inquiry into the Original of Our Ideas of Beauty and Virtue* (New York: Garland Pub., 1971), XVIII, 134, 270.

¹⁰ David Hume, "An Enquiry Concerning the Principles of Morals," in *David Hume: Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. Peter Harold Nidditch (Oxford: Oxford University Press, 1975), 173.

¹¹ Immanuel Kant, *Kritik der praktischen Vernunft, Akademieausgabe*, Vol. V (Berlin: Georg Reimer, 1913), 31.

¹² Kant, *Kritik der praktischen Vernunft*, 71 ff.

¹³ Kant, *Metaphysik der Sitten*, 399.

moral development as the basis for his understanding of the ontogenesis of moral thought.¹⁴ A further example is Rawls, who considered the idea of a parallel between Chomsky's concept of universal grammar and the sense of justice, although he did not develop this systematically after encountering substantial critique.¹⁵

This rich tradition of thought about the characteristics and the functionings of the mind and moral understanding comes as no surprise. The question of how human beings acquire moral knowledge and which cognitive mechanisms are used to put this knowledge into practice suggests itself too evidently to have been missed in the past.

Throughout much of this tradition, moral psychology was understood as equally important for the understanding and creation of law as for morality. The faculty of understanding that enabled human beings to cognize natural law or the law of reason was seen either as dealing with norms regarded as directly valid law or as constituting the foundation for the determination of the legitimate content of the law. The inclusion of moral psychology in theoretical reflections about the law thus also comes as no surprise. The discussion that follows consequently represents yet another small piece of reflection on a topic embedded in a rich history of creative thought. Despite its brevity, it hopefully will shed light on some problems of human moral understanding that hitherto have not been fully illuminated.

Some theories assume that human beings have a general learning ability that both allows them and at the same time makes it necessary to learn everything anew, including moral concepts. Some of these theories understand humans as protean beings who are infinitely malleable and whose mental makeup is determined by their social surroundings, as we have seen. It should be noted again that all such theories, despite their constructivist bent, formulate a substantial, empirical psychological hypothesis including assumptions about human nature, assumptions that are true or not, depending on evidence: Their thesis is that the human mind is structured in such a way that it can and indeed must learn everything anew. In this sense, consequently, even radically culturalist or constructivist theories are naturalistic or nativist because they presuppose an empirical thesis about the cognitive foundations of the development of morality in human beings. There is no escape from psychology and theories of the structure of human cognition.

The view that there are species-specific properties of the human mind other than general learning abilities that constitute the preconditions of the development of morality has gained considerable influence, although perspectives often differ radically about what this means in detail and whether this is good news for an egalitarian ethics of human justice and solidarity and for the political and

¹⁴ Jürgen Habermas, *Moralbewusstsein und Kommunikatives Handeln* (Frankfurt am Main: Suhrkamp, 1983), 127 ff.

¹⁵ Rawls, *Theory of Justice*, 46 ff.; Mikhail, *Elements*.

cultural project of human rights in particular, or whether it rather calls these projects into question.

It is easy to discern the reason for the view that such species-specific properties of human beings exist. Everybody agrees that human beings acquire a vast number of norms as they grow up, through their upbringing, education and socialization. These norms are products of their culture. At the same time, it is not really plausible that the complex cognitive capacities forming the foundation of morality are not species-specific: The cultural influences are obvious, “[b]ut the household pets growing up in the same cultural and religious contexts do not thereby become moral beings.”¹⁶ Therefore, it seems quite reasonable to assert that “there is absolutely no question that human children are biologically prepared for the process as well.”¹⁷

The inquiry into the cognitive preconditions of morality is important, regardless of whether we believe that any findings in this area have *normative* consequences in addition to *explaining* some of the cognitive mechanisms involved in moral understanding, or whether we believe that this would mean committing a naturalistic fallacy or neglecting the is/ought dichotomy.¹⁸ There are various reasons for this importance. First, the question of whether facts of moral psychology have normative consequences or not must be informed by what these facts actually are, and so we need to inquire into these facts. Second, the comprehension of moral cognition and its relation to the law is itself a valuable, even indispensable scientific goal if one wants to gain a profound understanding of the nature of morality and law. We clarified this in the Introduction of the present study. Third, the empirical theories of moral cognition very often do have a direct or indirect normative edge, because they influence the way human beings think about the mechanism that directs their volitions and actions. Just because there are good reasons to take the is/ought dichotomy seriously (as will be argued here) does not mean that everybody in fact agrees with or, even if they avowedly agree, adheres to this stance. Misconceived theories may have significant practical impact. What has been called “evoconservatism” – the justification of reactionary political visions by means of evolutionary theory – is just one such example we will discuss below.¹⁹ Finally, we have to answer the question of the relation between what seems right in normative theory and the facts about human moral cognition. To put it in concrete terms: Do human rights demand the impossible of human beings, as some influential voices argue? Thus, as things stand, the critical assessment of theories of human morality and the law by psychology and neuroscience is of great significance. A lack of interest in the

¹⁶ Michael Tomasello, *A Natural History of Human Morality* (Cambridge, MA: Harvard University Press, 2016).

¹⁷ Tomasello, *Natural History of Human Morality*, 134.

¹⁸ Cf. Selim Berker, “The Normative Insignificance of Neuroscience,” 293 ff.

¹⁹ Allen Buchanan and Russell Powell, *The Evolution of Moral Progress* (New York: Oxford University Press, 2018).

cognitive preconditions of human moral understanding is not something moral and legal theorists can afford.

As indicated above, these questions are not of minor concern. As Hannah Arendt correctly observed, a core problem brought to the fore by the horrors of the Holocaust is the question of whether human beings have a faculty of moral understanding that enables them to distinguish between right and wrong, even if this means setting themselves in radical opposition to their environment.²⁰ The assumption that everybody was able to know that Auschwitz was a crime, irrespective of the drums of Nazi propaganda, presupposes at least two things: firstly, that there *are* standards of right and wrong; and secondly, that people have the *ability to understand* them. The first question already occupied us in our discussion of the justification of human rights, with encouraging results. The second must be informed by what we know about human moral cognition, not least because some theories maintain that moral cognition does not facilitate but rather obstructs doing the morally right thing.

6.2 THE EPISTEMOLOGY OF MORAL COGNITION

There are at least two acts of cognition relevant to this inquiry. First, there are concrete assertions of rights along the line of “X has a right to Y.” These can take the form of assertions in concrete circumstances relating to specific people, such as Creusa’s complaint about the violation of her rights by Apollo. They can also take the form of more abstract and general assertions – for example, about the rights of women to sexual self-determination.

Second, there are acts of cognition that concern the justification of such rights assertions. The cognition of the justification of a right is achieved by a series of mental acts, by a set of cognitive judgments. The final judgment has the propositional content “right X is justified” – for example, expressions such as “freedom of speech is worth defending.” This proposition in fact relies on a number of complex reasons. It implies, as explained above, assumptions about the usefulness of a right such as free speech in human communities (for the expression of humans’ personality, the pursuit of truth, the protection of democracy and the like) and – although these are not always made explicit – anthropological claims about the importance of freedom for human beings. In the case of legal rights, further complications arise from their nature as positive law – for instance, as to their justiciability.

One further element of this bundle of predications is of special concern for normative theory: the predication of the *moral* justification of a right. Complex systems of moral legal rights, many details of which are historically contingent, cannot be derived directly from basic moral intuitions. As we have seen, substantial steps are necessary to move from concrete moral judgments to the formulation of

²⁰ Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Penguin Group, 2006), 294 f.

explicit moral human rights. Basic intuitions of justice or impermissible harm are one thing, the formulation and full justification of a norm in the form of an explicit moral human right, let alone in the technical form of a legal fundamental right and its regime of limitations, quite another.

This notwithstanding, making the moral case for human rights forms a necessary element of any justification of human rights.²¹ There are many considerations that speak for human rights, but these rights cannot be legitimate if they are not justified from a moral point of view. Nobody would have taken the UN General Assembly seriously if it had stated that adopting the *Universal Declaration* made good sense even though it was quite an unjust catalogue of rights. A human rights catalogue that does not claim to lay down a just and morally appropriate set of norms is not a proper human rights catalogue.

Legal human rights sometimes appear in very abstract form. More often than not, however, the respective norms are stated in more differentiated terms, including a regime of justified interferences. This is of crucial importance, because this regime of justified interferences determines the true content of the right, as we have seen. The final judgment “right X is justified” in this case thus encompasses intricate arguments not only about the *prima facie* scope of a right, but also about other values that justify restrictions of this right and the degree to which this is possible. In this context, principles of proportionality have become a core element of modern human rights catalogues,²² sometimes buttressed by the protection of the essence of a right.²³ The determination of a human right’s concrete content thus involves the recursive application of normative principles to limit the *prima facie* scope of the right and determine the meaning of the clauses of limitations, not least by weighing and balancing different rights. As explained above, similar considerations apply for moral rights.

The predication of the normative justification of a right in this sense by a moral judgment (as one element of a bundle of highly complex justificatory arguments about human rights) must be a principled mental act. Caprice cannot provide justification. Rather, the application of normative principles to the case evaluated – the human right under consideration – gives rise to the judgment that this right is morally justified (or not). But what are these principles? And how are these principles, used to justify norms such as human rights, justified themselves? These two major questions of practical philosophy lead us straight to the heart of

²¹ For a derivation of a wide range of human rights from basic moral judgments, albeit not from principles of justice and altruism as proposed above and discussed in detail below, John Mikhail, “Moral Grammar and Human Rights: Some Reflections on Cognitive Science and Enlightenment Rationalism,” in *Understanding Social Action, Promoting Human Rights*, eds. Ryan Goodman, Derek Jinks and Andrew Woods (Oxford: Oxford University Press, 2012), 196 ff.

²² Cf. Art. 36 Constitution of the Republic of South Africa, December 18, 1996; Art. 36 para. 3 BV.

²³ Cf. Art. 36 para. 4 BV.

the normative component of the theory of human rights. Accordingly, over the course of the history of ideas, many principles have been formulated that seek to pin down at least some core elements of morality, from Socrates' belief that it is better to suffer injustice than to do injustice to Kant's categorical imperative and beyond. Given what has been said above about the justification of rights, it seems plausible that principles of equality and thus of justice, duties of human care and respect for the worth of human beings play a crucial role in this respect.

It seems likely that these principles also guide moral judgments about concrete claims, the first category of cognitive acts mentioned above, and not only the normative justification of these intuitions following their transformation into an explicit concept of human rights. Such concrete claims are regularly based on intuitions about justice and the moral obligations of others in concrete circumstances that give rise to specific rights and duties. For example, Creusa's rebellion against her abuse implies that what Apollo has done to her is unjust. These kinds of claims, too, originate in the justificatory principles of justice, benevolence and respect.

The problems of the content and justification of the normative principles guiding moral judgment in the context of concrete claims and the justification of rights lead to the core question of interest here. *Is the content of these justificatory principles perhaps dependent on the structure of human thought?* Are these principles possibly even in one way or another determined by the properties of the human mind? Is the nature and structure of human thought a factor that defines the results of reflection – in the moral domain for moral judgments? This is, as indicated above, the Cartesian, Lockean, Humean and Kantian question about the preconditions of the possibility of human moral cognition. Are there such *Verstandesbegriffe*, such “concepts of understanding” in the moral domain, to use a Kantian term? If so – what is their effect? Are the particular nature and structure of human thought crucial to objective, foundational moral insight? Or are they indispensable elements of moral cognition that, however, do not reveal “morality in itself” (to modify another Kantian idea), do not reveal the true, objective morality? Is there thus a parallel to theoretical cognition, in the same way that human thought, in Kant's view, never grasps the nature of the “*Ding an sich*,” the thing in itself?²⁴ Or – a third possibility – do such structures of the mind exist and decisively influence human moral thought but create nothing but cognitive illusions, the moral equivalent to the Müller-Lyer visual illusion mentioned in the Introduction?

These questions lead us to another important challenge to human rights to be considered in this study. A particular line of cognitive psychology and neuroscience has formulated exactly the latter thesis. It holds that the nature and structure of the human mind is indeed decisive for human moral reasoning, but that the moral

²⁴ Immanuel Kant, *Kritik der reinen Vernunft* (2. Ed. 1787), Akademie Ausgabe, Vol. III (Berlin: Georg Reimer, 1911), 16 f., 202 ff., 207.

judgment thus determined yields not insight, but error. Accordingly, the idea of human rights is part of this erroneous reasoning, and thus is not qualified to decidedly influence human affairs. This challenge will be considered over the next sections of this chapter, before we turn our attention to the fertile research in behavioral law and economics and then to a related challenge from evolutionary psychology in Chapter 7. Finally, we will consider other potentially very fertile approaches to the relation between mind and rights in the last chapter of this book.

6.3 THE NEUROSCIENTIFIC ATTACK ON HUMAN RIGHTS: HUMAN RIGHTS AND THE MENTAL GIZMO THESIS

The thesis to be examined here accepts as its general framework the dual-process model of the mind. This model holds that there are two kinds of mental processes: “thinking fast” and “thinking slow,” to use the popularized terms.²⁵ Thinking fast means using heuristics, framing operations or biases to solve everyday problems. These mechanisms are hardwired into the human brain. Humans cannot but use them and do so intuitively and unconsciously.²⁶ Thinking slow means using reflective rationality that abides by standards of logic.²⁷ Thinking fast works well for many aspects of everyday life but is skewed in important regards. Human decision-making is thus not fully rational. It is possible to become aware of the factors that skew human rationality, such as heuristics, framing effects and biases, and overcome their influence through slow thinking. However, this may not necessarily happen, because the control system of slow thinking may fail in its task.²⁸ This dual-process model of the mind has become highly influential beyond psychology, inspiring research in many other fields, not least behavioral law and economics. Its pioneers have also explored the place of moral reasoning within this model.²⁹ The mental gizmo thesis, however, aims to provide more substantial guidance. Most importantly for our inquiry, human rights are prominent topic of the discussion.

The mental gizmo thesis runs as follows: Moral cognition is part of the dual-process mind.³⁰ Deontological judgments are part of fast thinking. There is a mental “gizmo” that yields such judgments involuntarily, unconsciously, as a product of the

²⁵ Kahneman, *Thinking, Fast and Slow*.

²⁶ Kahneman, *Thinking, Fast and Slow*, 19 ff.

²⁷ Kahneman, *Thinking, Fast and Slow*, 19 ff.

²⁸ Kahneman, *Thinking, Fast and Slow*, 39 ff. This idea is the origin of nudging, the idea that one can systematically exploit these factors for the benefit of others in the framework of a “libertarian paternalism.”

²⁹ Cf. for a summary of the research Daniel Kahneman and Cass R. Sunstein, “Cognitive Psychology of Moral Intuitions,” in *Neurobiology of Human Values*, eds. Jean-Pierre Changeux, Antonio Damasio, Wolf Singer and Yves Christen (Berlin: Springer, 2005), 91 ff., arguing that indignation is key to explaining the outrage heuristic, the centrality of harm, the role of reference states, moral framing and the act–omission distinction.

³⁰ Greene, *Moral Tribes*, 15, 105 ff.

fast, automatic and emotional operations of the human mind. These deontological judgments are like heuristics or biases: They are useful in certain respects but should be disregarded as general guides for moral judgments because they systematically skew human moral rationality.³¹ The mental gizmo thus causes “moral illusions”³² (in the same way that the Müller-Lyer illusion causes visual illusions), and one prime example of a product of its operations is Kant’s principle of humanity that one should never use other human beings merely as means and not as ends.³³ This in itself is already important for the topic of human rights, because this principle is widely regarded as an important element of the concretization of guarantees of human dignity in various national, international and supranational legal systems.³⁴ Human dignity, in turn, is a constitutive part of the whole architecture of human rights. If human dignity is a moral illusion, then this architecture starts to totter. However, the mental gizmo thesis goes even further than this. Human rights as such are themselves seen as products of the mental gizmo, useful as rhetorical devices and exploitable for good causes but without any claim to rationality as such and often quite harmful in their effects.³⁵ Instead, for truly rational moral thinking one needs to resort to utilitarianism. Utilitarianism constitutes slow thinking, which is what should govern human moral reasoning in the last instance.³⁶

What is the evidence for the mental gizmo thesis? Some of many variants of the familiar trolley problem form the starting point of the analysis.³⁷ In the so-called bystander case, a bystander can turn a switch so that a runaway trolley is redirected with the consequence that it kills not five persons on one track but one person on another track. In the variant called the “footbridge case,” a person is thrown down onto a track from a bridge to stop a runaway trolley, thus saving the five people working on the track. Proponents of the mental gizmo thesis argue that a proper analysis of cases where deontological judgments seem to be at play, because the observers judge it not to be permissible to sacrifice the life of one person to save five (footbridge case), shows that this judgment is in fact determined by hardwired emotional reactions.³⁸ This analysis is supported by neuroimaging studies. These

³¹ Greene, *Moral Tribes*, 132 ff.

³² Greene, *Moral Tribes*, 252.

³³ Greene, *Moral Tribes*, 105 ff., 115.

³⁴ Cf. Mahlmann, *Human Dignity and Autonomy*, 370 ff.

³⁵ Greene, *Moral Tribes*, 302 ff.

³⁶ Greene, *Moral Tribes*, 290 ff.: Utilitarianism is called “deep pragmatism.” Greene sums this thesis up: “The Central Tension principle: Characteristically deontological judgements are preferentially supported by automatic emotional responses, while characteristically consequentialist judgements are preferentially supported by conscious reasoning and allied processes of cognitive control,” Joshua Greene, “Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)Science Matters for Ethics,” *Ethics* 124, no. 4 (2014), 699. For an endorsement cf. e.g. Peter Singer, “Ethics and Intuitions,” *The Journal of Ethics* 9, no. 3/4 (2005), 331 ff.

³⁷ Greene, *Moral Tribes*, 105 ff.

³⁸ Greene, *Moral Tribes*, 119 ff. Greene, “Why Cognitive (Neuro)Science Matters for Ethics,” 698 states that mechanisms of fast thinking do not need to be “hard wired.” In his discussion,

studies, it is argued, show that when deciding on these cases, the ventromedial prefrontal cortex (VMPFC) is active, an area of the brain associated with the production of emotion.³⁹ The VMPFC is active because these cases involve the agent directly; they are “personal” and thus trigger emotional reactions.⁴⁰

In other cases that are “impersonal,” where the judgment reached is different and participants consider it permissible that one person dies and five persons are saved (bystander case),⁴¹ the dorsolateral prefrontal cortex (DLPFC), the center of cognitive control in the brain, is active, showing that this utilitarian judgment is a rational, not an emotional one.⁴² Further evidence is provided, it is argued, by studies showing that if the VMPFC is damaged, utilitarian judgments are made by the participants in both the bystander and footbridge cases.⁴³

From this perspective, deontological arguments are nothing but the post-hoc rationalization of hardwired emotional reactions.⁴⁴ This is not only the “secret joke of Kant’s soul,”⁴⁵ but, one might add, the secret joke of the souls of a great many thinkers on moral issues from antiquity to the present. Thus, say, Plato’s dialogues, the *Critique of Practical Reason* and *A Theory of Justice* are all exercises in self-delusion on the part of their authors. Plato’s defense of a nonconsequentialist concept of justice, the formulation of the categorical imperative in its formal and material versions and John Rawls’ nonutilitarian principles of justice are ultimately expressions of the secret workings of the emotional mental gizmo, rationalized post hoc and writ large.⁴⁶ Human rights are part of these post-hoc rationalizations: “Rights’ are nothing short of brilliant. They allow us to rationalize our gut feelings without doing any additional work.”⁴⁷ The many people concerned with human rights, such as lawyers, judges, activists, politicians and – most importantly – the people claiming human rights,

the mental gizmo appears, however, throughout to be “hard wired,” a given of human nature, shared by Kant, Rawls and us.

³⁹ Greene, *Moral Tribes*, 121 ff., Joshua Greene et al., “An fMRI Investigation of Emotional Engagement in Moral Judgement,” *Science* 293, no. 5537 (2001).

⁴⁰ Greene, *Moral Tribes*, 121 ff.

⁴¹ In the bystander case, the bystander can turn a switch so that a runaway trolley is redirected with the consequence that it does not kill five persons on one track but one person on another track.

⁴² Greene, *Moral Tribes*, 120.

⁴³ Greene, *Moral Tribes*, 124 ff. For more studies taken as support for this thesis, Greene, “Why Cognitive (Neuro)Science Matters for Ethics,” 700 ff.

⁴⁴ Greene, *Moral Tribes*, 298 ff., 300: “The moral equivalent of confabulation is *rationalization*. The confabulator perceives himself doing something and makes up a rational sounding story about what he’s doing and why. The moral rationalizer *feels* a certain way about a moral issue and then makes up a rational-sounding justification for that feeling” (emphasis in original).

⁴⁵ Greene, *Moral Tribes*, 301, quoting Nietzsche: “In other words, Kant has the same automatic settings as his surrounding tribespeople. But Kant, unlike them, felt the need to provide esoteric justifications for their ‘popular’ prejudices.”

⁴⁶ On Rawls, *Theory of Justice* as being another product of rationalization of the working of the mental gizmo, Greene, *Moral Tribes*, 333.

⁴⁷ Greene, *Moral Tribes*, 301 ff., 302.

fighting for these rights to be respected, hoping, sometimes desperately, for their protection, are all under the spell of a “moral illusion.” Given the importance of human rights in practice, this “moral illusion” has massive consequences that dwarf any practical effects that other elements of a skewed rationality, such as framing effects, may possibly have. These consequences are quite harmful: “Rationalization is the great enemy of moral progress, and thus of deep pragmatism.”⁴⁸

The mental gizmo thesis is part of a wider trend in contemporary neuroscience that can be called neuroscientific emotivism. This trend uses the means of cognitive psychology to defend traditional emotivism’s idea that human morality is no more than the expression of certain emotions of appraisal and disgust. It represents a very radical critique of human rights: Its punchline is not to deny the cognitive reality and impact of deontological judgments and the idea of human rights, but radically to reinterpret their status and meaning. In this view, deontological judgments and human rights are not manifestations of practical reason, but on the contrary are some of the causes of human moral irrationality. This irrationality has such far-reaching detrimental consequences that it needs to be overcome for the sake of the survival of the species. Only utilitarianism, only slow thinking, it is maintained, is able to solve the great problems of humanity and transcend the parochial moralities of the human tribes created by the mental gizmo,⁴⁹ “to free philosophers from the ups and downs of their automatic settings,”⁵⁰ and to free all the rest of us, of course, who also often suffer from the “moral illusion” of human rights.

How to answer this interesting challenge, which presumably will be paradigmatic for quite a few discussions to come?

6.4 THE MENTAL GIZMO THESIS RECONSIDERED

One fundamental problem is that the analysis of the trolley problems (and its predecessors, such as those formulated by the German criminal lawyer Hans Welzel) underlying the mental gizmo thesis is deficient.⁵¹ The cases that are taken to prove the workings of emotional gut reactions (instances of the footbridge case) in

⁴⁸ Greene, *Moral Tribes*, 301.

⁴⁹ Greene, *Moral Tribes*, 289 ff.

⁵⁰ Greene, “Why Cognitive (Neuro)Science Matters for Ethics,” 720.

⁵¹ The usual reference for the origin of this problem is Philippa Foot, “The Problem of Abortion and the Doctrine of the Double Effect,” *Oxford Review* 5 (1967); Judith Jarvis Thomson, “The Trolley Problem,” *Yale Law Journal* 94 (1985); Judith Jarvis Thomson, “Killing, Letting Die, and the Trolley Problem,” *The Monist* 59, no. 2 (1976). In fact, core elements of the problem were already formulated previously, cf. Hans Welzel, “Zum Notstandsproblem,” *Zeitschrift für die gesamte Strafrechtswissenschaft* 63 (1951), 47, 51 (with trains, not trams as in Foot). It is interesting that almost no reference to this earlier paper is found in current international debates, although Welzel was a well-known criminal lawyer and the article has been a standard in German-language criminal law discussions ever since its publication. It would be interesting to know whether Foot was familiar with it. Welzel discusses the concrete case of Nazi doctors’ culpability for the mass murder committed as part of the so-called euthanasia program.

fact show something quite different, namely the relevance of the means/side effect distinction in explaining the empirical patterns of moral evaluation observed and, more precisely, the prohibition of the instrumentalization of human beings.⁵² A proper analysis of the trolley problems therefore in fact vindicates the relevance of Kant's famous formula of the principle of humanity and thus the relevance of a crucial element of the idea of human rights. In addition, it is far too rash to conclude that a utilitarian reasoning is at play in the other cases where choosing the death of one person is taken to be permissible if the alternative is the death of five (or more victims, as in Welzel's initial train case). Thinking that it is permissible to choose the lesser of two unavoidable evils, though perhaps feeling at the same time that this is a tragic choice, is one thing, endorsing utilitarianism in the sense that it is always permissible simply to count lives quite another. The analysis of the trolley problems needs to be much more sophisticated and transcend such simple dichotomies.⁵³

⁵² The most advanced analysis of the trolley problem is provided by Mikhail, *Elements*, including the introduction of formal modes of representation of structures of the human actions evaluated, such as act trees, cf. Mikhail, *Elements*, 118 ff. Greene discusses Mikhail in some detail, cf. Greene, *Moral Tribes*, 230 ff. As Mikhail correctly argues, the core of the footbridge scenario is the use of the patient as a means. More precisely, in Mikhail's analysis, a battery (not the death) is a means to stop the trolley, not just the foreseen side effect of an action taken to save five, Mikhail, *Elements*, 123 ff. One can question whether the means to stop the trolley is just a battery or the death of the person, and this may be a significant difference. The means–ends distinction is, however, of crucial importance in any case. An explanation relying on the alternative account of the personal character of the action (pushing the patient over the bridge) is unconvincing, given scenarios that remove this personal element, cf. Mikhail, *Elements*, 109 (drop man) and passim; Mikhail, "Moral Grammar and Human Rights," 183. Empirical evidence quoted by Greene, *Moral Tribes*, 215 ff. is inconclusive, given empirical evidence that on the contrary buttresses the relevance of the means/side effect distinction, Mikhail, *Elements*, 319 ff. and other studies with quite different results, cf. n. 63. The "loop case" does not call these findings into question. On this case adduced by Greene, *Moral Tribes*, 220 ff. as a counterexample, Mikhail, *Elements*, 336 ff., 359. In the latter case, the issue of what counts as the origin of data (i.e. the problem of the criteria for the selection of judgments taken as evidence) is of crucial importance, because some scenarios can be so complicated that their moral point becomes obscure; see the remarks on "considered judgments" below. The "modular myopia hypothesis" that Greene, *Moral Tribes*, 224 ff. formulates as an explanation and according to which the emotional (deontological) cognitive subsystem is blind to harmful side effects is therefore not convincing. Kahnemann and Sunstein, "Cognitive Psychology of Moral Intuitions," 102, are consequently mistaken in evading the question of whether there are principled interpretations of the trolley problems. This question is decisive for understanding how moral cognition operates.

⁵³ Cf. for an attempt to move forward in this respect Mikhail, *Elements*; on alternative explanations for the reactions to trolley problems, cf. Guy Kahane and Nicholas Shackel, "Methodological Issues in the Neuroscience of Moral Judgment," *Mind & Language* 25, no. 5 (2010); Guy Kahane, "Sidetracked by Trolleys: Why Sacrificial Moral Dilemmas Tell Us Little (or Nothing) about Utilitarian Judgment," *Social Neuroscience* 10, no. 5 (2015): 555 f.; Guy Kahane et al., "Beyond Sacrificial Harm: A Two-Dimensional Model of Utilitarian Psychology," *Psychological Review* 125, no. 2 (2018): 132 ff. As Zamir, *Law, Psychology, and Morality*, 188, has pointed out, given sufficiently large "net good outcomes," subjects may reason as if they were consequentialists. One important question in this context is whether there is a significant moral difference between cases where harm is inflicted on others and are

Another problem of the mental gizmo thesis is that it is self-refuting. The reasons for this are as follows. Utilitarianism in all its classical and contemporary rule- or action-based variants is guided by the principle of utility: Any rule or action that creates the greatest happiness of the greatest number is normatively justified.⁵⁴ The foundation of this principle is the idea that each person's happiness counts equally because the persons who are happy count equally. This egalitarian thrust of utilitarianism explains its persistent attraction and forms the core of what is truly admirable in this line of thought.⁵⁵

This maxim rests on two pillars: the equality of persons and the prescriptive principle that equal persons ought to be treated equally. Only given these two propositions does it follow that everyone's happiness shall count equally, as presupposed by the principle of utility. Consequently, the obligation to respect the equality of equal persons by counting their happiness equally in the utilitarian calculus is not the consequence, but the precondition of utilitarianism. The obligation to respect the equality of persons is not and indeed cannot be derived from the application of the principle of utility: Being its foundation, the obligation to treat equals equally cannot be the consequence of that latter principle. Rather, the obligation to respect the equality of persons is a principle that forms the nonconsequentialist normative precondition of consequentialism. There is thus a deontological residue at the core of utilitarianism, because the principle of the obligatory equal treatment of equals (the reason for being obliged to value everybody's happiness equally, as presupposed by the principle of utility) is the foundation of any utilitarian argument.⁵⁶

widely regarded as (morally and legally) justified and cases such as the bystander case in the trolley problem. The perception that the latter case involves tragic choices implies that there is such a difference. Shaun Nichols and Ron Mallon, "Moral Dilemmas and Moral Rules," *Cognition* 100, no. 3 (2006) distinguish between broken (apparently explicit) rules and permissibility all-things-considered. This distinction raises interesting and important issues. However, the issue discussed here as an incident of tragic choices arises precisely in the case where an action is regarded as permissible all-things-considered and consequently has to be accounted for.

⁵⁴ Cf. Bentham, *Introduction to the Principles of Morals and Legislation*, I, n. 1.

⁵⁵ Cf. Bentham, *Introduction to the Principles of Morals and Legislation*, I, 13 n. d; Mill, "Utilitarianism," V, 200 and 198, on the principle of equality: "It is involved in the very meaning of Utility, or the Greatest Happiness Principle. That principle is a mere form of words without rational signification, unless one person's happiness, supposed equal in degree (with proper allowance made for kind), is counted exactly as much as another's." Cf. Chapter 4.

⁵⁶ There is the argument that utilitarianism is not about the equality of persons but only about the equality of happiness, cf. for instance Hart, *Essays on Bentham*, 98. In this case, too, a prescriptive rule is implied, however – the rule that you ought to treat equal matters of fact (e.g. happiness) equally by including any equal amount of happiness on an equal footing in the utility calculus. Moreover, to value any degree of happiness equally seems to imply the equality of the persons experiencing it. The happiness of a king is therefore not worth more than the happiness of a beggar. Greene realizes that the foundation of utilitarianism is such a principle of equality, Greene, *Moral Tribes*, 163, 170: "The second utilitarian ingredient is impartiality,

This analysis of utilitarianism leads us to the central catch of the mental gizmo thesis. The catch is this: Either it is true that utilitarianism is slow thinking (as utilitarianism presupposes deontological principles of equality, it follows in that case that such deontological principles are slow thinking, too, because these deontological principles of equality are the normative core of what is regarded as slow thinking) or it is true that such deontological principles are fast thinking (as utilitarianism presupposes these deontological principles of equality, which is fast thinking, utilitarianism must constitute fast thinking, too). In either case, the mental gizmo thesis is refuted by internal contradictions.⁵⁷

the universal essence of morality, that's distilled in the Golden Rule. Having added this second ingredient, we can summarize utilitarianism thus: Happiness is what matters, and everyone's happiness counts the same."

⁵⁷ Greene does not provide any justification of the foundational principle of utilitarianism that he identifies (impartiality, the Golden Rule, cf. Greene, *Moral Tribes*, 163, 170) and draws no consequences from it, even though this evidently calls his analysis in question. What is "impartiality" or the "Golden Rule"? Slow thinking? Why is it not just another one of those ethical principles that he derides as dangerous rationalizations of gut reactions? What is the difference in this respect between "impartiality" or the "Golden Rule" and, say, the categorical imperative or Rawls' principles of justice? This question needs to be asked not least because the point of the categorical imperative or Rawls' principles of justice is precisely to reach "impartiality" by universalization or by deliberation behind a "veil of ignorance." The "Golden Rule" is certainly related to one of the key ideas of the categorical imperative, namely the idea of universalization. Greene's argumentation is thus circular: Principles such as the categorical imperative are criticized as dangerous rationalizations of emotional gut reactions on the basis of principles such as the Golden Rule that are in fact similar to the very principles criticized.

Greene, "Why Cognitive (Neuro)Science Matters for Ethics," 717, claims that his argumentation "favors consequentialist approaches to moral problem solving, ones aimed solely at promoting good consequences, rather than deontological approaches aimed at figuring out who has which rights and duties, where these are regarded as constraints on the promotion of good consequences." The critique developed here can be stated in the terms of constraining rights and duties, too: The principle of equal treatment ("impartiality") at the foundation of utilitarianism implies that all human beings have a *right* that their happiness should count equally and that others have a *duty* to count their happiness equally. Only given these normative constraints is the application of the principle of utility legitimate for utilitarianism. The doctrine of rights and duties is thus criticized using a doctrine that itself relies on very important rights and duties. Again, the circle is complete. Greene even states that utilitarianism presumably rests on an "affectively based evaluative premise," Greene, "Why Cognitive (Neuro)Science Matters for Ethics," 724. That this "affectively based evaluative premise" is supposed to be a high-level intuition does not change the fact that – by relying on such an *affectively* based premise to replace emotional gut reactions – the theory has become quite visibly inconsistent. It is a useful exercise to reconsider on the basis of this observation the meaning for Greene's argument concerning the studies (interesting as they are) listed in Greene, "Why Cognitive (Neuro)Science Matters for Ethics," 701 ff. Another example of these kinds of contradictions is the following statement, referring to emotional and cognitive neural structures: "It seems that healthy humans engage both responses and that there is a higher-order evaluation process that depends on the ventromedial prefrontal cortex, a structure that across domains attaches emotional weight to decision variables. In other words, the brain seems to make both types of judgement (deontological and consequentialist) and thus makes a higher-order judgement about which lower-order judgement to trust, which may be viewed as a kind of wisdom (reflecting virtue or good character)," Joshua Greene et al., "Embedding Ethical

When thinking about these kinds of claims advanced by certain psychological theories, we should bear one thing in mind: While neuroimaging studies have provided many fascinating results in recent years, as yet there are limits to these insights. These limits are not only the product of the constraints imposed and problems implied in the methods of neuroimaging, such as the level of the spatial and temporal resolution of the methods applied, “voodoo correlations,”⁵⁸ circular analysis,⁵⁹ the effects of the statistical “smoothing” of results and so on.⁶⁰ For example, the way that complex mental phenomena are realized in the human brain still has not been clarified. One question to ask in this context is whether the long-dominant focus on the localization of functions in the brain is fruitful or whether the research on patterns of activation is not more promising.⁶¹ This has

Principles in Collective Decision Support Systems,” *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* (2016), 4148. It is noteworthy that the last-instance arbiter of human moral judgment in this passage is a region of the brain that is associated with emotional, deontological judgment, albeit here couched in the terms of some kind of virtue ethics. It is unclear how this can be reconciled with Greene’s theses on fast and slow moral thinking.

Kahane et al., “Beyond Sacrificial Harm,” convincingly argue that two elements of utilitarianism need to be distinguished: The element of impartial beneficence and the element of permissible or even required instrumental harm (i.e. harm inflicted on persons for the greater benefit of others). It is, however, also crucial to include in the analysis of moral cognition the insight highlighted here that the principles of equality at the heart of utilitarianism do not distinguish utilitarianism from deontology because they are utilitarianism’s deontological base. The principles of equality together with other deontological principles like the prohibition of instrumentalization and its deeper normative foundations may also help us to understand such findings as reported in Kahane et al., “Beyond Sacrificial Harm,” 152 ff., 155: While impartial beneficence is associated with greater emphatic concern, with helping with greater generosity, with greater welfare-based concern for the environment and with greater identification with the whole of humanity, these measures were either not or negatively associated with the moral acceptance of inflicting instrumental harm (the latter, however, positively associated with psychopathy, *ibid.* 151). These findings may be best understood as the expression of an underlying universalist, egalitarian, emphatic, other-regarding moral framework within which individuals are respected and valued for their own sake.

These problems affect other studies that build on these assumptions, too – for example, on the effects of variation in the oxytocin receptor gene on moral judgment, cf. Regan Bernhard et al., “Variation in the Oxytocin Receptor Gene (OXTR) Is Associated with Differences in Moral Judgement,” *Social Cognitive and Affective Neuroscience* 11, no. 12 (2016). Another example is Kahnemann and Sunstein, *Cognitive Psychology of Moral Intuitions*. The study identifies morality with emotional intuitions but states that such intuitions can be transformed by conscious reasoning, *ibid.* 92, 103. Which moral principles are the basis of this transformation by conscious reasoning?

⁵⁸ Cf. Edward Vul et al., “Puzzling High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition,” *Perspectives on Psychological Science* 4, no. 3 (2009).

⁵⁹ Cf. Nikolaus Kriegeskorte et al., “Circular Analysis in Systems Neuroscience: The Dangers of Double Dipping,” *Nature Neuroscience* 12 (2009).

⁶⁰ Cf. e.g. Russell Poldrack, “The Future of fMRI in Cognitive Neuroscience,” *Neuroimage* 62, no. 2 (2012), 1216 f. on some statistical problems.

⁶¹ Cf. e.g. Poldrack, “The Future of fMRI in Cognitive Neuroscience,” 1216, 1217 f., on the move away from “blobology” to pattern analysis: “[T]he goal of finding blobs in a specific region can drive researchers into analytic gymnastics in order to find a significant blob to report. However,

consequences for the interpretation of patterns of neuronal activity that form the basis for the hypotheses of neuroimaging studies, because it is far from clear what a given observed neuronal activity really means. This is due not least to a classical problem of neuroimaging studies: reverse inference. One cannot conclude from the fact that a brain region is active when performing a certain task that whenever this brain region is active, this cognitive task is being performed. This is because a single brain region may perform many tasks – interacting with other areas of the brain, for instance. Reverse inference can produce research hypotheses but cannot provide conclusive evidence of how a mental function maps onto the brain.⁶² The fact that a brain region – say, the DLPFC – is active when performing certain tasks involving cognitive control thus does not entail that whenever the DLPFC is active, cognitive control tasks are being performed. The same is true of the VMPFC and emotional reactions. Consequently, if they stand the test of further research, the results of neuroimaging studies such as the ones referred to are in no way proof that deontological morality is the expression of emotion. In addition, there are many competing empirical findings,⁶³ interestingly – if critically interpreted – also in studies

for the last few years the most interesting and novel research has focused on understanding patterns of activation rather than localized blobs. The appreciation of patterns is happening at multiple scales. At the systems (whole-brain) scale, the modelling of connectivity and its relation to behaviour continue to grow. ... I think the jury is still out on how well fMRI can ever characterize neuronal connectivity; as we outlined in Ramsey et al. (2010), there are a number of fundamental challenges in using fMRI to characterize causal interaction between brain regions." For a similar assessment (from phenology to network theories) cf. Lutz Jäncke, *Kognitive Neurowissenschaften* (Bern: Hogrefe, 2013), 71 ff.

⁶² Cf. Russell Poldrack, "Can Cognitive Processes Be Inferred from Neuroimaging Data?" *Trends in Cognitive Science* 10, no. 2 (2006); Russell Poldrack, "Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding," *Neuron* 72, no. 5 (2011); Poldrack, "The Future of fMRI in Cognitive Neuroscience," 1216, 1218 f., on the (difficult) task of finding "a region that is engaged *selectively*, such that activation of the region is actually predictive of the mental process" (emphasis in original), as a precondition for overcoming the problems of reverse inference; Russell Poldrack, *The New Mind Readers* (Princeton, NJ: Princeton University Press, 2018), 20 ff.

⁶³ The insights gained through experimental work should not be overestimated. To take an example: There are studies suggesting that patients with lesions to the VMPFC are more vindictive in ultimatum games than normal subjects (cf. Michael Koenigs and Daniel Tranel, "Irrational Economic Decision-Making after Ventromedial Prefrontal Damage: Evidence from the Ultimatum Game," *The Journal of Neuroscience* 27, no. 4 (2007): 951), which seems to imply a less "utilitarian" and more "deontological," fairness-oriented outlook, while the same brain defect is used as an argument for the thesis that "deontological" judgments are emotional reactions emanating from the VMPFC; see above. This is not really convincing as "[s]uch patients exhibit both an abnormal utilitarian *and* an abnormal deontological tendency!" Kahane and Shackel, "Methodological Issues in the Neuroscience of Moral Judgment," 573 (emphasis in original). On the same problem cf. Aaron Duke and Laurent Bègue, "The Drunk Utilitarian: Blood Alcohol Concentration Predicts Utilitarian Responses in Moral Dilemmas," *Cognition* 134 (2015): 121, 124: "Alcohol intoxication is associated with increased emotional reactivity and selective attention towards emotional cues, which according to Greene's dual process conceptualisation, should lead to increased deontological (non-utilitarian) inclinations, the opposite of what was observed here." On another study with the result that "there is little

coauthored by Greene himself.⁶⁴ In light of what has been said, there is thus ample reason to reinterpret such findings and what they tell us about the workings of the mind, considering closely the analysis above of cases such as trolley problems and what they mean for plausible theories of moral judgment.

In this context, we should stress that no one doubts that emotions are a central part of moral evaluation. The question is, however, whether such emotions constitute moral (deontological) evaluation, as emotivists contend, or play a

relation between sacrificial judgements in the hypothetical dilemmas that dominate research, and a genuine utilitarian approach to ethics," Guy Kahane et al., "Utilitarian' Judgments in Sacrificial Moral Dilemmas Do Not Reflect Impartial Concern for the Greater Good," *Cognition* 134 (2015): 193. A related debate appears in Jorge Moll and Ricardo de Oliveira-Souza, "Moral Judgments, Emotions and the Utilitarian Brain," *Trends in Cognitive Science* 11, no. 8 (2007); Joshua Greene, "Why Are VMPFC Patients More Utilitarian? A Dual Process Theory of Moral Judgement Explains," *Trends in Cognitive Science* 11, no. 8 (2007); Jorge Moll and Ricardo de Oliveira-Souza, "Response to Greene: Moral Sentiments and Reason: Friends or Foes?" *Trends in Cognitive Science* 11, no. 8 (2007); another move is to reinterpret findings on the trolley problem in the framework of "intuitive/counterintuitive judgments," cf. Guy Kahane et al., "The Neural Basis of Intuitive and Counterintuitive Moral Judgment," *Social Cognitive and Affective Neuroscience* 7, no. 4 (2011); on some reinterpretations of the role of the VMPFC in moral decision-making, Joshua Greene, "The Cognitive Neuroscience of Moral Judgement and Decision Making," in *The Cognitive Neurosciences*, eds. Michael Gazzaniga and George Mangun (Cambridge, MA: MIT Press, 2014), 1017 ff. Cf. for some more possible functions of brain regions associated with moral judgment, Joanna Demaree-Cotton and Guy Kahane, "The Neuroscience of Moral Judgement," in *The Routledge Handbook of Moral Epistemology*, eds. Aaron Zimmermann, Karen Jones and Mark Timmons (New York: Routledge, 2019), 84–104, 92 ff. One can conclude from these debates that moral and legal theory is urgently needed to create a theoretical framework in which experimental findings can be designed and interpreted more successfully, including a much more finely grained account of the structure and content of morality and the role of emotions as a precondition and consequence of moral judgment than is sometimes used in these experiment-based debates.

⁶⁴ Cf. Karen Huang, Joshua Greene and Max Bazerman, "Veil-of-Ignorance Reasoning Favors the Greater Good," *Proceedings of the National Academy of Sciences* 116, no. 48 (2019). The authors investigate the effect of prior veil-of-ignorance reasoning on subsequent moral judgments about moral dilemma situations, including the footbridge case. They find that such prior veil-of-ignorance reasoning increases the number of people opting for pushing the man off the bridge to stop the trolley in the footbridge case. The study does not discuss an evidently important factor, namely that self-interest may dominate over moral considerations that still continue to influence human beings – for instance, through the indirect effects of a bad conscience. If you imagine yourself in veil-of-ignorance conditions as one of the persons on the track (a 5 in 6 chance) compared to the person pushed onto the track (1 in 6 chance), you may opt for pushing the person because of fear for your life but still find it morally problematic to do so. It is a lot to ask others to opt to endanger their own life for moral reasons. Criminal law often accounts for such cases with elements of exculpation (not justification). Interestingly, the number of participants finding it morally acceptable to push the person still remains low – 38 percent in comparison to 24 percent in control conditions. The majority, thus, even under the threat of their own death, would not opt to instrumentalize the person on the bridge to save themselves. These results confirm the importance of this principle of noninstrumentalization, which is at the heart of the footbridge case, as we have seen. The other cases imply further problems that the studies do not address – the earthquake case raises the problem of threshold deontology and the autonomous vehicle cases lead to the question of the effect that the responsibility for creating a risk (using an autonomous vehicle) has on the decision about the distribution of risk.

different role. Again, the theoretical power of imagination evident in emotivist accounts seems far too limited to explain the complexity of the human moral world.⁶⁵ One important point is these accounts' analytical failure to distinguish emotions that are the *consequences* of moral judgment from emotions that *constitute* moral judgment. Consider the case of outrage after witnessing a grave injustice. Here, the cognition of injustice is the precondition of and thus not identical to the feeling of moral outrage. The fact that moral sentiments arise is predicated upon certain preconditions, such as the indignation about an injustice upon the unequal treatment of two persons who are equal in the respects that are relevant in the situation at hand. Only if the agent thinks that these elements of an immoral act exist will the respective feeling ensue. This is why such feelings disappear when one understands that the facts actually were different – for example, that there was indeed a relevant difference between the persons being unequally treated.

For the study of the neurological basis of morality, this means that it would be very surprising if brain functions relevant for human sentiment (whatever they turn out to be) were not engaged when an agent evaluates a situation in moral terms. However, these moral sentiments do not constitute moral judgment, but are the consequence of an analysis of structural elements of the action (agency, patients of action, intentions and their kind and object, relations of equality, etc.).

In addition, as our analysis of the concept of fundamental or human rights and the history and justificatory theory of human rights have shown, the preconditions of the predication of rights are highly complex both in form and in substance. Equating such complex judgments with an emotional gut reaction does not seem a very promising approach to the matter.

All of this is nothing more than a reminder that the interpretation of empirical data is dependent upon theory: Data only have meaning within a theoretical framework. In our concrete case, the value of neuroimaging studies of the neurophysiological basis of moral judgment is dependent on the merits of the theoretical framework in which these studies are developed. If this framework is deficient, the interpretation of the data will be insufficient, too.⁶⁶

In addition, identifying rationality with utilitarianism seems somewhat naive.⁶⁷ The question is: Why should the scope of practical reason, to use a traditional term,

⁶⁵ Cf. Matthias Mahlmann, "Ethics, Law and the Challenge of Cognitive Science," *German Law Journal* 8, no. 6 (2007): 586 ff.; Pardo and Patterson, *Minds, Brains, and the Law*, 58 on emotions accompanying moral judgments.

⁶⁶ Current debates often criticize "armchair philosophers" for their naivete, sometimes with good reason. One should, however, not overlook the deficits of some of the experimental work, which would benefit a great deal from more preliminary theoretical work (cf. n. 63).

⁶⁷ Greene, "Why Cognitive (Neuro)Science Matters for Ethics," 696, posits that slow thinking is "a general-purpose reasoning system, specialized for enabling behaviors that serve long(er) term goals." This overlooks the theory of justice's insight that equality as a normative principle is not the same as general rationality, cf. Gosepath, *Gleiche Gerechtigkeit*. Berker, "The Normative Insignificance of Neuroscience," 293 ff., underlines correctly that a normative argument is

not be wider? In the history of thought, it was the default assumption that human thought is made up not only of some kind of instrumental rationality, but that there are other, qualitatively different yardsticks, most importantly those of justice and moral goodness. This is the common denominator of much of the greatest thought on these matters. Why should this be wrong? Why are deontological principles a priori not rational, or reasonable, if you prefer? What is intrinsically better about the principle of utility (forgetting for a moment its deontological foundations) as compared to the prohibition of instrumentalization, the principle of the justice of the equal treatment of equals or the obligation to care for others?⁶⁸ A strange impoverishment of the richness of human thought is at play in such theories that fails to live up to the insights of practical philosophy and legal theory.

It is important to emphasize that the mental gizmo thesis is not a necessary consequence of the dual-process model of the mind as such. It is possible to believe that this model describes an important aspect of human cognition without finding the mental gizmo thesis convincing. Deontology could be part of slow thinking; there is no a priori reason why this could not be the case. Consequently, not being convinced by the mental gizmo thesis does not say anything about the value of the dual-process model of the mind. The mental gizmo thesis is just an implausible thesis developed within this model of the human mind.

Nor do neuroscientific or psychological approaches to ethical and legal issues as such wed us to a certain perspective on the cognitive origin of ethics and law. In particular, nothing in the theory of mind, in neuroscientific research or in psychology forces us to develop an impoverished account of human practical thought. The answer to psychological skepticism is consequently neither to ignore neuroscience and psychology, nor to escape into a normative theory where psychology, whatever it says, simply does not count. Rather, the answer is to develop a substantial concept of human moral cognition as an element of a wider theory of human rights that is more plausible than its skeptical alternatives.

6.5 RIGHTS AND BEHAVIORAL SCIENCE

One large and creative area of research concerns the analysis of law on the basis of behavioral science. Classical law and economics operating within the parameters of rational choice theory assume that human beings are rational maximizers of their expected utility in absolute terms. This assumption is not only meant to be

needed to justify the conclusion that deontological judgments use morally irrelevant criteria, whereas utilitarianism does not. These normative criteria cannot be drawn from neuroscientific research as such, *ibid.* 326.

⁶⁸ Greene, *Moral Tribes*, 136: "Reasoning, as applied to decision making, involves the conscious application of decision rules." Why, according to this rather broad definition, is the principle of utility (or the principle of impartiality or the "Golden Rule," see n. 56) a candidate for reasoning but the categorical imperative and Rawls' principles of justice are not?

descriptive, but often forms part of a normative theory: Justified decisions must be based on such rational choices.

Behavioral law and economics take their impetus from systematic differences between the assumptions of rational choice theory and actual human decision-making.⁶⁹ One influential element is prospect theory.⁷⁰ The central thesis of prospect theory is that people make choices not on the basis of utility determined by a final state, but rather in the light of changes relative to their specific situation or reference point. Losses and gains are central elements of perceived utility. People are loss averse – the utility of gains is perceived to be smaller than the disutility of losses, even if gains and losses are equal in absolute terms. When making decisions, people use heuristics, general decision principles that may work well in many cases but may yield results that are irrational – for example, the representational heuristics that ascribe properties, sometimes falsely, to individuals because of the class to which they belong. Judgments are also influenced by biases, such as the omission bias – all things equal, people prefer not to act. Framing effects are a third example: The way a decision situation is framed – for instance, in terms of losses and not of gains – influences the decision, even though the outcome is the same in absolute terms.

Within this conception of the mind, theories of rights are developed. One such theory explores the possible effects of loss aversion on conceptions of rights. The distinction between civil and political rights on the one hand and social and economic rights on the other, or between so-called first- and second-generation rights, is of interest in this context. A key question is whether or not there are reasons to protect both kinds of rights equally or whether only the former qualify as true human rights. A possible approach to this question is not to engage in the debate about the justification of these different rights, but instead to try to explain the reasons for the widespread perception (be it justified or not) that there is such a difference.⁷¹ One way to do so is to refer to loss aversion as a psychological mechanism. Taking this approach, “the crucial distinction is not necessarily between governmental acts and omissions, but rather between government giving and not taking.”⁷² Civil and political rights are perceived as being about the government not taking something, such as the unrestrained possibility of free expression. Social and economic rights, by contrast, are about giving something to the rights-holders. As losses loom larger than gains, the former kind of rights enjoys greater plausibility than the latter. Reference points are one major factor in analyzing this problem. The fact that many legal systems demand that if social or economic benefits are provided then this must be done without discrimination is based on such a reference point: Not receiving the benefit is experienced as a loss,

⁶⁹ Cf. for a restatement Eyal Zamir and Doron Teichman, *Behavioral Law and Economics* (Oxford: Oxford University Press, 2018).

⁷⁰ Cf. for a summary Kahneman, *Thinking, Fast and Slow*, 109 ff.

⁷¹ Zamir, *Law, Psychology and Morality*, 140.

⁷² Zamir, *Law, Psychology and Morality*, 143.

because those similarly situated persons who do receive the benefit serve as reference points.⁷³

This view is of considerable interest when attempting to explain why civil and political rights are perceived as being different from social and economic rights, at least in modern debates. Whether this difference also exists in this form from a historical perspective is far from clear, given the prominent place of rights to material goods in older reflections.⁷⁴

This approach leads to the following question, however: Why does loss aversion (assuming that it plays a role in this context, if only for the sake of the argument) not settle the case? Why do questions about social rights not only arise at all but do so very powerfully, and arguably since the beginnings of the discourse on rights? This seems to indicate that certain normative principles have an influence on human perceptions of the justification of human rights beyond loss aversion, including principles of justice, which are particularly relevant for social and economic rights.

This does not speak against an impact of loss aversion on such debates, but rather against assuming that this is the decisive or even only psychological influence on the perception of whether these rights are justified.

Other examples seem to confirm this kind of analysis. One such example is affirmative action, which is a controversial issue in the interpretation of equality guarantees and thus in the interpretation of an important element of human rights. Legal systems apply affirmative action to benefits that people do not yet possess.⁷⁵ Accordingly, affirmative action programs concern, for instance, access to university or employment for people who do not enjoy this access yet, but do not demand that others relinquish places at universities or jobs they already hold. This can be explained by loss aversion: The losses of losing a benefit (admission to university, employment, etc.) loom larger than the gains for those who are not enjoying this benefit yet.⁷⁶

There are interesting questions to be asked beyond loss aversion in this context, including the legitimate protection of trust and the reliance on certain decisions for allocating the goods of students admitted to university programs. Revisiting decisions on university admission should only be possible under very restrictive conditions, such as applicants intentionally providing wrong information, because they will have based many subsequent decisions on their admission. Such considerations are mirrored in some administrative law.

But even if one disregards this for the sake of the argument, an answer still is needed to the question of why the desire for affirmative action arises in the first place. This desire is connected with the idea of justice as equality. The ultimate

⁷³ Zamir, *Law, Psychology and Morality*, 143.

⁷⁴ Thomas Aquinas' theory of strong obligations of mutual help is but one example discussed in Chapter 3.

⁷⁵ Zamir, *Law, Psychology and Morality*, 148.

⁷⁶ Zamir, *Law, Psychology and Morality*, 144 ff.

aim is, after all, to achieve equal access to the benefits of society independently of characteristics that are irrelevant for the allocation of such goods, as skin color is for admission to university. For this reason, minorities who were excluded previously are treated preferentially in order to overcome traditional patterns of exclusion. The question is not whether this connection between affirmative action and justice exists, but rather whether considerations of this sort can trump the principle of equal treatment for a transitional period in order to achieve this aim.⁷⁷

It is argued that loss aversion is compatible with common-sense morality – which adheres to threshold deontology.⁷⁸ Classical distinctions in ethics between doing and allowing or between intending (in the sense of purposeful action) and foreseeing mirror loss aversion, it is argued: The duty to avoid doing harm is stricter than the duty (if it exists at all) to help others (or to not allow harm to happen to them). The two distinctions are not the same, as illustrated by cases in which one intentionally allows harm or harms somebody by an action, but only as a side effect.⁷⁹

The problem with this account is that loss aversion may be compatible with a version of threshold deontology but clearly fails to adequately specify the relevant moral principles that are the real core of the matter. These principles are more complex than the doing/allowing or intending/foreseeing distinctions suggest.

Take the trolley problem: Mere loss aversion gives no reason to judge flipping the switch in the bystander case to be permissible but throwing the person off the bridge in the footbridge case to be impermissible. In both cases, the losses are the same – five people killed, one person not killed, or vice versa, depending on whether or not action is taken. Even loss aversion supplemented with the distinction between doing and allowing is not enough. Consider a variant of the footbridge case: The person on the bridge is a toddler, Hannah, on her tricycle who will fall on the track as the railing is damaged, stopping the train, if the other person on the bridge who is entrusted to care for the toddler does not prevent the child's fall, which the other person is easily capable of doing. It seems impermissible to let the toddler fall in order to stop the train, even though one is not doing anything, only allowing something to happen. The reason has already been pointed out: More complex principles than loss aversion or the distinction between doing/allowing determine the evaluation of such cases. What is key here is an ends–means distinction and the prohibition of using human beings (merely) as means for other ends, be it by doing or allowing.

Loss aversion consequently has limited power to explain the problems investigated here. However, studies in experimental psychology and behavioral economics deal directly with normative principles relevant for our argument, looking at intuitions about justice and benevolence – the next topic to which we will turn.

⁷⁷ This insight can be translated in the terms of prospect theory, cf. Zamir, *Law, Psychology and Morality*, 228.

⁷⁸ Zamir, *Law, Psychology and Morality*, 177 ff.

⁷⁹ Zamir, *Law, Psychology and Morality*, 182.

6.6 JUSTICE AND BENEVOLENCE

Research on justice has a substantial tradition in social psychology, encompassing research on relative deprivation, distributive justice and the fairness of outcome distributions, procedural justice and corrective justice.⁸⁰ One prominent example from the current debate is the extensive discussion of fairness or, to use another term, inequality aversion. Yet another example is altruism, which can be seen, for example, in the case of contributions to social goods or in the (related) form of altruistic punishment. Importantly, altruism is understood in the sense of strong reciprocity; that is, behavior that does not lead to individual economic benefit of the agent.⁸¹ There is cross-cultural research on such attitudes.⁸² Another aspect of this research is the development of such patterns of behavior in children.⁸³

From the perspective of our inquiry, one problem of these studies is that they are predominantly concerned with patterns of behavior (e.g. distributive acts, punishments, rewards) and not with the internal mental states of the agents. In particular, they are not concerned with the reflective evaluation of actions by an agent or observer that has deontic, prescriptive consequences either from the first-person perspective of the agent or the third-person perspective of the observer. Reflective evaluation is crucially important for human morality, however, as moral judgment involving a prescriptive dimension is the core of the matter, as already indicated. Morality is not concerned simply with acting in conformity with certain (other-regarding) standards; it is concerned with a moral evaluation that yields prescriptive propositions such as “You should not bombard hospitals in civil wars” and possibly gives rise to action on the grounds of and motivated by such prescriptive propositions.

It is important to remain aware of a traditional insight of moral philosophy in this respect: There is no necessary or deterministic connection between moral evaluation and moral behavior. An agent may very well perceive the morality or immorality of an action but nevertheless fail to act accordingly due to intervening interests, weakness of moral will and so forth. Agents’ failure to display altruistic or just

⁸⁰ Cf. Tom Tyler et al., *Social Justice in a Diverse Society* (New York: Routledge, 1997), 11 ff., for an overview of earlier research.

⁸¹ Cf. Ernst Fehr and Urs Fischbacher, “The Nature of Human Altruism,” *Nature* 425 (2003).

⁸² cf. e.g. Joseph Henrich et al., “‘Economic Man’ in Cross-Cultural Perspective: Behavioural Experiments in 15 Small-Scale Societies,” *Behavioural and Brain Sciences* 28, no. 6 (2005).

⁸³ Cf. e.g. Kristina Olson and Elizabeth Spelke, “Foundations of Cooperation in Young Children,” *Cognition* 108, no. 1 (2008) (three- and five-year-olds); Marco Schmidt and Jessica Sommerville, “Fairness Expectations and Altruistic Sharing in 15-Month-Old Human Infants,” *PLoS ONE* 6, no. 10 (2011); Ernst Fehr, Helen Bernhard and Bettina Rockenbach, “Egalitarianism in Young Children,” *Nature* 454, no. 7208 (2008); Ernst Fehr, Daniela Glätzle-Rützler and Matthias Sutter, “The Development of Egalitarianism, Altruism, Spite and Parochialism in Childhood and Adolescence,” *European Economic Review* 64 (2013). On the effect of self-reflection in the framework of identity utility, cf. Christoph Engel and Michael Kurschilgen, “The Jurisdiction of the Man within – Introspection, Identity, and Cooperation in a Public Good Experiment,” *MPI Collective Goods Preprint* (2015).

behavior does not allow direct conclusions to be drawn as to the principles governing their moral judgment. They may simply not be acting on the basis of their moral judgment.

There is another problem concerning the theoretical and conceptual framework of such studies and thus the determining framework for the interpretation of data. Concepts such as “inequality aversion” or “other-regarding preferences” play a central role in some of these studies. These concepts do not fully fathom the intricacies of moral judgments, however. Aversions and preferences describe inclinations to act. An aversion to asparagus means that one has no wish to eat this vegetable if it can be avoided. Moral judgment concerns something qualitatively different, as we just have seen, namely the *reflective evaluation* of an intention or action (e.g. based on such inclinations). This evaluation has deontic, prescriptive consequences, a moral ought. This “ought” is categorically different from a mere aversion or preference. It does not *incline* – it *obligates* persons, whatever their inclinations may be. The possible conflict between prescriptions that obligate and inclinations to act, such as preferences, or not to act, such as aversions, reveals the difference between the two categories, showing that we should not mistake the one for the other. Moreover, there is the basic deontic category of permissions, which is important for the concept of rights and cannot be translated into preferences or aversions either. Whether you have an aversion or preference to do something is irrelevant for the question of whether you are permitted to do it.

These methodological and theoretical problems notwithstanding, this area of research offers important findings. There are a large number of studies that provide empirical evidence about the egalitarian intuitions of human beings, prominently in the ultimatum game, for instance, which involves the following: A player receives a sum of money and distributes it between themselves and another player. If the recipient accepts the distribution, both keep the amount distributed; if not, nobody receives anything. There are many variants of this game – for instance, the dictator game. As a baseline, the results show that proposers offer an almost equal share and that responders will not accept just any distribution but reject very low shares.⁸⁴ We should note that neither an unequal proposal nor the acceptance of an unequal distribution means that the offer is considered just. Selfish impulses evidently have a strong hold on human beings (whether the proposer has a persistently good

⁸⁴ Cf. for instance Colin Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (New York: Russell Sage Foundation, 2003), 43 ff. Results for the dictator game (the proposer sets the share unilaterally) indicate similar patterns: The average given is about 42 percent, though about 36 percent maximize their own profit from the game: “Even generous subjects thus tend to have a selfish side,” Christoph Engel, “Dictator Games: A Meta Study,” *Experimental Economics* 14 (2011): 583, 607 concludes. Joseph Henrich, *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton, NJ: Princeton University Press, 2017), 193, reports different results for traditional small-scale societies.

conscience is, however, a different question), and the responder may consider receiving a smaller, unequal share to be a lesser evil than receiving nothing at all.

The most plausible interpretation of these results is that human beings are not (only) maximizers of utility, but that their evaluation of distributions is based on moral principles – for instance, of (proportional) equality.⁸⁵ If they were simple utility maximizers, they would accept any distribution that improves their situation, however small it might be. Importantly, the preservation of relations of equality seems to be a value in itself, and greater than at least some material benefits.

The results of studies on altruistic punishment point in the same direction: Human beings value certain standards and act to enforce them, even if this comes at a certain cost to themselves. Whether the reason for this kind of behavior is that they expect to benefit from the maintenance of such structures themselves at some point or that they consider defending certain normative principles to be of intrinsic value is quite another question.

There are other patterns of behavior that are discussed in connection with questions of social norms and morality. To take some examples:⁸⁶ Communication has the effect of increasing cooperative behavior. Many people are conditional cooperators – they cooperate in proportion to the cooperation of others. In finitely repeated public good games, levels of cooperation deteriorate over time, despite high initial cooperation rates. Stable group composition leads to higher cooperation rates. Framing a game as a community game has positive effects on cooperation in comparison to framing the same game as a stock market game. If the games are played sequentially, however, the effect of this framing disappears. Peer punishment leads to greater levels of cooperation, but its effects can be undermined if the punishment indicates selfish intentions. There may even be forms of punishment of cooperators. Punishment means a certain investment. Nevertheless, people prefer an environment with peer punishment. Rewarding cooperators increases cooperation. Bringing cooperative individuals together also augments cooperation.

How to explain such patterns of behavior? One way to approach the problem is to account for these patterns by the effects of social norms. This is how Ernst Fehr and Ivo Schurtenberger proceed, for instance. In their influential analysis, they supplement a direct social norm approach with the key idea of conditional cooperation and a

⁸⁵ Zamir and Teichman, *Behavioural Law and Economics*, 102, sum up research on social justice: “The most influential theory in the social-psychological study of substantive fairness has been *equity theory*. It posits that people perceive that they are treated fairly when the ratio between their received outcomes (for example their salary) and their input (e.g., the effort, talent and commitment they put into their work) is equal to the ratio between the received outcomes and the inputs of other peoples” (emphasis in original). This is evidently nothing but an evaluation on the basis of proportional equality and certain criteria of distribution (effort, talent, commitment) regarded as relevant for certain spheres of distribution, cf. Chapter 5.

⁸⁶ Cf. for the following list of behavioral patterns Ernst Fehr and Ivo Schurtenberger, “Normative Foundations of Human Cooperation,” *Nature Human Behaviour* 2, no. 7 (2018): 458–68, 459 ff.

set of particular psychological mechanisms. The direct social norm approach calculates the utility and disutility of following a social norm. The (dis)utility depends on an intrinsic desire to comply with norms.⁸⁷ The conditional cooperation approach postulates that people will cooperate dependent on others' level of cooperation. Because of the mechanism of conditional cooperation, the mentioned patterns of behavior arise, Fehr and Schurtenberger argue, with the exception of: peer punishment; punishment of cooperators that threatens the positive effects on cooperation; and the preference for environments with peer punishment. Here, Fehr and Schurtenberger's theoretical account relies on additional psychological mechanisms, such as social preferences for fairness/equity, reciprocity, a prosocial self-image or an aversion against guilt.⁸⁸ The social preferences for fairness and equity are themselves determined by social norms.⁸⁹ From these authors' perspective, institutional structures play an important role: Punishment is necessary to maintain cooperation and institutions that guarantee norm conformity.⁹⁰

It is true that "unconditional normative prescriptions like 'be selfless'" alone cannot account for such patterns of behavior.⁹¹ But this does not mean that the normative principles that we have discussed in the present inquiry are irrelevant or nonexistent and that the moral world is limited to conditioned cooperation and the particular understanding of fairness and equity considered by this experimental work. This is because a moral obligation provides a sometimes-powerful motivation but is always just one of other impulses that influence human behavior. According to the view defended here, human beings are not selfless beings, but beings who have the faculty to limit their many selfish impulses because of moral judgments and the volitional consequences of these judgments. Moreover, cooperation is a complex affair and evidently not just based on moral impulses of selflessness, justice and altruism.

Therefore, the patterns of behavior recalled as examples of this strand of research come as no particular surprise: Communication can help with cooperation for many reasons – for instance, defining one's mutual advantage, clarifying common interests or strengthening the will to adhere to moral intuitions. That levels of cooperation can be influenced by the declining cooperative behavior of others is easily reconcilable with a moral orientation, too, as the latter does not imply a readiness to be exploited. Even if we feel a moral obligation to act in certain ways, our preparedness to do so may diminish if we see others pursuing egoistic goals. The sobering effects of repeatedly played games with declining levels of cooperation may have related reasons but do not say anything about the reality of more exacting standards of behavior. In the same vein, partner matching or the framing of a game as community-oriented action can

⁸⁷ Fehr and Schurtenberger, "Normative Foundations," 461 f.

⁸⁸ Fehr and Schurtenberger, "Normative Foundations," 463.

⁸⁹ Fehr and Schurtenberger, "Normative Foundations," 463.

⁹⁰ Fehr and Schurtenberger, "Normative Foundations," 464.

⁹¹ Fehr and Schurtenberger, "Normative Foundations," 461.

help agents to adhere more faithfully to some moral standards and prudential principles that are also important for cooperation. The sequential playing of these games may have the mentioned sobering effect.

Altruistic and peer punishment has much to do with principles of justice. Distinguishing just punishment from unprincipled, potentially boundless revenge and determining the principles that make punishment just have been permanent themes of the theory of justice, dealt with in many contributions, including Aristotle's classic treatment of the matter. A preference for entrenched systems of punishment is a possible product of cool-headed assessments of the pacifying, conflict-reducing effects of such arrangements. That rewards may help to motivate people to do anything, including to cooperate, needs as little proof as the positive effects of working with like-minded cooperators.

The reference to fairness, equity and moral emotions like guilt takes us straight to the question of the structure underlying human moral cognition. Here, one has to be analytically precise and try to determine more concretely what a moral judgment is about, avoiding in particular the category error of mistaking a preference or aversion for an obligation or prohibition. We already tried to give some indications of what such an analytical theory of morality could look like, and we will proceed further on this path in the last chapter. Moreover, this body of empirical evidence does not conclusively answer the question of the origins of foundational moral intuitions; in particular, it cannot tell us whether they are culturally induced or based on the innate structure of the human mind. We will also ask, therefore, whether it is really true that ideas of fairness and equity are wholly the offspring of social norms or whether these social norms are rather in at least some part the expression of basic moral intuitions, not least in the light of child psychology.

The behavioral studies we have reviewed contain no findings that have a direct and robust bearing on the question of human rights. We identified some problems of the theoretical framework of the interpretation of the empirical findings. Nevertheless, the thrust of this important research underlines the fact that intuitions about justice such as equality and altruism are not mere theoretical chimeras, but psychological realities of substantial importance for human beings, their decision-making, the explicit norms they develop and their subsequent actions.