Routledge
Taylor & Francis Group

Philosophical Explorations

An International Journal for the Philosophy of Mind and Action

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rpex20

Against moral judgment. The empirical case for moral abolitionism

Hanno Sauer

To cite this article: Hanno Sauer (2021) Against moral judgment. The empirical case for moral abolitionism, Philosophical Explorations, 24:2, 137-154, DOI: 10.1080/13869795.2021.1908580

To link to this article: https://doi.org/10.1080/13869795.2021.1908580

Published online: 11 Apr 2021.

PHILOSOPHICAL EXPLORATIONS
2021, VOL. 24, NO. 2, 137–154
https://doi.org/10.1080/13869795.2021.1908580

Routledge
Taylor & Francis Group

Against moral judgment. The empirical case for moral abolitionism

Hanno Sauer

Utrecht University, Utrecht, Netherlands

ABSTRACT
In this paper, I argue that recent evidence regarding the psychological basis of moral cognition supports a form of (moderate) moral abolitionism. I identify three main problems undermining the epistemic quality of our moral judgments – contamination, reliability, and bad incentives – and reject three possible responses: neither moral expertise, nor moral learning, nor the possibility of moral progress succeed in solving the aforementioned epistemic problems. The result is a moderate form of moral abolitionism, according to which we should make fewer moral judgments much more carefully.

ARTICLE HISTORY
Received 6 July 2020
Accepted 12 February 2021

KEYWORDS
Moral cognition; debunking arguments; moral psychology; moral abolitionism

1. Introduction: rational pessimism[1]

In this paper, I argue that recent evidence regarding the psychological basis of moral cognition supports a form of (moderate) moral abolitionism. I identify three main problems undermining the epistemic quality of our moral judgments – contamination, reliability, and bad incentives – and reject three possible responses: neither moral expertise, nor moral learning, nor the possibility of moral progress succeed in solving the aforementioned epistemic problems. The result is a moderate form of moral abolitionism, according to which we should make fewer moral judgments much more carefully.

I will argue that the practice of moral judgement is deeply problematic, and that therefore, we should stay away from it as often as possible. This claim applies to most people, in most cases, or rather: to most people always, and to all people most of the time. This rather immodest thesis should of course only be taken half seriously. But at least *half*: in what follows I will show that there are good reasons to be skeptical of our practice of moral judgement – at least more skeptical than we are usually inclined to be. This skepticism is not grounded in moral nihilism, nor is it based on an error theory. Rather, it is based on a moral argument itself: moral judgements are important and highly consequential, which makes it important not to render such judgements carelessly, or in bad faith, or without proper reasons, or an appreciation of the facts, or all of the above.[2] This is the only way their potential harmful consequences can be avoided or minimized. The claim of this paper, in short, is that we should (largely) stop making moral judgments because we are bad at it.[3]

CONTACT Hanno Sauer h.c.sauer@uu.nl

The pessimistic perspective I offer is no mere misanthropy; rather, it has an empirical basis. I will offer a (more or less representative) overview of the most important research developments regarding moral judgement within recent moral psychology and cognitive science, and show that the clumsiness of our moral competence is an empirically well-supported fact which moral philosophers would be ill-advised to ignore.

The pessimism I advocate is, nevertheless, rationalist in nature. The dubiousness of our moral capacities has little to do with the fact that the hopes and promises of the Enlightenment, that morality could be grounded in pure reason, turned out to be indefensible (although they did). Empirical impurities are welcome – and in any case unavoidable. Empirically informed ethics (or empirical ethics, for short) does not show that practical reason is a fiction; nor does it justify any Humean enslavement to the passions. Empirical ethics shows that our moral rationality is real but fragile, and that in its fragility, it often works in disturbing and unexpected ways (Rini 2017, 1439–1458). It thus fosters a sense of modesty towards our moral cognition. There is no *a priori* guarantee we are any good at moral judgment. *How* good we are is determined by how the world actually works.

My paper has five parts. The first three outline my pessimistic argument. Much of what I have to say here will not be new to anyone familiar with the literature, and is merely supposed to set the stage for the rest of the argument, so I ask the reader to bear with me until that point. In section (1) I deal with the problem of reliability. (2) discusses the problem of contamination. Section (3) zooms in on the problem of motivation. In the fourth section I consider possible optimistic replies, according to which the phenomena of moral learning, moral expertise and moral progress may reanimate our hopes in the quality of moral cognition. Since only the latter strategy gives us some grounds for hope – though certainly not for enthusiasm – I will end by addressing how we should move on from there. Should we simply stop making moral judgements? Or continue as we did before? Here, I will advocate a moderate moral abolitionism, according to which we should make moral judgments much less often, and much more carefully.

2. The problem of reliability

A first reason for a pessimistic outlook on moral judgement has to do with its reliability: empirical studies portray the genesis of moral judgements in an unflattering light. When our moral convictions are influenced by a variety of factors of dubious epistemic quality, we have little reason, or no reason at all, to trust those judgements. This strategy is pessimistic, since it requires an uncomfortable reevaluation of our moral knowledge. At the same time, it is fundamentally rationalist, for without robust standards of reliability it is not possible to identify a cognitive process as unreliable (Rini 2016, 675–697).

i *Sentimentalism vs. rationalism.* One way of disputing the reliability of moral judgments is to undermine them on psychological grounds. The most popular version of this strategy is the attempt to work out the emotional – and thus supposedly non-rational – basis of moral judgements (Prinz 2006, 29-43). First, I will show why *this* path gives us *no* particular reason for pessimism.

Empirical ethics has revived the sentimentalism/rationalism-debate. Sentimentalists like to point out that emotions are either proximally or distally (or both) important for moral judgements: moral convictions depend either on occurrent emotions, or on emotional dispositions; moreover, affective experiences must *necessarily* be present ontogenetically for people to form mature moral capacities at all (Prinz 2007).

As evidence for this *necessity* thesis one can rely on studies establishing a correlation between emotional deficits and non-normative moral cognition.[4] For instance, apart from a frequently disastrous biography, some studies suggest that it is difficult for psychopathic individuals to understand the difference between moral and conventional rules (Blair 1995).

A second sentimentalist driving force are studies on the impact of so-called incidental affects on the moral judgements of experimental subjects. These concern affective reactions which are artificially generated 'in the laboratory', stand in no relation to the scenarios judged by the participants, and whose origins are generally hidden from the test subjects. A series of studies suggests, for example, that feelings of disgust can influence moral judgements. These can be elicited by orchestrating an unsanitary test situation, unpleasant smells, or the activation of certain vile associations. In cases like these, the participants are often asked to judge simple 'vignettes'. Compared to an experimental control group without manipulated emotions, increased disgust can lead to stricter moral judgements. The idea is that, if moral judgements and sentiments co-vary, the latter are not only necessary but also *sufficient* for our moral beliefs.

The conjunction of the necessity – and the sufficiency thesis initially appeared to support a strong version of sentimentalism with respect to the nature of moral judgements. After about a decade of sentimentalist dominance in empirical metaethics, however, the pendulum is now starting to swing back.

It increasingly becomes clear that a distinction between rationality and emotion, reason and sentiment either cannot be made sense of at all (for conceptual reasons), or in any case not strictly enough (for empirical reasons). The conceptual problem has to do with the fact that the concepts of rationality and emotion are orthogonal to one another: the predicate 'rational' normatively characterizes a belief or action as well-grounded or reasonable, while the predicate 'emotional' descriptively characterizes the kind of mental process involved in it, independently of the question whether that process is reasonable or in some other sense legitimate. Now emotional processes may even happen to be, by and large, less likely to be based on good reasons than non-emotional processes; this would, however, merely amount to an empirically contingent fact rather than a general indictment of emotionally charged cognitive outputs. For structuring a research debate, the alternative 'rationality or emotion?' is about as useful as asking someone if they preferred to meet in the city or in the afternoon. Cognitivist theories of emotions show that this reply has at least a minimum of conceptual plausibility (cf. e.g. Scarantino 2010, 729–768).

What is empirically problematic about the distinction between reason and sentiment is that in the actual world – for instance in the brain – nothing can really be found which corresponds to it. There seem to be no neural processes that can truly be mapped on this distinction, which has become canonical under the name of 'faculty psychology' (Holtzman 2018). Emotions are *part* of our practical reason, rather than its adversary.

Secondly, both the diagnosis and the explanation of psychopathic deficits had to be revised. It was shown that psychopathic individuals do in fact appreciate the difference between moral and conventional rules (Aharoni, Sinnott-Armstrong, and Kiehl 2012, 484). By now it is clear that their cognitive-motivational deficits – whatever they ultimately consist in – must, to a considerable extent, be explained in rationalist terms (Maibom 2005, 237–257). The psychopath cannot be invoked as a key witness of sentimentalism either.

Thirdly, in the meantime it has been shown that the impact of incidental emotions on our moral judgements is substantially less strong than previously assumed (May 2014, 125–141). If an impact of such changes on moral judgements can be detected at all, it usually remains restricted to particular groups (e.g. people who feel easily and/or intensively disgusted) or particular scenarios (e.g. only some of the experimental vignettes). Moreover, most of the studies yield small effect sizes: they do not succeed at transforming moral approval into disapproval or vice versa. Yet this is exactly what sentimentalists predict.[5] A current meta-analysis even suggests that the effect does not exist at all and is simply due to a *publication bias* for significant findings. If we consider unpublished studies which fail to reject the null hypothesis, the effect disappears completely (Landy and Goodwin 2015, 518–536).

In those cases, however, that do show a substantial emotional impact, it is usually not clear why this impact is supposed to be illegitimate (May 2018). The judgement that discrimination against a certain group is unjust is likely substantially influenced by a sentiment of indignation. But this seems to be perfectly fine, rather than a disturbing discovery. The emotional basis of our moral cognition thus hardly justifies a pessimistic attitude.

(ii) *Debunking arguments*. A second central debate around the problem of reliability of our moral judgement concerns the prospects of so-called debunking arguments [reference omitted]. This strategy, which has been pioneered by Nietzsche, Marx and Freud, extends the hermeneutics of suspicion to those psychological processes that end up determining the content of our moral convictions. I believe that X is morally wrong, and that Y is praiseworthy. But what if I am so convinced just because the functional imperatives of capitalism want it that way, or because the weak – out of resentment and fear – succeeded in convincing the strong of the value of equality, or because unconscious processes dictate orderliness and virtuousness?

Debunking arguments present epistemic defeaters that do not directly attack the truth of a moral judgment, but undermine their justification. Their structure is easily made transparent. Debunking arguments typically consist of an empirical premise, which outlines the causal origin of a moral belief (Nichols 2014, 727–749), and of a normative premise, which discredits this genealogy as epistemically unreliable (Kahane 2011, 103–125). Together they deliver the debunking conclusion:

(1) Our moral judgements M are formed on the grounds of process P.
(2) P is epistemically unreliable.
(3) Therefore, M is epistemically unjustified.

Debunking arguments can attack moral judgements either *globally* or *selectively*. In the above schema, *M* stands for a subset of moral judgements, such that {moral judgements based on disgust}, {evolutionarily formed moral judgements}, {deontological judgements} or even {all moral judgements} can be substituted for *M*. The scope of the debunking argument can be adjusted depending on wether all moral judgements (global) or merely some (selective) are targeted.

A central problem for every debunking attempt is to establish its own stability. Selective debunking is often threatened by an infinite regress, given that the unreliability of a given set of moral judgements can only be evaluated with a reference to another set of moral judgements, whose reliability then immediately becomes doubtful itself (Rini 2016, 675–697). Global accounts, on the other hand, tend to spin out of control, given that the properties which allegedly make all and *only* our moral judgements unreliable can easily be found among non-moral judgements as well (Kahane 2014, 327–341).

There are different types of *debunking* arguments, depending on where they locate the epistemic defect which contaminates a given process of moral belief formation. In some cases, the causal origin of a set of moral convictions is completely 'off track': there is simply no connection between the processes which produce a belief, and the facts the convictions are (or should be) about (Street 2006, 109–166). Evolutionary debunking arguments are frequently of that variety since the selective pressures that have shaped our values have nothing to do with the moral truth (cf. FitzPatrick 2015, 883–904). Radical debunking attempts of this sort, however, are difficult to pull off because the *joint* plausibility of the empirical and normative premise is hard to secure.

Other debunking attempts hold that cognitive processes, even if they *were* once reliable can *become* unreliable due to changes of context. To restrict ones cooperative dispositions to related and/or strategically relevant partners may have been adaptive in the past – for instance in the 'environment of evolutionary adaptation'. Under conditions of modern ultrasociality, however, such parochial altruism may lead to harmful 'us/them'-constructions, with various annoying social, political and economical consequences. The epistemic defect here lies in the obsoleteness of the processes which generate our judgements.

A third source of unreliability can be found in improperly calibrated cognitive processes. Looking back once again to the example of moral judgements based on disgust: even if incidental affects have no major impact on our moral convictions, it is plausible that a quite substantial amount of those convictions depend, in a more fundamental way, on reactions of disgust. Moral judgements which regulate the boundaries of decent behavior with regard to our all-too-human topics – sex, food, death – are hardly comprehensible independently of disgust (Kumar 2017).

The problem with disgust is that it is doxastically oversensitive und thus produces an undesirably high amount of false positives. This is probably connected to the mix of hesitation and enthusiasm with which omnivores like ourselves had to navigate the well-known asymmetry between missed advantages and risks, during foraging. Not to have eaten a seemingly rotten piece of meat, even if it was actually not dangerous, is regrettable, but in the end not fatal. To have eaten something toxic or pathogenic but seemingly safe is a disaster. Sentiments of disgust hence tend to be hypersensitive, and thus classify an excessively high amount of candidates as contaminated. This kind of hypersensitivity can often have horrible consequences (Nussbaum 2009).

Besides hypersensitive cognitive processes like disgust, there are also those which become epistemically unreliable due to their *hypo*sensitivity. Empathy has an immaculate reputation both within and outside the academics; however, there is increasing evidence that moral judgements based on empathy are severely biased and easily distracted (Prinz 2011, 214–233). Concrete empirical studies show that our moral sympathy is very easily exhausted. Phenomena like 'pseudoinefficacy' (Västfjäll, Slovic, and Mayorga 2015, 616) or 'compassion fade' (Västfjäll et al. 2014) suggest that we suffer from rapidly diminishing marginal empathy: we are prone to think that we can't help one because we can't help all (pseudoinefficacy) and we care much less for the marginal person in need than we do for the first (compassion fade). It is simply psychologically impossible for us to care about everyone equally.

Many of our moral judgements are influenced by the sheer contingency of where and when they were formed. *We* believe that autonomy and authenticity are important values; but *elsewhere* people do not believe that and hold harmony and tradition to be more important. *We* believe that one should try new things and cut old ties; however, *in the past* people did not believe that, new things were viewed with skepticism and people preferred the authority of the tried and true. *We* believe that one may do anything that isn't harmful to others; but *others* do not believe that and view this kind of moral minimalism as a sign of moral obtuseness.

These observations are not intended to support moral relativism, according to which when in Rome one should do as the Romans do. Rather, there is the epistemic problem that we have good reasons to consider our own convictions unjustified if there are actual or counterfactual epistemic peers who do not share our beliefs. Christians, for instance, must come to terms with the fact that, had they been born in a different time and/or place they would have believed in Thor or Vishnu rather than Jesus of Nazareth. Likewise, a different evolutionary trajectory would have led to different values. As long as there is nothing to break this epistemic tie, we need to reduce our confidence in our moral beliefs (Bogardus 2016, 636–661).[6]

The strength of debunking arguments consists in the fact that they are not, like some skeptical scenarios, presented as mere logically possible hypotheses, but rather that there is positive and usually very robust evidence that they actually obtain. The evolutionary source of our fundamental value settings is no far-fetched 'brain-in-a-vat' thought experiment, but rather a concrete probability that is strongly supported by evolutionary biology and psychology.

At the same time, the conclusiveness of debunking arguments should not be overestimated, independently of as how suggestive or surprising they appear. As soon as good reasons can be found to reconsider a conviction that was previously undermined, its epistemic status is quickly repaired. Debunking arguments are thus the strongest when those additional reasons are not currently present. For the same reason, the debunking strategy in moral philosophy turned out to be particularly fruitful: on the one hand, it is correct to say that any chain of justification finally ends with mere intuitions – it appears that *p* –; in ethics, however, those intuitive rocks that bend back the spade are usually reached far earlier than in other disciplines. Moreover, intuitions in ethics often play a major role, inasmuch as they directly contribute normatively contentful principles. The evidential force of an apparent self-evident moral intuition is easily neutralized, when its connection to truth is severed. After that, the burden of proof lies with the defenders of that intuition. Often, no good defense is forthcoming.

Debunking arguments often provide very good reasons to consider much of our moral cognition unreliable even if, as mentioned above, such reasons are not *conclusive*. Factors that provoke the derailment of our judgements are almost never introspectively accessible. At the same time, they are ubiquitous, so that we can never confidently rule out contamination. Accepting that many debunking arguments are successful, supports a rationalistic pessimism. Pessimism, because the empirical evidence in favor of debunking strategies exposes large parts of our practice of moral judgement as flawed as well as a great amount of our moral judgements as unsustainable. Rationalistic, inasmuch as this insight itself relies on the availability of at least some basic moral standards in light of which which this assessment is made.

3. The problem of contamination

Many of the most spectacular discoveries in empirical philosophy purport to show that our mind is influenced by factors about which we have no clue. Experimental philosophy has shown, for instance, that our attributions of intentionality happen 'downstream' from normative assessments (Knobe 2010, 315–329; [reference omitted]). That this could be so had not occurred to anyone over the course of 2500 years of professional introspection.

Generally speaking, the 'negative program' (Kauppinen 2007, 95–118) in experimental philosophy has taken up the task to check whether our epistemic, moral or metaphysical intuitions – which philosophers, in characteristic self-aggrandizement, had usually taken to be *a priori* – are influenced by 'irrelevant factors'. Here as always, the last word has not been spoken yet. However, recent systematic reviews suggest that the venerability of armchair philosophy is surpassed only by its naïvete. The reciprocally maintained reassurance among professional philosophers to be capable of the quite literally uninformed divination of the nature of the mind and the world is a phantasy that does not survive closer scrutiny. Instead, it has been shown that our intuitions are influenced, sometimes to a great degree, by our culture, gender, socioeconomic status, age, personality and other trivial, contextual factors like order and framing (Machery 2017, 86f).

In empirical ethics, the contamination strategy has found a home in the debate between deontological and consequentialist moral theories. Joshua Greene's research program attempts to show that the distinction between reliable and unreliable moral judgements can be mappend into the distinction between two types of moral judgements: consequentialist judgements are – at least in principle – trustworthy, deontological judgements are (typically) not.

In the first round of this research program, neuroscience was still expected to do much of the evidential heavy lifting (Greene et al. 2004, 389–400). The idea – roughly – was to demonstrate the contamination of deontological judgements via neuromaging. If, when people make deontological judgements, the lights go on in emotion-related brain regions and, while consequential judgments are made, they do so in the more 'cognitive' regions, the latter type of judgment must be superior to the former. This assumption seemed to enjoy further support from the fact that consequentialist judgments took longer to make, and could be reduced via additional cognitive load.

By now this neuroscientific approach has turned out to be of secondary importance. It is not clear to what extent fMRI results play an independent role in undermining deontological judgements (Berker 2009, 293–329). It is not even clear if there are any brain areas at all that are *specifically* responsible for the processing of emotions [reference omitted]. Moreover, the relevance – or in fact the existence – of variable reaction times has become doubtful. Primarily however, Greene's argumentation is still confronted with massive difficulties regarding the empirical operationalization of the difference between consequentialist and deontological judgements.

The first attempt in this direction was, to put it bluntly, a mess: one may deem consequentialists to be heartless bean counters, but even this theory does not deserve the allegation that one could justify hiring a rapist to restore a relationship in terms of utility-maximization (McGuire et al. 2009, 577–580). The tendency toward such pseudo-utilitarian judgements hardly appears to be connected to actions favored by genuine utilitarians, such as philanthropic donations or vegetarianism, but harmonizes more with various Machiavellian traits (Kahane et al. 2015, 193–209). Later attempts could not get rid of the problem that in the used material, the distinction between deontological and consequentialist judgements had been conflated with the distinction between *intuitive* and *counterintuitive* judgments. Once one considers that there can be counterintuitive deontological and intuitive consequentialist judgements, the envisioned correlations (intuitive/deontological vs. counterintuitive/consequentialist) turn out to be spurious (Kahane et al. 2011, 393–402).

What remains is that snapshot judgements, of any content whatsoever, are often problematic, at least when our automatic-intuitive cognition confronts an ethical problem for which it was not prepared by evolutionary or cultural or individual learning. This is certainly not nothing, but still a disappointing result for those consequentialists who, in light of the apparent denunciation of their theoretical opponents as touchy-feely confabulators, had already popped the champagne in the fridge.

Social Intuitionism extends Greene's suspicion of the illegitimate influence of epistemically dubious emotionally charged emotions on deontological judgements to *all* moral judgments (Haidt 2001, 814). The emotional dog waggles its rational tail, and certainly not on the contrary. Our moral intuitions emerge from six moral 'foundations', which owe their functionality to the evolutionary problems which they had to solve (Graham, Haidt, and Nosek 2009, 1029; [reference omitted]). Conscious moral reflection acts like a layer, constructing sophistic rationalizations of automatically generated intuitions. If this were not so, why don't we give up our moral intuitions when they out to be indefensible, and fall into a state of 'moral dumbfounding' instead (Kennett and Fine 2009, 77–96)?

Here, too, some problems soon became evident. First of all, one cannot stress enough that some participants in Haidt's experiments did question their own moral judgements (cf. also Paxton, Ungar, and Greene 2012, 163–177). People often have an instinct for good reasons, even when it is difficult to articulate them (Jacobson 2012, 289–316). The fact that everything went well *ex post*, does not generally mean that a reckless experiment was a good idea *ex ante*. Finally, one must not underestimate that many test subjects unconsciously refused to believe in the penny dreadful of the allegedly perfectly normal sibling couple deciding to have sex – *just for fun*, as it were – but rather suspected a more sinister background to the story (Royzman, Kim, and Leeman 2015; Guglielmo 2018, 334–337). Finally, in a social context, individually biased reasoning may be a useful part of a collectively efficient enterprise (Sneddon 2007, 731–748; [reference omitted]).

Despite their theoretical and empirical problems, both Greene's dual process theory and Haidt's Social Intuitionism have provided valuable ammunition for my pessimistic argument. Their most radical claims may be overstated. But, academic hairsplitting aside, Greene and Haidt are generally right that at least a sizable portion of our moral judgements is distorted by the influence of irrelevant factors and that at least a great part of our practice of giving and asking for moral reasons consists in the mere rationalization of preformed intuitions, amounting more to social currency and rarely to an openminded search for the truth (Stanley et al. 2018).

Finally, our moral judgements often fail because the necessary *mindware* to discipline them is missing, or because we operate on the basis of positively deceptive mindware [reference omitted]. One can think of the underdeveloped ability to adequately factor in probabilities in calculations of risk, as well as a missing vocabulary to characterize moral infringements appropriately (Fricker 2007) as missing mindware. Contaminated mindware consists in cognitive processes involved in the formation of our moral judgments which are contaminated theories and concepts that let epistemically unjustified intuitions impact our moral cognition. Zero-sum-thinking, in-group/out-group-constructions, or retributive tendencies are possible examples. More generally, qualified moral judgement often depends on so-called 'System III'-dispositions for critical thinking (that is, cognitive dispositions to override and question one's intuitions, [reference omitted]), which are always in short supply. In order to make proper moral judgements, it is not enough to *be able to* reflect on them; first the necessity of critical thinking must be *detected*; then, critical reflection must be *initiated*; finally, every reflective episode must be carefully monitored, executed and concluded. All of this is a time-consuming nuisance; moreover, few people enjoy looking into the eyes of the cold and hard truth. In light of this evidence, it would be implausible to deny that our moral judgements are, at least to a considerable extent, epistemically contaminated. This, too, supports a form of rational pessimism.

4. The problem of motivation

A third problem for the epistemic quality of our moral judgements consists in a toxic incentive structure. The problem of incentives has nothing to do with the fact that many people are lazy or stupid, however true that may be. Rather, it is based on the fact that even well-informed morally decent people with good intentions face incentives not to make well-intentioned, well-informed moral judgments, while having many problematic incentives to make bad judgements. Moreover, not everyone is well-informed and morally decent, and for those who aren't, making epistemically poor moral judgments is overdetermined.

The field of morality, and the political, social and economic domains it is closely related to, are fields in which the correct beliefs are often highly counterintuitive. Basic moral beliefs (*Killing is wrong*) seem relatively easy to grasp. But the trouble begins when these rather abstract general principles need to be applied to the real world and weighed against other principles. Sound reflection, particularly under modern conditions in large, dynamic societies, frequently requires deep critical thinking and override of our automatic moral intuition. This means that due to simple cognitive miserliness (that is, the fact that we are by default reluctant to critically reflect on whatever our beliefs are, Stanovich 2011), we have a standing incentive to make moral judgments on the basis of unreliable processes.

But this is not all there is to it: not only do we lack incentives to make sound judgments, we are indeed positively biased towards unsound ones. When it comes to most moral, political and social questions (such as immigration, technology, economics, or politics), we do not only have an inadequate cognitive toolkit, but operate with distortive cognitive patterns, thus essentially guaranteeing the emergence of false beliefs.

It is an important function of moral judgement not only to evaluate other people, but also to present oneself in a certain morally favorable light. This is justified to some extent (Levy 2020). But over time, this fact does not only help in contaminating public moral discourse through seemingly strict but actually disingenuous moral standards, but also to cultivate one's own moral sensibility in the service of social signals, resulting in a race to the bottom towards excessive moral sensitivity (Tosi and Warmke 2020). When this happens, people may end up employing increasingly stringent moral standards of wrongness and culpability that become more and more difficult to justify. Everyone wants to appear moral, because doing so sends signals of trustworthiness that can help foster mutually beneficial cooperation. So far, so good. But ideally, we would like to appear to be *slightly more* moral than the average person, which may prompt us to ratchet up the rigor of our moralizing. But if most people do this, the ensuing competition for the positional good of ranking more highly in moral purity than the next person can drive society towards exorbitant levels of moral punitiveness.

The logic of collective action applies to the practice of moral judgement in another sense: the overall quality of moral discourse is a public good, and as such tends to be underprovided by aggregated individual actions.[7] Consider, for instance, the objective seriousness and difficulty of topics such as the death penalty, the public health system, taxation, or the problems surrounding migration, and contrast the breathtaking *frivolity* with which basically everyone forms his or her own moral-political convictions about these. The problem is that the costs of incompetent moral judgement can be almost completely 'externalized', while its benefits are almost entirely absorbed by the subject. Most people invest considerably more time and effort in the choice of a new phone than in the choice of their own moral opinion, simply because in the case of the former, the costs and benefits fall on the person making the choice. We do not choose our moral beliefs the way we choose a cell phone, of course; but unless one wants to deny that we have any epistemic obligations whatsoever (or that ought implies can), we must, at least in principle, be able to exert some degree of influence over our moral judgments. This ability is rarely properly exercised though, because when it comes to bad moral judgements, most of their instrumental benefits accrue to the subject – i.e. to feel and appear moral, signal group membership, earn status, vent one's grievances, enjoy bonding over gossip – while the costs can be shifted onto others. Since this logic applies to everyone equally, public moral discourse is subject to a degenerative dynamic. *Each* of us consumes more moral unreasonableness than is good for *all* of us.

Those who reject that the logic of collective action applies to moral cognition in this way are invited to explain why it should be otherwise or, indeed, how it *could* be otherwise. The assumption that we have no reason for a pessimistic view on our moral cognition probably fails already because of simple problems of consistency. Why should we all be sinners in our actions, but saints in our judgments? Recently, behavioral symmetry arguments such as these have become prominent in political philosophy (Freiman 2017). If one thinks about institutional design – such as the justification of regulative interventions by the state – one should not assume without good reason that people act selfishly within markets, but not within politics. Asymmetrical motivational assumptions of that kind require a specific justification. The reasons why governmental action becomes necessary are very often the reasons why it might not work. It is a surprising observation that moral philosophers readily admit that people often *act* immorally or irrationally, but hesitate to concede the same point about moral *judgement*. Why is there a need to engage in moral judgement in the first place? Because people are often weak-willed, short-sided, uninformed, or downright malicious. But if this is true, why should we suspect that the same people are reasonable, decent, unbiased, diligent, informed and precise when it comes to moral judgement?

5. Rationalist optimism

In order to reject the pessimistic argument developed above, there are main three strategies: first, a reference to the impact of *rational learning processes*; second, a recourse to the possibility of *moral expertise*; and third, trust in the epistemic force of *moral progress*.

A first possible optimistic reply aims to restore the good reputation of our moral cognition in terms of rational learning processes [reference omitted]. Recently, such moral learning accounts have turned out to be the most promising version of rationalism about the psychology of moral judgment. Peter Railton, for instance, has shown that deontological judgements can be accounted for with the distinction between model-based and model-free reinforcement learning (Railton 2017, 172–190). Some cognitive processes operate on the basis of 'cached' responses, which encode alternative courses of action as simple situation/action-pairs; other processes of decision – und judgement-making, however, operate with more or less complex causally branched out models of the world. The widespread asymmetry of our moral intuitions in Trolley-dilemma scenarios is due to the fact that in one case, such a cached response is available (Should I push a person to his death? Never!), whereas we revert to model-based decisions in the other case. Moreover, Shaun Nichols and his team succeeded in explaining the principle of double effect, or rather the differentiation between acting and allowing, as a result of rational learning processes (Nichols et al. 2016, 530–554), while some had assumed it to be an innate component of our universal moral grammar (Mikhail 2007, 143–152). Fine-grained rules such as the principle of double effect are difficult to articulate for almost everybody, yet we easily comply with them, even though no one is explicitly taught this rule. This seemed to suggest that the rule had to be nature rather than nurture. Nichols et al. were able to show, however, that this and other rules – *pace* poverty of moral stimulus arguments – can be acquired on the basis of implicit statistical learning processes acting on relatively scant input.

Despite the impressiveness of these results, rational optimists should not become too hopeful. The basic problem with theories of rational learning is that – independently of how refined the learning processes may be – the quality of the learning *results* is always determined by the particular learning *inputs*: garbage in, garbage out. Rational learning theories cannot, as a matter of principle, evade this problem. Since the learning processes which Railton or Nichols et al. refer to are not only open to good moral reasons, but absorb all kinds of data, there is no guarantee that rational learning will deliver acceptable results. Even homophobic, racist or anti-semitic prejudices can be refined through Bayesian updates (Gjesdal 2018). The point here is that whatever improvements to the quality of our moral judgments can be made, these improvements will have to come from *what* we learn (the input) rather than how we learn (rationally or not). If we rationally refine our bad impulses, that does not constitute any improvement at all: it makes things worse. Evidence for rational learning is thus inherently ambivalent.

A second possible optimistic reply states that moral convictions by laymen may be unreliable, yet that we can become ethical experts through training, such that our moral judgements will then not be subject to similar distortions anymore.

That professional moral philosophers make better moral judgements than non-professionals is partly an empirical thesis, which can be tested with empirical methods. Now the first results are in, and it does not look great for us. It is largely uncontroversial that moral judgements should not be influenced by the order in which they are presented (cf. however Horne and Livengood 2017, 1189–1218). But they are, and even philosophical expertise does not seem to change anything about it (Schwitzgebel and Cushman 2012, 135–153). Worse, studying ethics does not even seem to have a beneficial influence on people's actual behavior (Schwitzgebel and Rust 2014, 293–327).

Others counter by arguing that it's not a big problem if philosophers' raw intuitions are as unstable as those of philosophical laymen, since the epistemic advantages of expertise start to show only in the experts' *considered* judgements (Rini 2015, 431–452). It is not clear, however, why raw intuitions and considered judgements should not, in the case of ethicists, be much more congruent, since ethicists presumably already have an opinion about well-known known dilemmas. And even if moral expertise were real, the question of how this expertise is transmitted to non-experts remains open, and it is doubtful that such a form of trickle-down epistemology can be defended. Machery sums it up nicely: 'It is time to call the Expertise Defense what it is: a myth' (Machery 2017, 169).

A final possible reason against pessimism about moral judgment may be found in the human potential for moral progress. For those who aren't moral nihilists, it is very difficult *not* to consider various important developments in recent human history moral progress. It is almost irresistible to welcome the abolition of slavery, the introduction of female suffrage, the implementation of civil rights for minorities, or the increasing recognition of the rights of non-human animals as a progressive development (Buchanan and Powell 2015, 37–67).

How can these huge improvements be reconciled with pessimistic arguments? The conflict, in my opinion, is only apparent. Overall, the glacial pace of moral progress does not really support an optimistic view of our moral capacities. The measure of progress remains the century, or at least the generation. Scientific knowledge or technological innovations are often adopted much more quickly, which suggests that this is a problem that is specific to our moral beliefs: we are especially reluctant to accept changes to our norms and values, even when such changes seem eminently justified. Most of the arguments against slavery had been well-known for a very long time before slavery was finally abolished. But people were reluctant to *accept* these arguments due to bias, ignorance, selfishness, or simply because they could get away with it. And when new moral insights are implemented, this usually happens in only a few places. Note that my pessimistic argument does not entail that we never change our mind about moral matters, only that we are very often very bad at it. The prediction that results from this is that we should see some moral progress, but that it should be hard-won and painfully slow. The structure of moral progress that we see in the world – its tardiness and regionality – more or less exactly confirms pessimistic predictions.

6. What now? Perspectives

I have argued that our practice of moral judgement is highly deficient. Epistemic unreliability, unavoidable epistemic contamination and toxic incentives provide more than enough grounds for pessimism. Now what? In this last section, I want to draw some pessimistic conclusions and outline some alternatives to the status quo going forward: quietism, elitism, and abolitionism. I will advocate a moderate form abolitionism – which I will refer to as 'eliminativism' – according to which we should moralize much more rarely and responsibly.

A first possible answer to the 'Now what?' question is quietism. Following Wittgenstein, (moral) philosophy could leave everything as it is. If Joshua Knobe is right, we are 'moralizers through and through' anyway (Knobe 2010, 315–329). We could not stop making moral judgment even if we wanted to. Even our non-moral judgements are so deeply 'suffused' with morality that the hyperactivity of our moral cognition – according to the 'principle of minimal psychological realism' (Flanagan and Flanagan 2009) – remains without alternatives. It is certainly true that a complete end to our moralizing remains unavailable. This, however, has not been my proposal, and the anticipated objection is thus only partly relevant. It is an empirical question by which amount *x* our practice of moral judgment can be reduced. The rational pessimism I advocate then recommends that moral judgment be reduced by that amount *x*.

The elitist alternative states that the (moral, not legal) right to judge morally should be limited to those people which are *de facto* competent. But a moral 'epistocracy' (cf. Brennan 2016) of this kind may seem easier than it is. A limitation of the right to vote regulates people's access to the ballot box every four years; a limitation to the moral right to make moral judgments regulates – what exactly? Furthermore, the idea that there is something like an 'elite' of subjects capable of competent moral cognition is quite doubtful. In fact, it is a central feature of the epistemic deficits identified above that no one is immune to them.

The eliminativist alternative simply consists in the claim that, if we are not able to make reliable moral judgements, we should make none at all. I would like to stress once again that a selective eliminativism about moral judgment is a common idea: Joshua Greene, for instance, wants us to eliminate (or at least significantly reduce) deontological judgements, Peter Singer seeks to eliminate speciecist judgements, other authors described the extensive moral neutralization of various issues such food, sex or clothing as moral progress that consists in an elimination of 'surplus moral constraints' (Buchanan and Powell 2017, 108–135).

The suggestion to thoroughly revise our practice of moral judgement may cause discomfort. But this does not mean that it is wrong. There is already a series of respectable proposals of that sort with regard to issues other than moral judgment. It has been suggested that we should stop attributing freedom of will; many are eliminativists towards theistic or scientifically obsolete vocabulary ('God', 'Phlogiston', and the like). Eliminativism about moral judgment is thus in good company.

My proposal may, if by anyone, be taken to heart by the wrong people. We cannot afford, this argument goes, to let the epistemically diligent and morally scrupulous unilaterally lower their moral voice, while simpletons and charlatans predictably will do no such thing. This objection is partially legitimate. However, I want to stress that, in this paper, I care about the *ethical* question of what our obligations are rather than the *empirical* question of whether it would be wise to announce these ethical obligations to the wider public. Eliminativism may, for moral reasons, best be treated with discretion.

One has to distinguish the eliminativism proposed here from other forms of eliminativism.[8] Some authors argue that we must eliminate moral judgement simply because it is systematically erroneous. Others suggest that moralizing, independently from the question about truth, is more harmful than useful – even truthful or well-grounded judgements do more harm than good. Both are not my position. The eliminativism presented here is epistemic and consequentialist: it is based on the double thesis that moral judgments *can* create harm, and probably *do* – namely when they are based on dubious cognitive processes. It may *also* be true that moral values polarize and that they lead to terrorism and cruelty. The list of those who were killed in the name of morality is no doubt long. However, this was not my argument here.

A moral argument against moral judgment seems self-defeating. Shouldn't the very same skepticism that I recommend we apply to moral judgment in general apply to the moral judgments contained in this paper? This, it seems, sends us off into a problematic justificatory regress. However, I do not think that this objection succeeds, for various reasons having to do with the difference between everyday moralizing and moral *philosophy*. (I am not suggesting, of course, that moral philosophers are smarter or better people, and I have no doubt that when ethicists leave their study, their moral thinking suffers from the very same problems diagnosed above.) For one thing, the problem of bad incentives is much smaller, and may in fact be the opposite. Moral philosophers get *rewarded*, by and large, for not being obviously misinformed about the moral facts, and for first demonstrating a familiarity with opposing points of view and the state of the art of the debate on the respective issue they are writing about. Moreover, the obligations moral philosophers have when it comes to the moral claims and arguments they develop in a scholarly context are objectively different from the obligations faced by ordinary people in everyday contexts. In some respects, their obligations are stricter: as mentioned above, there are penalties involved for being ignorant of the facts and the best available arguments in the field. In some respects, their obligations are more lenient, because the judgments moral philosophers make are embedded in an adversarial institution in which the potential costs of advancing risky or counterintuitive moral claims are offset by the group-level epistemic benefits of staging a competition between disagreeing moral points of view.[9] A lawyer may have a strong duty to defend the innocence of their client even when the evidence speaks decisively against the accused. In everyday contexts, this may well be the wrong thing to do.

It may seem obvious that the best solution to the problem that the epistemic quality of our moral judgments is low would be to *improve* their quality. This is in some sense true:

when we make moral judgments, we should strive to make them more carefully, be better informed, and stay open to revising our point of view. Whenever this is genuinely possible, this alternative may be preferable to eliminativism. However, I am skeptical of this ameliorative route, largely due to the arguments sketched in this paper: the epistemic problems affecting our moral beliefs are not going anywhere, even with our best efforts at improving them. But it remains true though that the point of this paper could be rephrased as such: (1) People ought to try harder to make better moral judgments; (2) in those cases in which (1) is not feasible, people should refrain from moralizing. The evidence, I believe, suggests that (1) is, as a matter of empirical fact, very difficult to achieve. So (2) – the eliminativist solution – is *in most cases* the preferable one.

A strong eliminativism would entail that we stop making moral judgments entirely. This seems neither possible nor desirable to me. Yet a more moderate eliminativism, according to which we should drastically cut back on moralizing appears plausible. The argument is simple: we have grounds to suppose that unreliable moral judgements can cause tremendous harm. Recent evidence from moral psychology and cognitive science shows that many of our moral judgements are in fact based on such unreliable processes. That is why we should often not engage in moral judgment, simply because we are no good at it. At least this would be the right thing to do.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Horizon 2020 Framework Programme: [grant number 851043].

Notes on contributor

Hanno Sauer is a philosopher working at Utrecht University in the Netherlands. His main research interests are in metaethics and moral psychology. His three most recent books are Moral Judgments as Educated Intuitions (MIT Press 2017), Moral Thinking, Fast and Slow (Routledge 2018) and Debunking Arguments in Ethics (Cambridge University Press 2018).

Notes

1. This paper draws on material from Moralischer Rationalismus. Eine pessimistische Verteidigung (Paulo and Bublitz 2020).
2. For a similar strategy, see Brennan, Jason & Freiman, Christopher (2020). Moral philosophy's moral risk. *Ratio*.
3. (cf. Garner and Joyce 2018)
4. Regarding the negative consequences of emotional deficits for our general decision-making ability, also cf. Damasio (2006). Descartes' error. Random House.
5. Cf. however Prinz (2016). Sentimentalism and the moral brain; Liao (2016).
6. Much of the current debunking debate is about whether our moral convictions can be immunized against the problem of unreliability by employing a metaethical maneuver (Street 2006, 109–166). Here, the fundamental idea is that the success of debunking arguments depends on realistic premises, which can be easily sacrificed. Our moral judgements cannot be off track, if there is no track to be *on*. The metaethical neutralization of the skeptical force of

debunking arguments rests on the assumption that it would be so unattractive to epistemically undermine *all* of our moral beliefs that rather than countering the challenge, we may change the subject. On this second-order level, there are no discussions about which ethical convictions are justified or not, but rather about the very nature of our moral convictions and values, and here, etiological arguments have no substantive relevance anymore. This issue is to be settled internally, within moral discourse itself. Against this suggestion one can object that, first, such an evasive maneuver curiously becomes viable only at the point where *all* moral judgements become the goal of debunking. With each further subset of the superset {all moral judgements} that is identified as unreliable, the investment in the skeptical capital of debunking becomes more profitable; only to fall back to zero, as soon as *all* moral judgments become the target. This is a surprising result, to say the least. Secondly, it means that the meta-ethical turn is not an available answer to selective debunking. As long as only a part of our moral judgements are threatened – even if it is a large part – there will be no problem of a general moral skepticism, and thus no justification to divert the threat to a meta-ethical territory. Most debunking attempts – among them some of the most plausible ones – are selective and hence remain a substantive, first-order problem.

7. In developing the arguments of this section, I heavily draw on the literature on "rational irrationality" and apply it to moral cognition. See, for instance, Caplan (2011), Bogardus (2016) or Somin (2016) who develop similar arguments for people's political beliefs and voting behavior.
8. Discussions of different eliminativistic positions in relation to moral judgements can be found in Garner and Joyce (2018). *The end of morality: Taking moral abolitionism seriously*. Routledge.
9. Brennan, Jason & Freiman, Christopher (forthcoming). Moral philosophy's moral risk. *Ratio*.

References

Aharoni, E., W. Sinnott-Armstrong, and K. A. Kiehl. 2012. “Can Psychopathic Offenders Discern Moral Wrongs? A new Look at the Moral/Conventional Distinction.” *Journal of Abnormal Psychology* 121 (2): 484–497.

Berker, S. 2009. “The Normative Insignificance of Neuroscience.” *Philosophy & Public Affairs* 37 (4): 293–329.

Blair, R. J. R. 1995. “A Cognitive Developmental Approach to Morality: Investigating the Psychopath.” *Cognition* 57 (1): 1–29.

Bogardus, T. 2016. “Only all Naturalists Should Worry About Only one Evolutionary Debunking Argument.” *Ethics* 126 (3): 636–661.

Brennan, J. 2016. *Against Democracy*. Princeton, NJ: Princeton University Press.

Brennan, J., and C. Freiman. 2020. “Moral Philosophy's Moral Risk.” *Ratio* 33 (3): 191–201.

Buchanan, A., and R. Powell. 2015. “The Limits of Evolutionary Explanations of Morality and Their Implications for Moral Progress.” *Ethics* 126 (1): 37–67.

Buchanan, A., and R. Powell. 2017. “De-moralization as Emancipation: Liberty, Progress, and the Evolution of Invalid Moral Norms.” *Social Philosophy and Policy* 34 (2): 108–135.

Caplan, B. 2011. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies-New Edition*. Princeton, NJ: Princeton University Press.

Damasio, A. R. 2006. *Descartes' Error*. New York City, NY: Random House.

FitzPatrick, W. J. 2015. “Debunking Evolutionary Debunking of Ethical Realism.” *Philosophical Studies* 172 (4): 883–904.

Flanagan, O. J., and O. J. Flanagan. 2009. *Varieties of Moral Personality: Ethics and Psychological Realism*. Cambridge, MA: Harvard University Press.

Freiman, C. 2017. *Unequivocal Justice*. New York, NY: Routledge.

Fricker, M. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Garner, R., and R. Joyce. 2018. *The end of Morality: Taking Moral Abolitionism Seriously*. New York, NY: Routledge.

Gjesdal, A. 2018. “Moral Learning, Rationality, and the Unreliability of Affect.” *Australasian Journal of Philosophy* 96 (3): 460–473.

Graham, J., J. Haidt, and B. A. Nosek. 2009. “Liberals and Conservatives Rely on Different Sets of Moral Foundations.” *Journal of Personality and Social Psychology* 96: 5.

Greene, J. D., L. E. Nystrom, A. D. Engell, J. M. Darley, and J. D. Cohen. 2004. “The Neural Bases of Cognitive Conflict and Control in Moral Judgment.” *Neuron* 44 (2): 389–400.

Guglielmo, S. 2018. “Unfounded Dumbfounding: How Harm and Purity Undermine Evidence for Moral Dumbfounding.” *Cognition* 170: 334–337.

Haidt, J. 2001. “The Emotional dog and its Rational Tail: a Social Intuitionist Approach to Moral Judgment.” *Psychological Review* 108 (4): 814–834.

Holtzman, G. S. 2018. “A Neuropsychological Challenge to the Sentimentalism/Rationalism Distinction.” *Synthese* 195 (5): 1873–1889.

Horne, Z., and J. Livengood. 2017. “Ordering Effects, Updating Effects, and the Specter of Global Skepticism.” *Synthese* 194 (4): 1189–1218.

Jacobson, Daniel. 2012. “Moral Dumbfounding and Moral Stupefaction.” In *Oxford Studies in Normative Ethics*, edited by Mark Timmons, vol. 2, 289–316. Oxford: Oxford University Press.

Kahane, G. 2011. “*Evolutionary Debunking Arguments*.” *Noûs* 45 (1): 103–125.

Kahane, G. 2014. “*Evolution and Impartiality*.” *Ethics* 124 (2): 327–341.

Kahane, G., J. A. Everett, B. D. Earp, M. Farias, and J. Savulescu. 2015. “‘Utilitarian’ Judgments in Sacrificial Moral Dilemmas do not Reflect Impartial Concern for the Greater Good.” *Cognition* 134: 193–209.

Kahane, G., K. Wiech, N. Shackel, M. Farias, J. Savulescu, and I. Tracey. 2011. “The Neural Basis of Intuitive and Counterintuitive Moral Judgment.” *Social Cognitive and Affective Neuroscience* 7 (4): 393–402.

Kauppinen, A. 2007. “The Rise and Fall of Experimental Philosophy.” *Philosophical Explorations* 10 (2): 95–118.

Kennett, J., and C. Fine. 2009. “Will the Real Moral Judgment Please Stand up?” *Ethical Theory and Moral Practice* 12 (1): 77–96.

Knobe, J. 2010. “Person as Scientist, Person as Moralist.” *Behavioral and Brain Sciences* 33 (4): 315–329.

Kumar, V. 2017. *Foul Behavior*, Philosophers’ Imprint 17.

Landy, J. F., and G. P. Goodwin. 2015. “Does Incidental Disgust Amplify Moral Judgment? A Meta-Analytic Review of Experimental Evidence.” *Perspectives on Psychological Science* 10 (4): 518–536.

Levy, N. 2020. “Virtue Signalling is Virtuous.” *Synthese* 1–18.

Liao, S., ed. 2016. *Moral Brains: The Neuroscience of Morality*, 45–73. New York City, NY: Oxford University Press.

Machery, E. 2017. *Philosophy Within its Proper Bounds*. New York City, NY: Oxford University Press.

Maibom, H. L. 2005. “Moral Unreason: The Case of Psychopathy.” *Mind & Language* 20 (2): 237–257.

May, J. 2014. “Does Disgust Influence Moral Judgment?” *Australasian Journal of Philosophy* 92 (1): 125–141.

May, J. 2018. *Regard for Reason in the Moral Mind*. New York City, NY: Oxford University Press.

McGuire, J., R. Langdon, M. Coltheart, and C. Mackenzie. 2009. “A Reanalysis of the Personal/Impersonal Distinction in Moral Psychology Research.” *Journal of Experimental Social Psychology* 45 (3): 577–580.

Mikhail, J. 2007. “Universal Moral Grammar: Theory, Evidence and the Future.” *Trends in Cognitive Sciences* 11 (4): 143–152.

Nichols, S. 2014. “Process Debunking and Ethics.” *Ethics* 124 (4): 727–749.

Nichols, S., S. Kumar, T. Lopez, A. Ayars, and H. Y. Chan. 2016. “Rational Learners and Moral Rules.” *Mind & Language* 31 (5): 530–554.

Nussbaum, M. C. 2009. *Hiding from Humanity: Disgust, Shame, and the law*. Princeton, NJ: Princeton University Press.

Paulo, N., and Ch. Bublitz. 2020. *Empirische Ethik*. Berlin: Suhrkamp.

Paxton, J. M., L. Ungar, and J. D. Greene. 2012. “Reflection and Reasoning in Moral Judgment.” *Cognitive Science* 36 (1): 163–177.

Prinz, J. 2006. “The Emotional Basis of Moral Judgments.” *Philosophical explorations* 9 (1): 29–43.

Prinz, J. 2007. *The Emotional Construction of Morals*. New York City, NY: Oxford University Press.

Prinz, J. 2011. “Against Empathy.” *The Southern Journal of Philosophy* 49: 214–233.

Prinz, J. 2016. *Sentimentalism and the moral brain*.

Railton, P. 2017. “Moral Learning: Conceptual Foundations and Normative Relevance.” *Cognition* 167: 172–190.

Rini, R. A. 2015. “How not to Test for Philosophical Expertise.” *Synthese* 192 (2): 431–452.

Rini, R. A. 2016. “Debunking Debunking: a Regress Challenge for Psychological Threats to Moral Judgment.” *Philosophical Studies* 173 (3): 675–697.

Rini, R. A. 2017. “Why Moral Psychology is Disturbing.” *Philosophical Studies* 174 (6): 1439–1458.

Royzman, E. B., K. Kim, and R. F. Leeman. 2015. “The Curious Tale of Julie and Mark: Unraveling the Moral Dumbfounding Effect.” *Judgment & Decision Making* 10 (4): 296–313.

Scarantino, A. 2010. “Insights and Blindspots of the Cognitivist Theory of Emotions.” *British Journal for the Philosophy of Science* 61: 4.

Schwitzgebel, E., and F. Cushman. 2012. “Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and non-Philosophers.” *Mind & Language* 27 (2): 135–153.

Schwitzgebel, E., and J. Rust. 2014. “The Moral Behavior of Ethics Professors: Relationships among Self-Reported Behavior, Expressed Normative Attitude, and Directly Observed Behavior.” *Philosophical Psychology* 27 (3): 293–327.

Sneddon, A. 2007. “A Social Model of Moral Dumbfounding: Implications for Studying Moral Reasoning and Moral Judgment.” *Philosophical Psychology* 20 (6): 731–748.

Somin, I. 2016. *Democracy and Political Ignorance: Why Smaller Government is Smarter*. Stanford: Stanford University Press.

Stanley, M. L., A. M. Dougherty, B. W. Yang, P. Henne, and F. De Brigard. 2018. "Reasons Probably won’t Change Your Mind: The Role of Reasons in Revising Moral Decisions." *Journal of Experimental Psychology: General* 147 (7): 962.

Stanovich, K. 2011. *Rationality and the Reflective Mind*. Oxford: Oxford University Press.

Street, S. 2006. “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies* 127 (1): 109–166.

Tosi, J., and B. Warmke. 2020. *Grandstanding: The Use and Abuse of Moral Talk*. Oxford: Oxford University Press.

Västfjäll, D., P. Slovic, and M. Mayorga. 2015. “Pseudoinefficacy: Negative Feelings from Children who Cannot be Helped Reduce Warm Glow for Children who Can be Helped.” *Frontiers in Psychology* 616: 1–12.

Västfjäll, D., P. Slovic, M. Mayorga, and E. Peters. 2014. “Compassion Fade: Affect and Charity are Greatest for a Single Child in Need.” *PloS one* 9 (6): e100115.