Moral Judgement and Moral Progress
The Problem of Cognitive Control

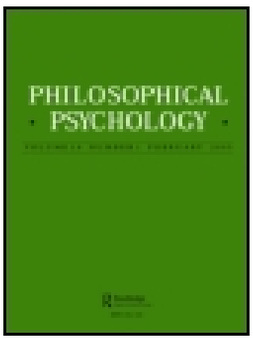
Klenk, Michael; Sauer, Hanno

DOI
10.1080/09515089.2021.1931670

Publication date
2021

Document Version
Final published version

Published in
Philosophical Psychology



Philosophical Psychology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cphp20

Moral Judgement and Moral Progress: The Problem of Cognitive Control

Michael Klenk & Hanno Sauer

To cite this article: Michael Klenk & Hanno Sauer (2021): Moral Judgement and Moral Progress: The Problem of Cognitive Control, Philosophical Psychology, DOI: 10.1080/09515089.2021.1931670

To link to this article: https://doi.org/10.1080/09515089.2021.1931670

Published online: 02 Jul 2021.

PHILOSOPHICAL PSYCHOLOGY
https://doi.org/10.1080/09515089.2021.1931670

Moral Judgement and Moral Progress: The Problem of Cognitive Control

Michael Klenk[a] and Hanno Sauer[b]

[a]Department of Values, Technology and Innovation, TU Delft, Delft, The Netherlands; [b]Department of Philosophy and Religious Studies, Utrecht University, Utrecht, The Netherlands

ABSTRACT

We propose a fundamental challenge to the feasibility of moral progress: most extant theories of progress, we will argue, assume an unrealistic level of cognitive control people must have over their moral judgments for moral progress to occur. Moral progress depends at least in part on the possibility of individual people improving their moral cognition to eliminate the pernicious influence of various epistemically defective biases and other distorting factors. Since the degree of control people can exert over their moral cognition tends to be significantly overestimated, the prospects of moral progress face a formidable problem, the force of which has thus far been underappreciated. In the paper, we will provide both conceptual and empirical arguments for this thesis, and explain its most important implications.

ARTICLE HISTORY
Received 15 August 2019
Accepted 13 May 2021

KEYWORDS
Moral progress; moral psychology; situational factors; moral development

Two hundred years ago, almost everyone in the world was poor; the US economy thrived on slavery; millions of Chinese girls' feet were bound, children were working long days in factories, and English gentlemen were killing each other in duels; women's right were nowhere to be found; animal rights were unheard of. Today, poverty is in decline and democracy on the rise. Civil rights are solidifying, and the idea that animals ought to be treated humanely is gaining popularity. Foot binding and dueling are recognized for the bizarre cruelties that they are. In many places, child labor has long been abolished.

Many would classify these developments – even if they are not happening everywhere, too slowly, and never perfectly – as instances of moral progress.[1] Moral progress occurs when social change is headed toward moral improvement. But are there limits to these developments? Recently, a variety of "conservative" challenges to the prospects of moral progress have gained increasing attention. The basic idea behind these challenges is that there are certain constraints on how much progress can be achieved.

CONTACT Michael Klenk m.b.o.t.klenk@tudelft.nl Department of Values, Technology and Innovation, TU Delft, Jaffalaan 5, Delft 2628 BX, The Netherlands

Stronger forms of conservatism suggest that, in fact, we may have already reached the limits of moral progress. The strongest of them hold that we have already begun to overstretch them.

In some cases, such feasibility limitations on the possibility of moral progress are thought to be due to the hard-wired constraints of our evolutionarily inherited psychology. If, for instance, we are innately prepared to carve up the social world in terms of *us* and *them*, then certain "inclusivist shifts" (Buchanan & Powell, 2016) toward an expanding circle of moral concern (Singer, 2011) may be out of reach. Alternatively, if modern forms of hyper-sociality are underwritten by a set of culturally evolved social institutions whose functioning remains causally opaque to us, then trying to bring about progressive change through deliberate social design may promise more harm than good (Kelly & Hoburg, 2017).

We take no stand on the plausibility of these challenges here. Instead, we aim to dig deeper, and propose a more fundamental challenge to the feasibility of moral progress. The most prominent extant theories of progress, we will argue, assume an unrealistic level of cognitive control people must have over their moral judgments for moral progress to occur. Moral progress depends at least in part on the possibility of individual people improving their moral cognition to eliminate the pernicious influence of various epistemically defective biases and other distorting factors. Since the degree of control people can, in fact, exert over their moral cognition tends to be significantly overestimated, at least by theories of moral progress, the prospects of moral progress face another formidable and thus far underappreciated obstacle. In what follows, we will provide both conceptual and empirical arguments for this thesis, and explain its most important implications.

Our paper has five sections. Section 1 introduces the problem of cognitive control about moral judgment for theories of moral progress. Section 2 motivates the control requirement for moral progress. In section 3, we argue that two prominent classes of views about moral progress, what we call end-point views and functionalist views, entail that moral progress requires individual's to control their moral judgments. In section 4, we provide a systematic assessment of the empirical evidence to show that there is a *pervasive* influence of *morally irrelevant* situational factors on moral judgment. Hence, the control requirement for moral progress, posed by end-point views and functionalist views, is not met. In section 5, we consider objections to our view.

1. The problem of cognitive control and its relation to moral progress

How does learning about the psychological causes of our normative beliefs affect our self-understanding of moral progress? Recent findings in

cognitive science and moral psychology suggest that we are "strangers to ourselves" (T. Wilson, 2004). For instance, in one well-known experimental setup, subjects have to judge whether it would be permissible to redirect a trolley that threatens to kill five workers on a railroad track so that it kills a single worker instead (Greene et al., 2001). A consistent finding is that seemingly irrelevant influences sway people's moral judgments, such as the order in which different cases are presented (e.g., Petrinovich & O'Neill, 1996). We aim to show that these experimental findings have skeptical implications that threaten the possibility of moral progress.

Our argument is based on the claim that moral progress requires the possibility of *controlling* our moral judgments so that they become less susceptible to normatively irrelevant influences. The idea that increased control over our moral cognition is needed to improve our moral judgments is endorsed even by those whose primary aim is to debunk, or undermine, our moral cognition as largely unreliable or confabulatory. For example, Jonathan Haidt (2001) and Joshua Greene (2008), both prominent debunkers, argue that gaining greater insight into our judgments will allow us to control them better in the future, presumably to make *better* moral judgments. Appiah (2009) and Bloom (2016) hint at a similar *control requirement for moral progress*: empirical research on the nature of moral cognition will allow us to adjust our moral judgments in light of new findings and make better moral judgments. Though we might not be able to get rid of some bugs (compare: seeing heatwaves on the asphalt as water), we might be able to control their influence (compare: not seeking out the heatwaves in search of water).

We will zoom in on the control requirement as a precondition for moral progress. Must we free ourselves from the influence of morally irrelevant situational factors to make moral progress? This issue immediately raises two questions. First, how strong is the influence of irrelevant situational factors? And second, what does this influence have to do with moral progress? Our answer to the first question is based on a recent review and meta-analysis of the empirical literature regarding the influence of situational factors in moral judgment (Klenk, 20212021), which shows that this influence is strong and pervasive.

The value of that observation for those already convinced of the pervasive influence of situational factors on moral judgment is twofold. First, we infer three general points about the nature of the influence of situational factors from the review. Second, we consider how the influence of situational factors matters for theories of moral progress, which is of philosophical interest even if all points about the influence of situational factors are already well established.

To answer the question of what the significance of situational influences on moral judgements is for the possibility of moral progress, we will

consider various contemporary accounts of moral progress and argue that all involve a commitment to the control requirement. We will argue that there is a thus far unexplored connection between the evidence for irrelevant influence and current debates about moral progress (e.g., Buchanan & Powell, 2018; Jamieson, 2002; C. Wilson, 2010): for moral progress to occur, individuals need (among other things) to make more controlled (in the sense of being insensitive to irrelevant situational factors) moral judgments. This requirement turns out to be harder to satisfy than commonly thought.[2]

2. Motivating the control requirement for moral progress

Consider the abolition and moral condemnation of chattel slavery. Slavery stands today as a paradigmatic moral wrong. Importantly, as Anderson writes, the abolition of slavery was underwritten by a "transformation of moral consciousness" (Anderson, 2014, p. 2). There was a time when people judged that slavery was not morally objectionable. The fact that people started objecting to chattel slavery on moral grounds was, though surely not sufficient, undoubtedly one reason for its demise (Buchanan & Powell, 2018; J Cohen, 1997). Nowadays, slavery is practically universally regarded as morally wrong. Correct moral judgments, therefore, seem to make a crucial contribution to progress. And in order for people to be able to make better moral judgments, they must control the unwanted influence of various discriminatory biases.

It may seem as if we are conflating two different senses of moral progress here. For one thing, progress could refer to change for the better in the norms and values that regulate individual and collective behavior. For another thing, moral progress could consistof making better – more informed, more evidence-sensitive, less biased – moral judgments. The feasibility challenge we want to put forward is that for moral progress to occur, people need to be able to make more accurate, more coherent and more controlled moral judgments, which speaks to the second sense of moral progress. However, the instances of moral progress identified above – the abolition of slavery, women's suffrage, the gradual disappearance of child labor – concern the first sense of moral progress (improvements in the norms and values governing social practices). It, therefore, seems that even if we could show that people have a hard time satisfying the control requirement, this would contribute little or nothing to showing that the most critical forms of moral progress – in terms of people making better judgments – are infeasible.

However, we believe that the two notions of moral progress are linked in important ways: we hold that certain epistemic improvements in exerting control over the influence of various distorting played *a crucial role* in bringing about the substantive normative changes that we want to classify

as progress. Thus, if it turns out that we are bad at controlling our biases, this will also mean that we are proportionally bad at promoting or continuing significant progressive developments. If the control requirement is not satisfied, it may still be possible in principle for normative moral progress to occur, and there are some historical examples for this. But without control over our cognitive biases, the feasibility of such progressive changes is, if not wholly impossible, at least significantly impeded.

An analogy with science supports the claim that making correct judgments is critical for moral progress. Correct judgments are those that either better represent reality or better cohere with the rest of one's judgments. Thus, progress in science can be understood in terms of making unbiased judgments: judgments that are guided by the evidence alone. To the extent that scientific progress requires less contaminated beliefs, we should defeasibly expect the same within the domain of morality.[3]

These examples are supposed to lend some initial credibility to the view that moral judgments that are influenced by irrelevant situational factors are deficient and that moral progress thus depends on controlling the influence of such factors. Ideally, a moral judgment should respond only to the morally relevant features of a situation. Reducing the influence of situational factors on our moral judgment will facilitate moral progress by getting rid of irrelevant influences on our moral judgment.[4] This view is also widely endorsed by other writers (Anderson, 2014; Buchanan & Powell, 2018; Jamieson, 2017; Macklin, 1977; Shermer, 2016; C. Wilson, 2010).

3. Moral progress requires cognitive control

Two prominent views of moral progress, *end-point views* and *functionalist views*, entail that morally irrelevant influences (and the lack of cognitive control induced by them) on moral judgment are detrimental to moral progress.

Before discussing both sets of views in greater detail, an important clarification is in order about the relation between moral judgment and moral behavior. The control requirement for moral progress, as well as the experimental findings that we will review below, concern moral judgment, not behavior. However, there is no immediate connection between moral judgment, and moral behavior. Critics might claim that only the latter matters for moral progress. Someone's judgment that it is acceptable to pull a switch in a trolley scenario is, in itself, no information about whether the person *would* pull the switch. The worry is that behavior rather than judgment is what matters for moral progress. If that were so, then focusing on moral judgment, as we do in this paper, may be neither here nor there.

The worry can be dispelled because even if moral behavior is necessary for *full* moral progress, we can show that moral judgment *also* plays a role

on any sensible view of moral progress. That is, though correct moral judgments alone may be insufficient for full moral progress (cf. Tam, 2019), our argument merely requires that correct moral judgments are of *some* importance for full moral progress. By "full moral progress", we do not mean progress where a final state of moral perfection is reached such that no further moral improvements are possible. Rather, full moral progress involves progress not just in our beliefs, but in our beliefs and our actual behavior. Improved beliefs and improved conduct are both part of moral progress, but only the realization of *both* amounts to full moral progress. So our view is that, since improvements in people's moral beliefs are a necessary component of full moral progress. Since the control requirement matters for the possibility of improving people's moral beliefs, the control requirement ends up mattering for moral progress. Some moral improvements in people's behavior – namely those that are possible without requiring said doxastic improvements – may still occur, but these will be limited in scope and degree.

Here, the distinction between improvement from a moral point of view and *robust* moral progress is important (cf. Buchanan & Powell, 2018). Improvement from a moral point of view is when a new state "conforms better to valid moral norms or better realizes sound moral values" (Buchanan & Powell, 2018, p. 48). That state, however, need not be brought about by human capacities: it might be an entirely fortuitous circumstance that, say, there are fewer pathogens in the air and thus fewer people die of infectious diseases. Hence, robust moral progress must also involve better, more accurate moral judgments. As Buchanan and Powell have it, moral progress at least partly depends on "the exercise of or improvements of the moral powers" (Buchanan & Powell, 2018, p. 51). Among these states, however, there might be changes that involve human motivational capacities, and they might lead to outcomes that are improvements from a moral point of view. For example, humans might stop all wars amongst themselves if faced with alien aggression. This peace would be moral progress, on one level, but the behavior might be purely instrumental, and there is a clear sense in which making peace for the right reasons would constitute even greater progress. For the purposes of this essay, the relevance of improvement from a moral point of view, or of a telic (outcome-oriented) conception of progress need not be disputed. It suffices to note that there is a form of moral progress, which we can call process-related moral progress, that requires correct moral judgments.

We will discuss the two main process-related accounts of moral progress in what follows.[5] Both entail that reducing the influence of situational factors is required for moral progress. More schematically, we defend the following argument:

P1. End-point views (EPVs) and functionalist views (FVs) of moral progress entail that full moral progress requires cognitive control. That is, according to EPVs and FVs, moral progress depends on individuals making moral judgments that are not influenced by morally irrelevant situational factors.

P2. Morally irrelevant situational factors pervasively influence human moral judgments.

P3: People are unlikely to be able to remedy the influence of morally irrelevant situational factors on their moral judgments.

C: Therefore, on EPVs or FVs, full moral progress is unlikely.

Our conclusion casts a into doubt the prospects of moral progress: the uncontrolled influence of morally irrelevant factors on our moral judgments constitutes a problem for two promising classes of theories of moral progress. We will explain and defend the argument's premises over the course of the rest of the paper.

3.1. End-point views

End-point views of moral progress posit a target state, or end-point, that must be approached (in an appropriate way) for moral progress to occur. End-point views differ in what they take 'approaching' to consist in (e.g., the making of different moral judgments, or different moral actions) and how they define the end-point. The most relevant end-point views posit *epistemic, formal*, or *substantive* criteria for moral progress.[6]

Consider the *epistemic* version first. Uncovering new or modified moral truths, in analogy to scientific progress, is what constitutes moral progress (Boyd, 1988).[7] Approaching the end state is done by finding out more about the world. C. Wilson (2010), in criticizing the metaphysical commitments incurred by this epistemic conception, argues that the moral improvement exhibited by one state of affairs over another can be conceived of without incurring metaphysical commitments. Wilson's proposal draws on the idea that moral principles, which are assumed to be true, provide the criteria for moral progress: applying the principle(s) in a wider range of cases is what constitutes progress (compare Macklin, 1977). Of course, as Roth (2012) notes, these accounts are just the epistemic criterion conception in disguise: the account merely assumes the truth of some moral principles. Another instance of the end-point view is the epistemic view discussed by Moody-Adams (1999). She suggests that deepening the grasp of a concept and its application in practice constitutes moral progress. Moody-Adams's account is interesting because she does not restrict *how* our grasp of the concepts ought to be deepened. Haslam's (2016) account of what he calls 'concept creep' would fit her view well: concepts like 'abuse' or 'bullying' are, according to Haslam, now applied to more cases and for, arguably, more minor

transgressions. For example, he discusses how in the early 19th century, the concept RAPE was deemed appropriately applied to a case of sexual abuse only if penetration occurred. Today, the fact that no penetration occurred during a case of sexual abuse is no legitimate reason not to call it a rape. Though Haslam is somewhat critical of some developments related to concept creep, if Moody-Adams is right, then he is describing an instance of moral progress. We are, after all, now realizing the true nature of (thick) normative concepts such as RAPE, and their implications.

The epistemic end-point view entails the control requirement because the moral truths or principles at the heart of such accounts will be truths according to a given substantive, normative theory. They will be truths of utilitarianism or deontology, for example. At the very least, epistemic end-point views require for moral progress that at least one individual makes a correct judgment about the moral truths, unperturbed by morally irrelevant situational factors. Now, there is a question, of course. of whether *one* individual's correct moral judgment would be enough for moral progress to occur or whether a sufficient number of individuals need to make sufficiently many correct judgments over time for moral progress to occur. We lean toward the latter interpretation, but we need not settle that question in this paper. For the purposes of our argument, it is sufficient to note that either option will require at least one individual to make correct moral judgments.[8]

Next, *formal* criteria of moral progress belong to the family of end-point views. Consider Singer's (2011) account of the expanding circle: moral progress is taken to be constituted by expanding morality's scope, authority, and domain. The formal end point view relies on a normative theory to justify its formal criteria. Singer, for example, argues that utilitarianism is true and that, therefore, we morally ought to extend our realm of consideration to non-human animals. Similarly, Nagel has argued that moral progress occurs when objective, impersonal, agent-neutral reasons replace subjective, personal, or agent-relative reasons for action (Nagel, 1986, ch. 9).

Formal criteria also entail the control requirement: the formal criteria need to be met in practice by individuals making moral judgments. They must, for example, judge that they have as much reason to help those in far-away countries as they have to help their close friends. To arrive at that judgment, they must not be perturbed by morally irrelevant situational factors, such as in-group favoritism. Likewise, Nagel's view requires people to be sensitive in their judgments to the right kinds of reasons. This, too, implies that people's judgments must be controlled.[9]

Finally, *substantive* end-point criteria for moral progress are evaluative bases that inform us how far we have come in terms of progress. A recent example is the objective list theory of Jamieson (2017). These theories are premised on the idea that there are some *actions* or *states of affairs* that are

positively evaluated by all moral theories. This is a less ambitious version of the idea that showing the consistency of all major normative theories (Parfit 2011).[10] In any case, substantive views quite obviously entail a commitment to substantive normative theories and hence to the control requirement: people must judge that the proposed substantive criteria are, say, good and in doing so they must be unperturbed by morally irrelevant situational factors.

Hence, all three major variants of the end-point view entail a commitment to the control requirement: in order to achieve moral progress in the sense described above, our moral judgments must be controlled so that they are not influenced by situational factors. The fact that moral judgments are pervasively influenced by situational factors is thus, insofar as end-point views are true, an obstacle to moral progress.

3.2. Functionalist views

The second relevant account of moral progress is functionalist. On such an account, morality is considered to have a function (indeed, that function is constitutive of morality). Functionalist views first identify what they take to be the constitutive function of morality and then define moral progress as increased efficacy in the performance of that function (cf. Buchanan & Powell, 2018).

For example, Kitcher (2011) argues that morality functions as a social technology for resolving altruism failures. Allowing the resolution of altruism failure is what morality *is*. Kitcher concedes that "in the course of progress, the background itself changes, generating new functions for ethics to serve, and hence new modes of functional refinement". In passages like these, it might seem as if there is no substantive definition of *what* function morality must fulfil, but rather morality is defined as fulfilling *some* function. Similarly, Roth (2012) proposes to measure progress by asking whether the initial problem one started with has diminished or disappeared and whether the number of problems has decreased. Railton (1986) can also be considered as a defense of a functionalist view of moral progress.

There are two ways of interpreting functionalist accounts: as *constitutive* or *normative* accounts. Consider the latter first. In that case, we run into a problem that is akin to the "open question argument". It always makes sense to ask whether we *ought* to fulfil whatever function morality has. As Buchanan and Powell (2018) note, we do and should care *how* altruism failures are resolved: does resolving altruism failures involve coercion? Should we accept some degree of altruism failure so that we do not have to use coercion? On a normative interpretation, functionalist views advance a substantive normative thesis about what we ought to do, in which case they entail the control requirement.

The constitutive interpretation of functionalism does not, on the face of it, entail the control requirement. That is because morality might fulfil its function without any conscious effort, or exercise of human "moral consciousness". To take an extreme case, consider Kitcher's proposal that morality's function is the resolution of altruism failures. Fulfilling this function does not require us to make correct, controlled moral judgments. The annihilation of humanity, for example, would ensure that there are no altruism failures anymore and thus there would not be any problems left for morality to solve. In some sense, morality would have fulfilled its function, though in the wrong way. There are two options now for proponents of a constitutive interpretation of functionalism: accept or deny that there are constraints on fulfilling morality's function. If they deny that there are constraints, then their view becomes susceptible to the same problems that make perspective-dependent views unattractive: anything goes. If they accept that there are constraints to fulfilling morality's function, then they are committed to a substantive normative theory to vindicate whatever constraints they envision. For example, Kitcher would probably claim that annihilating humanity is not an acceptable remedy to altruism failures; but his claim will have to be grounded in an appeal to human dignity, the value of life, or some such substantive normative view. That substantive normative commitment both motivates their view and turns out to be relevant for making moral judgments: if life is valuable, then we have to pick up on that in making moral judgments. Hence, even constitutive interpretations of functionalism are committed to the control requirement.

4. The pervasive influence of irrelevant influences on moral judgments

We have sketched how both end-point views and functionalist views of moral progress are committed to the control requirement for moral progress. Accordingly, full moral progress requires people to control for the influence of morally irrelevant factors on their moral judgment (since progress in our moral beliefs is a part of full moral progress). In this section, we will show that the control requirement is frequently and pervasively violated, creating yet another, hitherto underacknowledged, obstacle for full moral progress.

The last two decades have produced a tremendous amount of empirical research on the processes underlying moral judgments. This section presents and interprets the findings of a recent systematic review of the evidence for the influence of situational factors on moral judgments for the problem of cognitive control and moral progress.

Before we dive into the empirical evidence, it may be helpful to outline how the empirical evidence will bear on the theoretical arguments we discussed above, and what consequences we expect for normative ethics.[11]

Our theoretical considerations above suggest that full moral progress depends on people's ability to ward off the influence of irrelevant factors on their moral judgments. In short, they need to satisfy the control requirement. Our empirical considerations will suggest that the control requirement is frequently violated and so we uncover a novel obstacle for the viability of moral progress. For normative ethics, this implies not only that preconceptions of reliable moral judges may have to be reconsidered but also that we might have to reconsider applicable notions of moral progress, and the aim of normative theorizing in general: If normative theorizing aims at what we *ought to think* (cf. G. A. Cohen, 2003, p. 243) and yet we find that most people *fail to think what they ought to think even if they try*, then our findings may raise questions about that conception of normative theorizing. For moral psychology, our argument implies that finding ways to enhance control of moral judgment gain wider importance in the context of the desirability of and threat to moral progress.

With our argument outlined and situated, we can now turn to the empirical evidence. The standard way of testing alternative descriptive theories of moral judgment is by asking subjects to evaluate (amongst others) trolley dilemmas, which pit one moral principle against another. Consider the standard 'switch' trolley scenario that Greene et al. (2001) adopted from Foot (1967) and Thompson (1985)[12]:

Switch: You are at the wheel of a runaway trolley quickly approaching a fork in the tracks. On the tracks extending to the left is a group of five railway workmen. On the tracks extending to the right is a single railway workman. If you do nothing the trolley will proceed to the left, causing the deaths of the five workmen. The only way to avoid the deaths of these workmen is to hit a switch on your dashboard that will cause the trolley to proceed to the right, causing the death of the single workman. Is it appropriate for you to hit the switch in order to avoid the deaths of the five workmen?

In the switch dilemma, the utilitarian response is to pull the switch since doing so would save a greater number of people. The classification of this response as utilitarian is apt, though it should not obscure that it represents a rather rudimentary version of utilitarianism.[13]

In contrast, the following standard 'push' dilemma, adopted by Greene et al. (2001) from Foot (1967), and Thompson (1985), was explicitly designed so that the utilitarian response would conflict with the intuitions of many people:

Push: A runaway trolley is heading down the tracks toward five workmen who will be killed if the trolley proceeds on its present course. You are on a footbridge over the tracks, in between the approaching trolley and the five workmen. Next to you on this footbridge is a stranger who happens to be

very large. The only way to save the lives of the five workmen is to push this stranger off the bridge and onto the tracks below where his large body will stop the trolley. The stranger will die if you do this, but the five workmen will be saved. Is it appropriate for you to push the stranger on to the tracks in order to save the five workmen?

Again, deciding to act in 'push' is considered a utilitarian response since it would result in more life's saved. Deciding not to act, in contrast, can be classified as a deontological response. Trolley dilemmas, and comparable cases of harm-based moral dilemmas are 'sacrificial' dilemmas because they involve sacrificing (at least) one person to save a higher number (Kahane, 2015). The most common finding in the literature on situational influences is that the majority of people approve of intervening in Switch but disapprove of intervening in Push; that is, they "think" like utilitarians in Switch but like deontologists in Push (cf. Klenk, 2021).

Sacrificial dilemmas are used to test what we might call the *Standard Assumption*, namely that moral judgments are predominantly driven by morally relevant factors. Both cases, and the innumerable variations of them are designed to pit two competing normative theories, utilitarianism and deontology, against one another. The original philosophical use of trolley dilemmas (which is often obscured in psychological studies) was to test the respective normative theories against considered intuitions (that is, whether the action recommended by, say, utilitarianism in either case intuitively seems to be the right or most reasonable thing to do). The psychological use was to test quite a different thing: the cases are used to test whether moral theories drive people's moral judgments and, if so, which moral theory. For example, it is assumed that a person with utilitarian tendencies would make moral judgments in accordance with utilitarian moral theory. The standard assumption is not often made explicit, but it explains well the practice of testing normative theories of morality against intuitions. Normative theories, on this view, are implicit in normative concepts and in testing normative theories against our intuitions, we come to appreciate the moral theories we all hold implicitly. Of course, it is not required for the view that people *always* judge in accordance with the normative theory toward which they lean.

As the evidence shows, people's moral judgments are ill-explained by the implicit operation of any moral theory.[14] The average person would seem to hold inconsistent normative theories or no normative theory at all. In either case, the standard assumption fails. But if the standard assumption fails, then why do people tend to make utilitarian moral judgments in Switch and deontological moral judgments in Push? Numerous studies have uncovered a number of personal (e.g., Bartels & Pizarro, 2011) as well as situational

factors (cf. Klenk, 2021) that correlate with utilitarian and deontological moral judgments. What is a situational factor, however?

4.1. Irrelevant situational factors

Switch and *Push* are different – the action type in the former case, to name just one example, is switching a lever in the former, while it is pushing a person in the latter. Should we count features of the case such as 'action type' as situational factors? Alternatively, must situational factors be independent of the case description? This paper will make use of the following definition of a situational factor (Klenk, 2021, p. 2):

> **Situational factor**: Factor f is a situational factor in case c with answer options a1 (classified as utilitarian option), a2 (classified as deontological option) and a_n (for some other moral theory) if and only if f does not legitimately affect the respective normative classification of a1 . . . a_n vis-à-vis the normative theory in question and f is not a dispositional factor of the agent.

Hence, 'action types' are situational factors, as are stress-levels, or wordings of the case, but anti-social disorders or intelligence on part of the experimental subjects are not. The reason is that the latter vary by person, and are thus better classified as personal factors (i. e. individual differences). Classifying a situational factor does not require that we decide on what the correct normative theory is, e.g., deontology or utilitarianism, because, as the next section shows, there are situational factors which count as irrelevant on either theory.

Some of the effects of certain situational factors are well explained by dual-process theories of moral judgment, which predict that utilitarian judgments are typically the products of controlled cognitive processes that override automatic emotional responses (cf. Conway et al., 2018). For example, presenting cases in a foreign language increases the frequency of utilitarian responses by stimulating cognitive control (e.g., Corey et al., 2017).

At the same time, researchers have also consistently found effects that are not accounted for by dual-process theory, such as order effects (e.g., Petrinovich & O'Neill, 1996). The typical trolley response pattern does not occur when subjects are first presented with Push and then with Switch (in which case acting in Switch is also judged impermissible). Effects of the order of presentation do not obviously fit with dual process theory: we should not expect that the order of presentation increases affect, nor that it decreases cognitive control, both factors that would play a role on a dual process theory of moral judgment. The lack of a unified theory to explain the impact of situational factors is crucial because in lieu of such a theory, there

is no reason to suppose that those factors are really all morally relevant (Horne & Livengood, 2017).

4.2. Empirical evidence

The influence of situational factors on moral judgments is indeed pervasive. In this section, we analyze the findings of a recent systematic review of the empirical literature on the influence of situational factors on moral judgment in sacrificial dilemmas to support premise 1 in our argument (cf. Klenk, 2021 n.d.).

We envision two audiences of our argument in this section. First, there are some who are skeptical of findings from sacrificial dilemmas to yield valid information about the influence of situational factors on moral judgment (for recent book-length criticisms, see May, 2018; Pölzler, 2018; Sauer, 2018). To those, we offer a partial corroboration of the external validity of findings concerning the pervasive influence of situational factors on moral judgments. The situational factors that influence moral judgments are realistic, and they work in in realistic, and common scenarios, not only in highly contrived trolley cases, as critics sometimes allege. Second, there are some who are already convinced of the pervasive influence of situational factors on moral judgments. To those, we show that the influence of situational factors is less systematic than commonly thought. This makes it potentially harder to reign in and control our moral judgments, which corroborates the problem of cognitive control for moral progress. The best available evidence suggests that the cognitive control problem *for moral progress* is real, which should be of interest to either party.

The 53 articles analyzed in the meta-analysis report a total of 36 different situational factors tested with N = 11,575 subjects; effect sizes range from small to medium effects. Given extant concerns about the reliability and validity of the findings of such experiments (cf. Appiah, 2009), the reviewed literature strongly suggests that the influence of situational factors on moral judgments in both replicated and slightly amended experimental settings is considerable, which is already an important finding given the recent replication crisis in psychology.

A striking feature revealed by the meta-analysis the wide variety of different situational factors that were found to have an influence on moral judgments. The situational factors fall into two sets, based on the independent variable(s) directly manipulated in the respective study. Factors pertaining to the presentation of the dilemma altered features of the situation, such as the order and language of presentation, the time allowed for subjects to make a decision, or the severity of the harm described in the scenario. Factors pertaining to the subject that made a choice in the dilemma manipulated the subject directly. Examples are incidental serotonin and

stress levels and incidental empathy. Factors of both sets had considerable effects on moral judgments (for a detailed discussion of each factor, Klenk, 2021 2021, supplementary materials).

The majority of the situational factors found to influence moral judgment are regular, normal features of everyday life. This finding is important in the light of plausible concerns about the ecological validity of experiments involving sacrificial dilemmas (cf. Pölzler, 2018, pp. 58–60; Sauer, 2018, ch. 6). Many of the situational factors that sway moral judgments plausibly play a role in situations where people have to make moral judgments outside the lab. For example, in institutional settings with tight agendas (e.g., policy decisions), stress and cognitive load plausibly play a role when making moral decisions. Similarly, public debate on morally relevant decisions pertaining to, for example, discussion of immigration policies are often conducted in a manner that plausibly seem to activate situational factors such as action-framing (e.g., when the choice is framed as (passively) *allowing* immigrants to enter vs (actively) *bringing them in*), psychological distance (e.g., when striking visual imagery is used), and relation-to-judge (e.g., whether immigrants are portrayed as socially connected to the decision maker or not). Thus, insofar as these factors play a role in ecologically valid experiments, we have good reason to suspect that moral judgments 'outside the lab' are influenced by those factors in similar fashion (Paulo, 2020; Klenk, 2021 2021, section 5).

Given the available evidence, it is plausible that the situational factors found to exert significant influence on moral judgment frequently play a large role in everyday moral decisions. Though the best known experimental vignettes, which we introduced above, do indeed feature rather outlandish scenarios about runaway trolleys, the majority of studies is done on vignettes that are closer to scenarios of ordinary moral decision making. For example, the typical 'Trolley' effect also occurs when subjects consider dilemmas in realistic scenarios in the context of vaccine policies, animals research, and personal ones, such as thinking through decisions about relationships. This observation further corroborates the ecological validity of experiments with sacrificial dilemmas, and, more importantly, it supports our claim that situational factors pervasively influence moral judgments outside the lab (cf. Klenk, 20212021). For example, whether people are happy (incidental positive affect, social connectedness), prompted to rely on system 2 processing (e.g., cognitive control, time delay before answer, psychological distance, accessibility of information), or how they relate to the people under consideration (relation to judge, self in danger) are variables that are part and parcel of the context of everyday moral decisions.

Finally, the reviewed literature supports the claim that we are dealing with *morally irrelevant* situational factors. Some may doubt this: after all,

many critics of trolleyology (or, rather, critics of attempts to debunk deontological moral judgments based on trolleyological findings) argue that it is often question-begging to simply assume that some situational factor is morally irrelevant in the course of a debunking argument (cf. May, 2018). However, we need not resolve this discussion here because our claim is not that *all* situational influences on moral judgment are morally irrelevant on any normative theory. Our argument only requires that there are *some* morally irrelevant situational influences vis-à-vis utilitarianism and deontology, and this claim is evident from the factors listed in (Klenk, 2021, Table 3). To illustrate, assume, for the sake of argument, that utilitarianism is the correct normative theory. Then all situational factors that decrease utilitarian judgments are potentially detrimental to moral progress. Of course, if deontology were true, the same conclusion could be drawn: many pervasive factors of everyday life that should not affect whether or not we judge an action licensed by the theory as acceptable do in fact affect our judgments.

Therefore, the crucial upshot of the review is the pervasiveness of the influence of situational factors that play a considerable role in everyday moral judgment. People make moral judgments under stress, when they are tired, in happy states, when they feel supported by friends and family; they do so when their judgment concerns their family, friends, or themselves, they consider the consequences of their actions vividly, or not, and they make judgments in foreign languages. Depending on the correct normative theory, some (though not all) of these factors will influence our moral judgments negatively: we do no arrive at the moral judgments that we should arrive at. Importantly, no matter what normative theory we pick, there are at least some detrimental situational influences. This finding provides good reason to think that moral judgments are often not properly cognitively *controlled* in the sense that they are too easily influenced by situational factors.

One might think that knowledge of the pervasive influence of situational factors lessens their potential relevance for moral progress. Especially if the influence of situational factors is systematic, it might easier to control for them. However, the review puts pressure on prominent attempts to explain the influence of situational factors, and it suggests that we are currently lacking a wholly satisfactory theory about the influence of situational factors on moral judgment. The most prominent attempt to explain the influence of situational factors on moral judgment is dual process theory (Greene & Nystrom, 2004). The theory predicts that situational factors that lead to higher affect (e.g., personal force) activate system 1 processing and thus lead to decreased frequency of utilitarian moral judgments. However, the review suggests that this prediction is affected by valence. Negative affect behaves in line with the prediction of dual process theory, but positive affect does not

(2021., 19ff). It is tempting to suggest that this effect may be explained by the activation of system 2 by positive affect, but future investigations have to test this empirically. But this just points to a further problem in our current understanding of the influence of situational factors on moral judgments. The currently available evidence does not distinguish between an inhibition of system 1 and an activation of system 2, and vice versa. However, untangling these causal pathways will be crucial for controlling the influence of irrelevant situational factors to improve cognitive control. So, a novel insight provided by the meta-analysis in relation to the problem of moral progress is that the causal pathways of the pervasive influence of situational factors are, to a significant extent ill-understood. So, even if we are aware of their influence, the questions left open by current explanations of their effect corroborate the cognitive control challenge for moral progress.

Thus, psychology provides us with a bleak picture of our capacity for accurate moral judgment. Granted the ecological validity of experiments involving sacrificial dilemmas, we have good reason to conclude that the moral judgments we make are very often influenced by factors that make no normatively relevant difference. Our judgments about situations with normative, moral significance often fail to identify the normatively relevant factors. The experimental literature on the influence of situational factors on moral judgment suggest that some popular grounds for optimism about cognitively controlled moral judgments are unsubstantiated: There are realistic, mundane situational factors, that have an effect in realistic scenarios. Doing away with these findings by pointing to the bizarre nature of a tiny subset of cases used in these studies is misguided. Insofar as moral progress, per end-point views and functionalist views of moral progress, requires us to control for the influence of such factors, the prospects for making full moral progress are dimmer than previously thought.

At the same time, the significance of these findings should not be overestimated. We are unable to tell, thus far, how these findings carry over from the laboratory into the real world. After all, deciding what one would do in some hypothetical scenario (albeit realistic and familiar) need not translate to what one would actually do. Sacrificial dilemmas undoubtedly represent only a fraction of people's everyday moral lives. There is no guarantee that situational factors play as big a role in other areas of moral judgment, like character evaluations. If it turns out that people judge differently outside the lab and beyond sacrificial dilemmas, then the relevance of the influence of situational factors for moral progress is hampered. At the very least, however, we can debunk the criticism that such studies lack import because people are confronted with *unfamiliar*, and *widely implausible* scenarios, which would arguably hamper their external validity. The available evidence, however, does not give us positive reason for thinking so. Since there is also ample evidence about the lack of control in moral cognition

from studies beyond sacrificial dilemmas, the overall evidence thus far counts in favor of taking the influence of situational factors to be a pervasive real world phenomenon (cf. De Freitas & Johnson, 2018).

One may think that, while psychological findings on moral cognition offer reasons for pessimism, they also offer countervailing reasons for optimism. Recent findings on debiasing strategies, for instance, (Lilienfeld et al., 2009; Sellier et al., 2019), suggest that we can regain control over our moral judgments by training ourselves in the appropriate debiasing techniques. The problem with this suggestion, however, is that successfully engaging in and carrying out these techniques suffers from the very same control problems, albeit at a higher level. We must be motivated to engage in debiasing, become aware that debiasing is needed in a given instance, and do it properly, timely, and to the right extent), all of which are difficult and costly to do. Of course, proponents of moral progress might object that it is *possible* to correct for the influence of situational factors, even though it is costly, strenuous, and perhaps unlikely to happen. Conceding this point is to admit that moral progress is logically possible, after all. However, this concession is largely independent of the question of whether moral progress is possible *given our psychological constraints*. As we discuss in the next paragraph, existing studies of expert judgment aimed at averting the influence of situational factors suggest that doing so is *not* possible (at least not for the moral experts we have today).

In relation to the previous point, it is also important to address the thought that proponents of end-point views of moral progress may have a straightforward answer to the situationist challenge to moral progress. After all, the meta-analysis findings apply on the aggregate level, whereas end-point views may only require that *at least one* individual makes correct moral judgments. Ultimately, however, we doubt that this objection is compelling because it provides us with insufficient reason to reject the significance of situational influences as an obstacle for moral progress. The objection drives on the possibility that there are at least some individuals whose moral judgments are not affected by situational factors. But there is evidence that even experts in moral reasoning, like professional moral philosophers, are subject to the same situational factors as non-experts (Schwitzgebel & Cushman, 2012), and that it persists despite training and conscious efforts to avoid such influenced (Schwitzgebel & Cushman, 2015). And even if experts are subject to *different* situational factors than non-experts, there is reason to doubt that this implies that their judgments are more accurate, which is what matters for current objection (Tobia et al., 2013). Of course, one might reasonably disagree that moral philosophers have expert status when it comes to morality (perhaps because one is skeptical about the notion of moral experts in the first place). That perspective supports our point. If it is unclear that there are other groups or individuals whom we can reasonably expect to be free from the influence of situational factors (whether that is

because of their expert status, or for some other reason), then we have reason to believe that situational factors pose a serious obstacle for moral progress. Hence, the influence of situational factors makes trouble even for undemanding interpretations of end-point views of moral progress.

5. Conclusion

Empirical research has documented that morally irrelevant situational factors pervasively influence moral judgments. The essay has shown that the currently dominant accounts of moral progress entail that this situation hinders moral progress.

This finding is significant for two main reasons. First, the finding is based on a pervasive, empirically confirmed fact about moral judgments. Insofar as the argument of this paper is correct, the paper prescribes a clear way forward to making moral progress: by gaining more control over our moral judgments. How this is to be accomplished is an exciting question in particular with a view to the (detrimental) influence of situational factors 'one level up' when we think about debiasing and control strategies. This question should be addressed with a view to education, public discourse, and psychology.

Second, the paper suggests a further implication that might give rise to an interesting line of research. The relevance of situational influences on our moral judgments is often taken to undermine the reliability of moral judgments. We have no reason to suppose, the argument goes, that our moral judgments are capable of tracking the moral truth (in whatever form), given the pervasive influence of morally irrelevant situational factors. Insofar as some of the most prominent accounts of moral progress may be shown to rely on a normative theory, the psychological challenge might imply that our epistemic access to judging whether moral progress occurred is less than optimal.

Notes

1. Anderson (2014); E.g., Appiah (2009).
2. The claim is that making better moral judgments at least means to achieve what Moody-Adams has dubbed 'local moral progress.' Whether it also constitutes global moral progress, which indicates progress across (virtually) all subdomains of moral concerns and aspects of moral functioning, would require a more comprehensive treatment of the notion of moral progress that cannot be delivered in this paper (Moody-Adams, 1999, p. 169).
3. Recent forms of populism in public discourse seem aptly characterized by the view of moral progress at issue in this paper. The quality of public discourse starts to erode when people increasingly respond to illegitimate influences that sway their judgments instead of forming their beliefs based on processes that track the truth. For example,

when politicians tout punitive tariffs, voters' positive evaluation of the proposal is often guided by nationalist sentiments rather than proper economic analysis Caplan (2008). Similarly, when people consider immigration policies, their judgment is easily clouded by fear or group affiliation Achen and Bartels (2016). The problem here is not emotional influence per se; indeed emotions are required for sound judgment in very many cases cf. Roeser (2014), but that factors that have nothing to do with the truth cloud people's judgments.

4. Do we see things differently, i.e., do we not see the irrelevant factors anymore, or are we just able to control for situational factors better? Both views are possible, and both are compatible with the view insofar as the moral judgment is ultimately correct. As will become clear in section 2, the empirical evidence suggests that people will have to control for situational factors, hence I will be talking of the control requirement.
5. Looking at the process-related conception of moral progress, we can distinguish two broad families of views: *endpoint* views and *functionalist* views. There are perspective-dependent criteria, too, of course, but here we will assume that they cannot deliver a viable criterion of moral progress. This is a controversial assumption, but one that seems warranted in the present context. To see why, consider a prototypical perspective-dependent criterion of moral progress: state A represents progress over state B insofar as A is preferred by some individual, or group, over B. The problem with such views is not (only) that they introduce a raging subjectivism into the notion of moral progress. The decisive problem is that they align progress with preference and the latter need not be fixed. Hence, state A does and does not constitute progress over state B, depending on the judging individual or group, which is not very illuminating. Apart from perspective-dependent views, there are then two main families of views about moral progress: end-point views and functionalist views.
6. Some authors have also mentioned *quantitative* criteria as a separate type of account of moral progress; e.g., Musschenga and Meynen (2017) Examples of quantitative criteria are the number of right actions or the increase of the correlation between moral judgments and moral actions. However, these criteria are only derivatively quantitative: they rely on a substantive criterion about the unit of measurement. We need to know, in other words, *what* right actions are to use quantitative criteria, hence they can be subsumed under substantive criteria.
7. The end-point view of course raises metaethical questions and appears to reduce to a questions about the plausibility of moral realism: are these moral facts compatible with an acceptable metaphysics and epistemology? The problems of end-point views are not, however, a concern of this paper.
8. Our conception of endpoint views is undemanding, for the purposes of this paper, as a concession to proponents of the possibility of moral progress. If only one individual's correct moral judgments would already constitute moral progress, proponents of the possibility of moral progress have an easier case to prove compared to the (more plausible, but more demanding) requirement that a sufficient number of individuals must repeatedly and reliably make correct moral judgments across time. We return to this point in section 4 below.
9. To forestall the objection that agent-neutral reasons do not ultimately depend on a specific normative theory, consider that agent-neutral reasons exist only if you believe that the choice between saving your daughter or saving a bishop from a burning house just depends on the consequences of either act and that your personal relation with your daughter plays no role on this conception. Hence, Nagel is committed to a substantive normative theory, too.

10. There is a special problem for substantive views, of course, that has to do with incommensurability. The objective list views of moral progress that are exemplary of these kinds of views might indeed include values or aims that are positively evaluated by all moral theories; the question, of course, is whether all values can be realized to the fullest at the same time and what to do if value conflicts arise. Given the ongoing discussion of ethical dilemmas, it may be that proponents of substantive views have to settle for local views of moral progress since there will be incommensurability and dilemmatic trade-offs at the global level.
11. Thanks to an anonymous referee for prompting us to clarify these points.
12. The dilemma can be found in the supplementary material of Greene et al. (2001), available at http://science.sciencemag.org/content/suppl/2001/09/13/293.5537.2105.DC1.
13. Greene emphasizes that these judgments are merely *characteristically* utilitarian; Greene (2014).
14. Of course, there could be a theory that explains why you should act in the switch case but not in the push case, but there is not.

Acknowledgments

We are grateful to two anonymous reviewers of this journal for constructive feedback on this paper. We also thank the audience at the 2020 Descartes Lecture at Tilburg University, and Clemens Stachl, for helpful discussion.

Disclosure statement

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

Funding

Michael Klenk's work on this paper has been part of the project 'Design for Changing Values' that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No 788321. Hanno Sauer's work has been part of the project 'The Enemy of the Good' that has received funding from the the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, grant agreement No 851043.

ORCID

Michael Klenk http://orcid.org/0000-0002-1483-0799

References

Achen, C. H., & Bartels, L. M. (2016). *Democracy for realists: Why elections do not produce responsive government. Princeton studies in political behavior*. Princeton University Press.

Anderson, E. 2014. *Social movements, experiments in living, and moral progress: Case studies from Britain's abolition of slavery*. University of Kansas. The Lindley Lecture.

Appiah, A. (2009). *Experiments in ethics* . Harvard University Press. .

Bartels, D. M., & Pizarro, D. A. (2011). The mismeasure of morals: Antisocial personality traits predict utilitarian responses to moral dilemmas. *Cognition*, *121*(1), 154–161. https://doi.org/10.1016/j.cognition.2011.05.010

Bloom, P. (2016). *Against empathy: The case for rational compassion* . Ecco.

Boyd, R. (1988). How to be a moral realist. In G. Sayre-macCord (Ed.), *Essays on moral realism* (pp. 307–356). Cornell University Press.

Buchanan, A. E., & Powell, R. (2016). Toward a naturalistic theory of moral progress. *Ethics*, *126*(4), 983–1014. https://doi.org/10.1086/686003

Buchanan, A. E., & Powell, R. (2018). *The evolution of moral progress: A biocultural theory*. Oxford University Press. https://doi.org/10.1093/oso/9780190868413.001.0001

Caplan, B. D. (2008). *The myth of the rational voter: Why democracies choose bad policies*. . Princeton University. Press.

Cohen, G. A. (2003). Facts and principles. *Philosophy & Public Affairs*, *31*(3), 211–245. https://doi.org/10.1111/j.1088-4963.2003.00211.x

Cohen, J. The arc of the moral universe. (1997). *Philosophy & Public Affairs*, *26*(2), 91–134. https://doi.org/10.1111/j.1088-4963.1997.tb00078.x

Conway, P., Goldstein-Greenwood, J., Polacek, D., & Greene, J. D. (2018). Sacrificial utilitarian judgments do reflect concern for the greater good: Clarification via process dissociation and the judgments of philosophers. *Cognition*, *179*(1), 241–265. https://doi.org/10.1016/j.cognition.2018.04.018

Corey, J. D., Hayakawa, S., Foucart, A., Aparici, M., Botella, J., Costa, A., & Keysar, B. (2017). Our moral choices are foreign to us. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1109–1128. https://doi.org/10.1037/xlm0000356

De Freitas, J., & Johnson, S. G. (2018). Optimality bias in moral judgment. *Journal of Experimental Social Psychology*, *79*, 149–163. https://doi.org/10.1016/j.jesp.2018.07.011

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, *5*, 5–15.

Greene, J. D. (2008). The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.), *A bradford book: Vol. 3. The neuroscience of morality: Emotion, brain disorders, and development* (pp. 35–79). MIT Press.

Greene, J. D. (2014). The cognitive neuroscience of moral judgement and decision making. In M. S. Gazzaniga, G. R. Mangun, & S.-J. Blakemore (Eds.), *The cognitive neurosciences* (pp. 1013–1023). MIT Press.

Greene, J. D., & Nystrom, L. E., Engell, A. D., Darley, J. M, & Cohen, J. D. The neural bases of cognitive conflict and control in moral judgment. (2004). *Neuron*, *44*(2), 389–400. https://doi.org/10.1016/j.neuron.2004.09.027

Greene, J. D., Sommerville, R. B., & Nystrom, L. E., Darley, J. M, & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science*, *293*(5537), 2105–2108. https://doi.org/10.1126/science.1062872

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814–834. https://doi.org/10.1037/0033-295X.108.4.814

Horne, Z., & Livengood, J. (2017). Ordering effects, updating effects, and the specter of global skepticism. *Synthese*, *194*(4), 1189–1218. https://doi.org/10.1007/s11229-015-0985-9

Jamieson, D. (2002). Is there progress in morality? *Utilitas*, *14*(3), 318–338. https://doi.org/10.1017/S0953820800003630

Jamieson, D. (2017). Slavery, carbon, and moral progress. *Ethical Theory and Moral Practice*, *20*(1), 169–183. https://doi.org/10.1007/s10677-016-9746-1

Kahane, G. (2015). Sidetracked by trolleys: Why sacrificial moral dilemmas tell us little (or nothing) about utilitarian judgment. *Social Neuroscience*, *10*(5), 551–560. https://doi.org/10.1080/17470919.2015.1023400

Kelly, D., & Hoburg, P. (2017). A tale of two processes: On Joseph Henrich's the secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter. *Philosophical Psychology*, *30*(6), 832–848. https://doi.org/10.1080/09515089.2017.1299857

Kitcher, P. (2011). *The ethical project*. Harvard University Press.

Klenk, M. (2021). The influence of situational factors in sacrificial dilemmas on utilitarian moral judgments: A literature review and meta-analysis. Review of Philosophy and Psychology. https://link.springer.com/article/10.1007/s13164-021-00547-4 https://doi.org/10.2139/ssrn.3501289

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *4*(4), 390–398. https://doi.org/10.1111/j.1745-6924.2009.01144.x

Macklin, R. (1977). Moral progress. *Ethics*, *87*(4), 370–382. https://doi.org/10.1086/292049

May, J. (2018). *Regard for reason in the moral mind*. Oxford University Press.

Moody-Adams, M. M. (1999). The idea of moral progress. *Metaphilosophy*, *30*(3), 168–185. https://doi.org/10.1111/1467-9973.00120

Musschenga, A. W., & Meynen, G. (2017). Moral progress: An introduction. *Ethical Theory and Moral Practice*, *20*(1), 3–15. https://doi.org/10.1007/s10677-017-9782-5

Nagel, T. (1986). *The view from nowhere*. Oxford University Press. http://www.loc.gov/catdir/enhancements/fy0638/85031002-d.html

Paulo, N. (2020). Moral intuitions between higher-order evidence and wishful thinking. In M. Klenk (Ed.), *Higher-order evidence and moral epistemology*. Routledge.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology and Sociobiology*, *17*(3), 145–171. https://doi.org/10.1016/0162-3095(96)00041-6

Pölzler, T. (2018). *Moral reality and the empirical sciences*. Routledge. https://ebookcentral.proquest.com/lib/gbv/detail.action?docID=5391363

Roeser, S. (Ed.). (2014). Emotion and value. Oxford University Press

Roth, A. (2012). Ethical progress as problem-resolving. *Journal of Political Philosophy*, *20*(4), 384–406. https://doi.org/10.1111/j.1467-9760.2011.00400.x

Sauer, H. (2018). *Debunking arguments in ethics*. Cambridge University Press. https://doi.org/10.1017/9781108529181

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, *27*(2), 135–153. https://doi.org/10.1111/j.1468-0017.2012.01438.x

Schwitzgebel, E., & Cushman, F. (2015). Philosophers' biased judgments persist despite training, expertise and reflection. *Cognition*, *141*, 127–137. https://doi.org/10.1016/j.cognition.2015.04.015

Sellier, A. L., Scopelliti, I., & Morewedge, C. K. (2019). Debiasing training improves decision making in the field. *Psychological Science*. *30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429

Shermer, M. (2016). *The moral arc: How science makes us better people* . St. Martin's Griffin.
Singer, P. (2011). The expanding circle: Ethics, evolution, and moral progress. Princeton University Press
Tam, A. (2019). Why moral reasoning is insufficient for moral progress. *Journal of Political Philosophy*, 1–24. https://doi.org/10.1111/jopp.12187
Thompson, J. J. (1985). The trolley problem. *Yale Law Journal*, *94*(6), 1395–1415. https://doi.org/10.2307/796133
Tobia, K., Buckwalter, W., & Stich, S. (2013). Moral intuitions: Are philosophers experts? *Philosophical Psychology*, *26*(5), 629–638. https://doi.org/10.1080/09515089.2012.696327
Wilson, C. (2010). Moral progress without moral realism. *Philosophical Papers*, *39*(1), 97–116. https://doi.org/10.1080/05568641003669508
Wilson, T. (2004). *Strangers to ourselves: Discovering the adaptive unconscious*. Belknap.