Introduction: Debunking Arguments and the Gap

I.1 Cornflakes and Critical Theory

Many will know John Harvey Kellogg as the inventor of cornflakes, but perhaps his most profound influence on people's daily routines was due to his fierce advocacy against masturbation. Although "[c]overing the organs with a cage" (Kellogg 1887, 290–327), also proved effective, his favorite method for curing this pernicious habit was circumcision. Its punitive connotations were welcomed both by Kellogg himself and by his morally panicked contemporaries.

The medical rationale for circumcision was only "discovered" *post hoc* (Darby 2003 and 2005). However, its therapeutic benefits are now by far the most frequently cited reason for recommending the practice.

Suppose this historical sketch of how infant circumcision became widely practiced is true, as indeed it is. And suppose that you also find out that, as far as developed countries are concerned, routine circumcision of male infants is now performed almost exclusively in a country whose cultural climate has a reputation for being as prudish as a Victorian spinster. What, upon learning these facts, should we conclude about whether performing this operation is justified?

In the remainder of this book, I will not talk about breakfast and masturbation as much as I would like to. What I will talk about, however, is the style of argument illustrated here: there is frequently a contrast between an official story citing the reasons that would, in principle, be suitable for justifying a belief or practice and an unofficial one revealing its actual origins, the way it came about, and the forces that first caused and continue to sustain it.

Historical accounts of the – typically unobvious and opaque – origins of a belief or practice are often called *genealogies*. This book is about the epistemic role of such genealogies. What is their normative significance? Do they tell us anything of interest about whether a given belief or practice is justified, rational, or defensible? In other words: are genealogical accounts of why people think and do something fit to *debunk* what they think and do? Do they even have the power to do so?

Genealogies are typically deployed in a critical spirit, and genealogies that are supposed to have such negative, or undermining, epistemic significance are nowadays often referred to as *debunking arguments* (Nichols 2014). The question I am interested in is whether there are any successful arguments of this sort. Here, we have four main options:

(1) *Genealogies always debunk*. According to this thesis, it always undermines the justification of a belief or practice to find out that it has a certain causal origin. This thesis has been widely acknowledged to be implausibly strong. Call it the *genetic fallacy*.

(2) *Genealogies never debunk*. Some hold that learning about the historical background of one's beliefs has no epistemic significance regarding whether these beliefs are justified. That is, even if one found out that one has no good grounds for believing something and plenty of evidence that one's beliefs came about through untrustworthy means – think: wishful thinking, guessing, hearsay – this would entail nothing whatsoever for whether one is entitled to hold the belief or whether one would best abandon it. This thesis, too, seems implausibly strong and perhaps implausibly optimistic. Call it *historical obtuseness* (or, less incendiarily, naïveté).

(3) *Genealogies sometimes debunk, at least when the right conditions are met*. Here, the idea is that at least *some* historical accounts of how an individual came to believe or how a group of people came to practice something should make us suspicious. Bernard Williams (2002 and 2005), for instance, suggested that regimes of power must pass a "critical theory test" in order to count as legitimate. All beliefs and practices come about in some way; but when they came about through the oppression and coercion enacted by a powerful elite, and when the fact that said beliefs and practices are accepted and deemed legitimate is *due to the fact* that they came about through violent oppression, then we have reason to doubt their authority. Exposing the element of rational contingency in something's backstory can, and indeed should, shake our confidence in it.

(4) *Genealogies sometimes* justify, *at least when the right conditions are met*. Then again, some authors note that genealogies do not always have to constitute a net epistemic loss. In some cases, pointing out the historical origins of a belief or practice can make it appear in a more

> favorable light, especially when the reasons currently taken to justify it are reflected in its actual learning history (Kumar 2017).

In this book, I will focus on debunking arguments in ethics, where their particular popularity is perhaps best explained by the pleasant shudder of "doxastic embarrassment" (Rini 2017) they sometimes bring about. I will argue that there are both successful *debunking arguments* and, as I will refer to them, *vindicating* arguments in ethics, and thus that versions of (3) and (4) are true. But I will also argue that the structure, scope, depth, and indeed the very *point* of debunking arguments remain only poorly understood. This book wants to contribute to a better understanding of how debunking works, why it works, and when it works.

Against those who argue that genealogical arguments have no epistemic clout (Srinivasan 2015), I will argue that they do but only if understood correctly. Debunking arguments, I will show, are for the most part a burden-of-proof–shifting device that induces epistemic discomfort with one's intuitions. In ethics, which essentially always bottoms out in intuitions (Huemer 2005), this makes them a big deal, and as such, they do tremendously important epistemic work, even though there is a sense in which it remains true that they often don't do any of the epistemic heavy lifting. Debunking arguments clear the epistemic ground: they show who owes a plausible justification for their beliefs in the second round of inquiry.

Against many – Foucault's archeological endeavors come to my mind – who also wish to harness the debunking force of genealogical arguments, I will show that their epistemic significance is easily, and indeed frequently, overestimated. It is true that many things we do and believe did not come about in ways that make much sense. We may reconstruct a sense-making narrative in hindsight, but what we frequently find when we look at the actual historical record is that, for instance, the way we treat the mentally ill, delinquents, or other "deviants" has come about in fragmented, contingent, somnambulistic ways. Like those ancient buildings we find in the ground, many things we do or believe – and take for granted in doing and believing – have been pieced together without any central rational oversight. Accordingly, there may be little or nothing holding those pieces together besides sheer luck and a dash of cement. The result is that all too often, genealogical arguments are like New York City clubs: their window dressing and reputation are impressive, but upon entering, it quickly becomes clear that they don't live up to the hype and merely manage to survive by overcharging gullible tourists.

I wish to explore the prospects of the genealogical method. But I will not be content with speculative genealogies. Instead, I will focus on genealogical critiques for which there is actual evidence. That is, I will bring the tools of moral psychology as well as empirical and experimental philosophy to bear on the issue of genealogical debunking, which has fascinated people since the nineteenth century, or perhaps even since Xenophanes first argued that if horses had gods, they would imagine them to look like horses. And in the spirit of the genealogical method, it is perhaps most fitting to start with the past.

I.2 A Cold, Hard Look

For most of its history, philosophical moral psychology has been in bad shape. People were asking the right questions, but their methods were questionable: rampant speculation was revised in light of pure guesswork; guesswork had to be amended on account of arbitrary superstition; superstition was corrected by flimsy moralizing; and the whole thing was rounded off by a healthy dose of wishful thinking. Philosophical theories of human nature had to state how human beings ought to be rather than how they actually are.

It is not a good idea, generally speaking, to speculate about the nature of the moral mind without systematically investigating how the mind works. Why philosophers failed to appreciate this rather obvious truth is something I can only speculate about myself. The – arguably false – idea that the mind is transparent to itself and can thus be studied without external aid may have played a role. We now know that this type of self-transparency is an illusion and that expecting the mind to give honest answers when examined by introspection alone is hopelessly naive.

Perhaps I exaggerate, and it wasn't quite as bad. To find out how moral agents think and act, some philosophers like Aristotle, Hume, or Kant did consult the best science of their time. Then again, this did not necessarily amount to much. Others – Nietzsche comes to mind (Knobe and Leiter 2007) – were in fact pioneers and gave the field of empirically informed moral psychology, most of which was yet to emerge at the time, new directions to pursue and new questions to address. Yet all too often, philosophers "have been content to invent their psychology [. . .] from scratch" (Darwall, Gibbard, and Railton 1992, 189). A "cold, hard look at what is known about human nature" (Flanagan 1991, 15) seems to me to be the best cure for this affliction.

The main tension between philosophical and empirical accounts of human moral judgment and agency comes down to the fact that, at the end of the day, philosophers are interested in moral psychology for one thing and one thing only (I exaggerate again). They want to know what facts about the *psychological* foundations of morality can teach us about the foundations of morality, *period*: how facts about human nature bear on right and wrong, good and bad, just and unjust. This tension is further aggravated by the fact that many philosophers deem this to be a hopeless endeavor that is doomed to fail from the outset. The problem, these philosophers argue, is that there is no way (no legitimate and informative one, at any rate) to get from an *is* to an *ought*. Rumor has it that facts are different from values. Descriptive statements, it is said, do not entail prescriptive propositions. Empirical information, the story goes, has no normative significance. Nature allegedly has no moral import.

In what follows, I will refer to this problem as *the gap*. In the first section of this introduction, I will briefly explain what the gap is, why it is said to exist, and to what extent it is supposed to pose an obstacle to empirically informed theorizing about ethics. Most of this will be familiar to many readers.

In the second section, I will take a look at some of the most interesting recent developments in empirical moral psychology and explain what their normative implications are supposed to be to set the stage for the chapters to come. My selection of topics will be somewhat arbitrary and the discussion I provide by no means comprehensive. I am not attempting to give an overview of the whole field of contemporary moral psychology. This has already been done elsewhere by people more qualified to do this than myself (see Doris and Stich 2005, Appiah 2008, Alfano and Loeb 2014, Tiberius 2014, Rini 2015, Alfano 2016). Instead, I choose a more focused approach and look at the whole field from the perspective of what I take to be the main issue of philosophical interest: my aim is to illustrate how empirical moral psychology might be brought to bear on issues of normative significance – what the virtues are, what makes for a good life, whether free will exists, what role luck plays in morality, what constitutes an action, what it means to be a person, how people arrive at moral judgments, whether these judgments are relative, and whether we are at all competent to make them. My discussion will be arranged around four clusters: normative theory, moral agency, moral and nonmoral judgment, and moral intuition.

In the final chapter of this book, I will extract some lessons from this discussion. Are the skeptics right, and when it comes to figuring out what demands morality makes on us, empirical information remains thoroughly irrelevant? Or are there grounds for optimism, and empirically informed ethics may have a future after all? I will argue that the normative significance of empirical studies of human moral cognition and behavior, though always indirect, comes in essentially three forms: (i) by debunking the processes on the basis of which we make moral judgments and develop moral concepts; (ii) by debunking the empirical presuppositions of some normative theories, thereby possibly vindicating those of others; and (iii) by making information of type (i) and (ii) reflexively accessible, that is, by providing tools for the reflective improvement of moral judgment and agency by bringing to light the sometimes egregious mistakes that escape our powers of introspection and the empirically unaided mind.

Debunking arguments play a central role in all three of these ways of bringing empirical data to bear on normative issues. These arguments are uniquely equipped to bridge the is/ought gap, for they causally *explain* a judgment in a way that makes it appear normatively *suspect*. Typically, this involves showing that a person would believe something *even if it were not true*. In the absence of further grounds for holding the belief, this defeats a person's justification for believing it. Debunking arguments are thus perhaps the most promising tool for galvanizing the empirical and the normative.

I.3 The Gap

In philosophy, skepticism about the relevance of empirical facts for so-called *normative* questions – questions about right and wrong, permissible and forbidden, virtue and vice – can draw on two *loci classici*. One can be found in the third part of David Hume's *Treatise of Human Nature*, where he complains that

> [i]n every system of morality, which I have hitherto met with, I have always remarked, that the author proceeds for some time in the ordinary way of reasoning, and establishes the being of a God, or makes observations concerning human affairs; when of a sudden I am surprised to find, that instead of the usual copulations of propositions, is, and is not, I meet with no proposition that is not connected with an ought, or an ought not. (1739/2000, III.I.I)

Hume argued that this transition was as widespread as it was illegitimate; for in his view and the view of many others, there is no logically valid way to derive a proposition with normative content (It is not ok to lie; Drone surveillance is reprehensible; Chastity is a virtue; We have a duty to help others when doing so involves little cost to ourselves) from a set of premises with purely descriptive, factual content (People lie all the time; Drones are really useful; Your father wants you to be chaste; Helping others will make people like you). An inference is logically valid just in case the truth of its premises guarantees the truth of its conclusion. No such inference, Hume thought, could ever take you from an *is* to an *ought*.

The second go-to place for friends and foes of *the gap* is G. E. Moore's (1903) *Principia Ethica*. Here, Moore coined the term "naturalistic fallacy" (62) to refer to attempts to identify the property of being *good* with any natural property, such as being *useful* or *maximizing pleasure* or being *economically efficient* or being *sanctioned by the state*. Moore's point was that *good* and *bad* cannot be defined in natural terms, because if they could, then whenever we had found some action or event instantiating the natural property picked out by our definition (given that said definition is correct), the question whether the action or event is also good would necessarily be *closed* to anyone but the conceptually confused. Centaurs, and only centaurs, are creatures with an anthropic upper and hippic lower half; if I manage to show you such a thing, the question whether it is also a centaur is *closed*. Now Moore argued that for every proposed natural definition of the good – say "the good = that which maximizes pleasure" – it always remains possible to ask whether something instantiating the natural property specified in the definiendum is also good. "It maximizes pleasure, but is it also good?" Or: "It is loved by the gods, but is it also good?" Or: "It is useful for society, but is it also good?"; and so on. These questions all make sense, and so the property of being good cannot be conceptually reduced to other, natural properties. This is Moore's famous "open question argument."

The naturalistic fallacy is not, strictly speaking, a fallacy, and as we have seen, the term was originally supposed to refer not to *the gap* but to an entirely different, semantic point. Then again, people love to accuse one another of fallacious reasoning, and the term is catchy, so "naturalistic fallacy" stuck around and is now widely used for illicit attempts to bridge *the gap*. Examples for naturalistic fallacies are ridiculously easy to find and are especially common in debates on evolutionary psychology, sexual morality, and most other topics in applied ethics. I will not cite any sources here, as the research would have been too depressing. But I *can* give a few examples of the kind of reasoning I have in mind and which we are all too well acquainted with: evolution favors the selfish and competitive, so that is how we, too, ought to act; homosexuality is unnatural and should thus be banned; humans are the only animals with the power to reason, and so the rational life is best for humans; people have always killed animals for food, and women were always discriminated against, so clearly there is nothing wrong with those things. Never mind whether these inferences get the facts right or not – because even if they did, they would fail to establish their conclusion on account of *the gap*.

On the other hand, it seems hard to see how empirical facts could *always* remain *thoroughly* irrelevant to normative inquiry. Whether abortion is permissible and under what conditions will surely depend on what kind of being a fetus is and whether it can feel pain or has interests and conscious experiences. Likewise, my indignation toward the man I believe my wife cheated on me with and which I am about to punch in the face will readily switch its target once I have found out that *this* man isn't the culprit, but the pathetic scoundrel standing next to him. What should be done about climate change, or whether anything should be done at all, cannot be assessed without factual knowledge. And whether you should perform that tracheotomy to save your suffocating friend will depend on how likely it is that you will succeed. In all these cases, empirical facts have bearing on issues of normative significance, if only via the nonmoral facts upon which moral facts are grounded.

Moreover, many normative moral theories seem to make rather straightforward assumptions about what kinds of agents we are, assumptions that are far from empirically innocent. For instance, some Kantians argue that moral norms are prescriptive rules whose authority does not depend on whether one is already motivated to conform to them: these rules are supposed to be motivating *independently* of an agent's desires and goals simply in virtue of the fact that they specify what it means to be an agent (Korsgaard 1996, Velleman 2011). But what if this paints an unrealistic picture of how motivation works and of what constitutes an agent? Virtue ethicists often claim that a good person is a person with a coherent set of laudable character traits (Hursthouse 1999, Foot 2001). Does this account rely on an erroneous idea of how people function and how well their personalities are integrated? Some consequentialists hold that the right action – the one we ought to choose – is the unique action that has the best consequences. But what if figuring out which action this is is beyond human deliberative powers (Mason 2013)? In all these cases, normative theories make empirical presuppositions.

The question, then, is this: despite the fact that no ought ever follows from an is, and despite the fact that the concept of the good cannot be identified with any empirical property, how should we understand the *normative relevance of empirical facts* in light of the *empirical presuppositions of various normative commitments*?

I.4 Normative Theory

(i) *Consequentialism and Deontology*. Contemporary normative ethics is organized around a distinction that manages at the same time to be one of the least well liked and yet one of the most popular in all of philosophy: the distinction between *consequentialism* and *deontology*. Consequentialist moral theories hold that the rightness or wrongness of an action is determined *only* by its (actual or expected) consequences. Deontological moral theories deny this. Some deontologists hold that intentions matter for the moral evaluation of an action as well, while others argue that there are certain side-constraints (such as individual rights) on the maximization of the good, that it can make a moral difference whether one actively does something or merely allows it to happen or whether someone uses someone else as a mere means to an end rather than an end in herself. There is plenty of evidence that on an intuitive level, people take deontological considerations to be morally relevant (Young et al. 2007). Often, their judgments conform to deontological rules such as the doctrine of double effect (according to which harming someone can be permissible when it is an unintended but foreseen side effect rather than when the harm is directly intended; Kamm 2007, Mikhail 2007), even though such slightly more sophisticated principles may remain ineffable.

What about *the gap*? Can empirical data shed light on which theory is correct? One way to model the difference between consequentialism and deontology is to look at sacrificial dilemmas involving urgent trade-offs between harming an individual person and promoting the greater good and to see which conflicting actions consequentialism and deontology classify as right and wrong, respectively, when doing what's best overall clashes with certain intuitively plausible moral rules. Moral emergencies (Appiah 2008, 96ff.) of this sort form the basis of what is perhaps the single most thriving and controversial research program in normatively oriented empirical moral psychology: Joshua Greene's *dual process* model of moral cognition (Greene 2014). According to this model, cognitive science can show that one of the two normative theories is superior to the other. Consequentialism, the evidence is purported to show, engages more rational parts of the brain and more sophisticated types of processing than deontology, which is associated with more emotional parts of the brain and

Cambridge University Press
978-1-108-42369-4 — Debunking Arguments in Ethics
Hanno Sauer
Excerpt
More Information

more crude forms of cognition (Greene 2001 and 2004). When people judge it impermissible, for instance, to kill one person to save five others (thereby endorsing the deontological option), they arrive at this judgment via a more emotional and less calculating route. Deontological moral theory, then, amounts to little more than *post hoc* rationalizations of those brute, alarm-like responses (Greene 2008; see Chapters 6 and 7 for a more thorough discussion of the neuroscience of moral judgment).

The dual process model's main normative upshot is supposed to be a vindication of consequentialist and a debunking of deontological intuitions on the basis of empirical evidence regarding the cognitive processes that produce these two types of moral intuitions. But it remains unclear whether the way people arrive at their consequentialist responses deserves to be described as consequentialist reasoning at all rather than an ordinary weighing of competing considerations for and against a proposed action (Kahane 2012). Even worse, the consequentialist judgments some people end up endorsing do not seem to be based on an impartial concern for the greater good but on much more sinister dispositions (Kahane 2015). Perhaps most importantly, the connection between consequentialist judgments and controlled, System II processing on the one hand and deontological judgments and automatic, System I processing on the other hand (Evans 2008, Kahneman 2011, Stanovich 2011) seems to be due to the fact that in Greene's original studies, the consequentialist option always *happened to be* the counterintuitive one. When this confound is removed and counterintuitive deontological options are included, the pattern is reversed (Kahane et al. 2012; cf. Greene et al. 2014.) This pattern is corroborated by Koralus and Alfano (2017).

Dual-process theory continues to be haunted by *the gap*. Empirical data on which type of process or which brain region is involved in the production of a moral judgment tells us very little about whether this judgment is justified – unless we *already* know which processes are unreliable and which aren't, which we arguably do not. Now the dual-process model's two best shots are an *argument from morally irrelevant factors* and an *argument from obsoleteness*. First, it could be shown that regardless of whether people arrive at them through emotion or reasoning, deontological intuitions pick up on *morally irrelevant factors*, such as whether an act of harming someone has been brought about in a distal or proximal way. Such sensitivity to morally extraneous features is often sufficient to indict a particular type of judgment as unreliable. Second, one could argue that some moral intuitions are generated on the basis of processes that are unlikely to deliver correct results under conditions they have neither