T

Sarah Songhorian, Francesca Guma, Federico Bina,
Massimo Reichlin[1]

Moral Progress: *Just* a Matter of Behavior?

1. *Moral improvement as a requisite of moral progress*

One of the most relevant challenges, for empirically informed ethics, is to understand whether and how moral progress is feasible, given human beings' natural equipment (Klenk & Sauer 2021; Buchanan & Powell 2018). To answer this question, we need to clarify what is meant by "moral progress" and to suggest how it can be measured. The recent discussion on this topic mainly identified moral progress with the institution of collective moral practices which are considered better ones in virtue of their outcomes (Sauer 2019; Sauer et al. 2021); for example, because they better promote the well-being of people and/or sentient individuals. According to this approach, to see whether any moral progress has come about, we need to consider the outcomes that those practices and people's observable behavior actually produce.

However, if on the one hand moral progress refers to changes in collective behavior, on the other we believe that it must encompass individuals' moral improvement as well. We suggest the existence of a bijective relationship between a society's moral progress and the moral improvement of the individuals who are part of it. More precisely, we suggest a) that any progress in collective institutions and practices requires the active contribution of some individuals who have developed a sensitivi-

[1] Faculty of Philosophy, Vita-Salute San Raffaele University, Milan. Corresponding author: songhorian.sarah@unisr.it. Although all authors have collaborated in discussing and revising all parts of the paper, MR is mainly responsible for writing § 1, FG for § 2, SS for § 3 and § 5, and FB for § 4.

 DOI: 10.4454/teoria.v42i2.164

ty for the values at stake; b) moreover, that moral progress is inherently unstable unless enough individuals with better moral capacities promote and strengthen the new ideals and values through their beliefs and behaviors. These claims are partly in line with Buchanan and Powell's interest in highlighting the links between individual and socio-institutional moral change. However, despite touching on potentially interesting implications for an account of individual moral progress, Buchanan and Powell basically refer to morality and moral progress as social phenomena, not merely as individual ones; and they are mostly interested in individual changes in moral beliefs and attitudes «only insofar as these occur in sufficiently large numbers of people to effect social change» (Buchanan & Powell 2018, p. 47).

On the contrary, we want to single out the importance of the moral improvement of individuals who become sensitive to certain values. We agree that it is only with the spread of the new beliefs and values in a sufficient number of individuals that societal moral progress – i.e., a progressive change in common sense morality – is realized; and that such a progress is a factor in determining progressive changes in laws, or other established social practices and formal institutions. There are, of course, complex relationships among the three levels. And yet, the improvement of individuals is an important condition of societal moral progress, which in turn contributes to institutional progress. This must not be meant to exclude that institutional moral progress may also be accomplished independently, nor that it usually has feedback effects on both individual and societal moral progress.

If this picture is plausible, then it is reasonable to say that human moral progress depends at least in part on the possibility for individuals to improve their moral capacities, e.g., by reducing the influence of epistemically defectives biases and other distorting influences. Based on empirical research, some believe that the pervasiveness of morally irrelevant influences on moral judgments prevents moral progress (see Klenk & Sauer 2021, pp. 947-956). Quite the opposite, we believe that moral progress is possible – among other things – by enhancing our capacities to consciously control our moral judgments. Improving the capacity to produce consistent, accurate, and informed moral judgments may make individual improvement possible, thereby causing effects also in the social sphere (Campbell & Kumar 2012). While societal moral progress may be accomplished independently from individual moral improvement, improving the individual capacity for moral judgment is a relevant contribution to the promotion and

strengthening of progressive changes in institutional moral practices: better moral agents adopt better moral behaviors that may eventually be institutionalized, and they may actively promote progressive changes in social institutions.

2. *Moral progress beyond behavior*

As anticipated, many have identified moral progress in the outcomes that social structures and institutions produce, with particular reference to people's observable behavior. While this conceptualization might make its measurement easier, we will show that it overlooks relevant aspects of moral progress. To better illustrate our point, let us consider a few interesting experimental works that, although not primarily intended as contributions to the debate on moral progress, have nonetheless implications for it, since they consider changes in behavior in a direction the authors deem positive.

Schwitzgebel, Cokelet, and Singer (2020) have empirically tested whether ethics classes positively influence students' moral behavior, especially by looking at their meat consumption before and after an educational intervention. Since students attending ethics classes increased their vegetarian choices (as compared to a control group), they can be conceived as having improved, since reducing meat consumption is deemed a morally positive change by the authors. This study was extended and replicated, confirming that it is possible to influence students' attitudes and daily behavior through standard methods of university-level philosophy instruction (Schwitzgebel, Cokelet & Singer 2021).

In both cases, two factors are crucial for a moral improvement to occur. On the one hand, a change in the subjects' behavior or opinion is essential: individual moral improvement is, thus, substantive – i.e., it focuses on the observable behavior or on the content of normative judgments. On the other, the change has a precise direction decided from the beginning by the researchers. The goal is already clear in the title of the first paper: the authors want to study whether it is possible to influence students' behavior in a direction that is assumed beforehand as positive.

These studies have not been explicitly meant as a contribution to either define or promote moral progress, but simply as a test of the ability of ethics lessons to influence students' behavior. However, it is our opinion that, when placed in the context of identifying ways to stimulate moral progress,

this approach is exposed to several criticisms. First, such interventions can result in forms of indoctrination. Given that the experimenters aim at obtaining a certain response, can one really consider it an authentic moral improvement of the subject? If only the outputs are observed, it seems difficult to evaluate whether a change in moral opinion and/or behavior is the result of the acritical (and perhaps not fully conscious) assumption of an external point of view or the effect of a new personal way of thinking about the matter. Moreover, can these modifications be considered as stable improvements? The follow-up conducted in 2021 shows a fair amount of stability. However, if such results are the product of suggestion or indoctrination, the question remains not only whether the subjects would be able to personally formulate good reasons in favor of their new opinion, or whether they would merely repeat remarks that impressed them, but also how long these changes can last. Finally, is it possible to identify what produced the change and how it occurred? Focusing exclusively on the substantive component does not allow one to understand how the changes occurred. Considering only behavior prevents us from ascertaining whether these results are the effect of an increase in the individual capacity for moral judgment, a different personal way of judging, or just the effect of indoctrination. Influencing subjects to produce a particular behavior does not help illuminating the real factors producing that change.

In their second study, Schwitzgebel, Cokelet, and Singer refine the experiment to test whether behavioral change is caused mostly by elements of the instruction (e.g., by introducing non-vegetarian professors and by allowing only half of the students to watch a vegetarianism advocacy video). These modifications could decrease the likelihood of students' suggestibility. However, since these studies do not focus on the reasoning and justification processes that lead subjects to accept certain judgments, such considerations remain only hypotheses.

Focusing on a substantive component of moral improvement seems unsatisfactory. In particular, this approach ignores that a change in behavior or moral judgment, while being an indicator of moral improvement, should not be considered the only possible one, nor the most suggestive. For these reasons, and since we believe that philosophy is one of the means – among others – to achieve moral improvement, in this paper we suggest an alternative approach to the issue.

3. *A procedural moral improvement*

Given the difficulties of an account focused uniquely on behavior, we argue that moral improvement should be considered first and foremost as having a procedural rather than a substantive character (Schaefer & Savulescu 2019; Rawls 1951). On this account, we should not look at humans' actual behavior nor at the content of their moral judgments – although both are certainly relevant –, but rather at the abilities and faculties needed to ground them. What is relevant in this perspective is not what individuals do, judge, or believe; but rather the reasons and justifications they can provide in support of their actions, judgments, and beliefs. Thus, our goal is to underline the role of how a moral output is reached rather than simply focus on what that output actually is. Regardless of the content of a given moral output, we agree with Schaefer and Savulescu (2019) who provide a set of features that would make a moral justification a good one – i.e., logical, empirical, and conceptual competence; openness to the revision of one's opinions; sympathetic imagination; and the attempt to reduce one's biases. Once an output has been reached by making the best of these (and possibly other) abilities, then it should count as an improved one as opposed to an outcome that does not involve them at all.

We argue that there are at least two reasons for a procedural account to be preferable. First, it enables one to hold a pluralistic stance «thus avoiding many question-begging moral assumptions» (Schaefer & Savulescu 2019, p. 75). Several moral disputes are, in fact, so controversial that it is problematic to believe that one solution is certainly the true one, that everyone has reasons to accept. Second, a procedural account – especially one that is concerned primarily with how people justify their behaviors, decisions, judgments, and beliefs – is more suited to account for instances in which an individual might have come to a moral conclusion because of external or internal drives that would not count as an appropriate moral justification. Let us now delve a little bit more into these two issues.

As far as the latter is concerned, to say that individual moral improvement only consists in performing – or complying with – practices judged by a third party as "morally better" overlooks the possibility that behavior can be influenced or causally determined by manipulation or indoctrination. Since the latter are hardly considered appropriate sources of moral education, or of individual moral improvement, accounts that focus uniquely on behavior should show how they can be excluded. How can

we distinguish between someone who is getting rid of her biased behavior towards a social group because she has understood that it was grounded on faulty bases so that she now believes it was morally wrong to have that behavior to begin with, from someone else who does exactly the same just because it is fashionable to be seen as open-minded? In this case, the behavior change will certainly be relevant to account for a person's improvement, but it will not be sufficient. Indeed, it is difficult to say whether a change is determined by an effective, stable, and authentic moral improvement by only observing behavior: people could act in a certain way because they are influenced by internal or external stimuli, by their desire to be socially approved rather than by that of deserving approbation (Smith 1759, III.2.32), by morally irrelevant factors rather than by the morally pivotal ones. On the assumption that an authentic moral action involves a strong sense of agency of the subjects, focusing on how judgments are made and on how moral behavior is grounded can reveal a way to increase the agents' real moral capacity and the conscientiousness of their moral responses (Schaefer 2015).

Coming to the first issue, the adoption of a pluralistic stance drives us clearly away from accounts measuring moral change and moral improvement only in terms of their behavioral outputs. Although, as noted in § 2, Schwitzgebel, Cokelet, and Singer (2020, 2021) are not explicitly concerned with moral progress or improvement, their focus on the reduction of meat consumption after studying meat ethics is a concrete example of the substantive view that we deem problematic (especially when applied to more debated or controversial issues). There are, in fact, many contexts of choice where the issues involved are so disputable, and/or where no action is clearly recommended, that believing one particular behavior represents the right way to go means assuming a specific normative outlook, one that might not be universally shared. Is there a clear set of actions that we can universally conceive as the right one when dealing with issues such as, say, the scarcity of health care resources or global poverty? Since the answer is negative, it seems reasonable to focus on how people justify their often-divergent beliefs and behaviors; endowing people with a sensitivity for the reasons at stake and a capacity to respond to them helps reducing moral plurality by excluding those moral stances that do not pass the test of justification. Thus, improving the abilities and faculties that are involved in an appropriate moral justification should be the starting point to promote moral improvement and moral change. Schaefer and Savulescu's list (2019) is an interesting example of how one can and

should proceed, although we do not claim it is the only one nor it is necessarily complete.

By aiming to avoid the imposition of a substantive normative standpoint as the only right or best one, a procedural account lowers the risk of indoctrination, manipulation, and paternalism in the promotion and assessment of moral improvement and aims to track the path to enhancing moral agency. Focusing on individuals' abilities to provide reasons according to logical, empirical, and conceptual competence, openness to the revision of one's opinions, sympathetic imagination, and bias reduction – i.e., the abilities Schaefer and Savulescu focus on – is a good starting point to ascertain whether one is actually improving her moral stance. Thus, while a behavior or a judgment for which the subject can provide (convincing) reasons is certainly better than one for which no justification seems to be available to her, this clearly is not the end of the story nor a solution for every moral dispute. Much is yet needed for a complete account of individual moral improvement to be in place.

To pave the way for it, though, a procedural account like the one we have gestured towards here is required. In § 4, we will consider one of the most challenging objections to such an account: how can we be sure that improving someone's ability to provide reasons for her actions leads to a moral progress and not a regress? How can we be sure that a procedural account of moral justification has the resources to distinguish proper justification (or moral reasoning) from mere post-hoc confabulation (Haidt 2001; Greene 2008)?

4. *Acceptable moral justifications*

As mentioned, by avoiding any substantive commitment, our proposal risks considering an amelioration in the formal ability to rationalize any moral (or immoral) conclusion as a proper instance of moral improvement. In this section, we offer some replies to this concern by suggesting that not every reason-giving account counts as a proper moral justification, and by adding some considerations about the empirical and theoretical assumptions which may ground this worry.

This objection may stem from views sympathetic to Haidt's influential model of moral judgment (Haidt 2001). Drawing on empirical research, Haidt concludes that moral judgment is not the product of conscious reasoning, but the expression of automatic, unconscious, and affectively-laden

"intuitions" shaped by evolutionary, cultural, and social pressures[2]. Within this model, conscious reasoning intervenes only ex post by concocting reasons to support and socially justify fast and automatic reactions: «one feels a quick flash of revulsion […] and knows intuitively that something is wrong. Then, when faced with a social demand for a verbal justification, one becomes a lawyer trying to build a case rather than a judge searching for the truth» (Haidt 2001, p. 182). According to Haidt, the function of moral reasoning is to socially justify intuitive responses, but it has no power in shaping their content ex ante. In this framework, increased proficiency in the ability to provide socially acceptable justifications would just better perform the function of convincing others about the acceptability of conclusions that are essentially insensitive to rational scrutiny and revision.

However, not all justifications are equal. Following Schaefer and Savulescu (2019), we believe that satisfying certain procedural requirements makes certain reasons or justifications more intersubjectively acceptable than others, without committing to any substantive normative or metaethical view[3]. In particular, some justifications can be more consistent, more sensitive to empirical evidence and to others' perspectives and interests, and more open to revision than others. Acceptable moral justifications do not merely confirm one's opinions, intuitions, or feelings by effectively convincing other people about their soundness; they also express the effort of considering a broader spectrum of information, such as non-moral facts, or the interests and preferences of the individuals involved (including the agent's ones).

If, as we believe, Schaefer and Savulescu's criteria are reasonable and sensible, one can discriminate between different levels of reliability or appropriateness of moral justifications, distinguishing between confabulations (and the correlative phenomenon of moral dumbfounding), motivated or confirmatory rationalizations, and appropriate moral justifications.

In light of Haidt's work, a confabulation is the attempt to fabricate justifications for moral conclusions with clear fallacious results (e.g., blatant logical contradictions), pushed by the desire to hold and confirm one's feelings, intuitive judgments, and beliefs, even when put in front of inconsistencies and contrasting rational arguments (Festinger 1957; Kunda

[2] Haidt's idea of "intuition" differs radically from traditional (rationalist) conceptions of the term in the history of ethics.

[3] See §3 and below in this section – as well as Schaefer and Savulescu (2019) – for a more extensive discussion and defense of these criteria.

1990). In Haidt's famous experiments, some subjects try to rustle up support for their intuitive conclusions by offering fallacious and unsatisfactory justifications which, for example, patently clash with relevant information or just restate intuitive conclusions without justifying them at all (Haidt 2001, 2012). Therefore, we can conceive confabulation as a vicious kind of reason-giving, which lacks several features of an acceptable justification (such as empirical and logical consistency and openness to revision).

Rationalization can be conceived, more broadly, as the justification of behavioral outputs by offering reasons in their support "that would have made it rational" (Cushman 2020, p. 183), even if such reasons do not match the actual processes that led to that output. Many rationalizations can be more consistent and sensitive to logical reasoning and evidence than moral confabulation. However, providing reasons in favor of a moral judgment does not guarantee providing acceptable moral reasons because what is rational, e.g., from a self-interested point of view may not be so from a moral point of view. For example, a rationalization may be grounded on an astute selection of data, aimed to make the preferred conclusion plausible, while a proper moral justification does consider more morally relevant factors, such as the interests of other individuals involved. Also, while rationalization does not require critically examining one's own moral preferences, a good moral justification does. Moreover, even though rationalization requires paying attention to possible influences of biases on argumentation, it does not require taking seriously, for instance, the main moral reasons for and against available stances or lines of action. Supporting moral conclusions with acceptable moral justifications does not simply require a generic capacity to provide any kind of reasons in their favor, but to provide a much more specific kind of reason-giving account.

An acceptable moral justification, thus, requires adequately knowing the context of the situation under evaluation, along with one's and others' perspectives. Reasons for and against different conclusions should be balanced in light of available information, showing the attitude to evaluate potential alternatives with an open mind, and being disposed to reconsider one's opinions. The potential influences of biases or prejudices that might affect the evaluation should also be considered. To achieve this goal, it is important to avoid considering one's preferences as the right evaluative standard for the situation at hand, acknowledging and balancing the different interests at stake. Finally, acceptable moral justifications should satisfy standards of logical consistency. Improvement in these capacities would not only enhance the formal ability to justify any

possible moral judgment or behavior – as the objection we are addressing states – because if these requirements are satisfied the spectrum of reasonably acceptable moral conclusions shrinks significantly.

A couple of final remarks are in order. A strength of our view is that it stands even if Haidt's model of moral judgment is plausible. Even if in isolated, specific circumstances of choice explicit moral reasoning intervenes only after quicker psychological responses, improved justificatory abilities would not just better support a-rational outputs, but can be sensitive to independent relevant information. Nonetheless, there are several reasons to reject Haidt's thesis according to which moral reasoning has no causal influence on moral feelings, intuitions, and judgments. Critics have stressed the limits of Haidt's model, denouncing its rigid lack of interaction between controlled and automatic processes, as well as its blindness about the diachronic dimension of moral judgment (Campbell & Kumar 2012; Railton 2014).

Even if it does not come into play immediately before the expression of a moral conclusion at the time of decision, explicit moral reasoning can feedback on, inform, and improve people's future moral responses (Sauer 2017). If this is true, an appropriate moral justification can also reliably point out some of the reasons that informed one's intuitive judgment or behavior (Cushman 2020).

All these are not necessary requirements of motivated (or confirmatory) rationalizations. Therefore, we conclude that acceptable moral justifications can be distinguished from other reason-giving accounts. This allows us to reject the objection accusing our position of considering mere improvements in the capacity to rationalize as proper moral improvements.

5. *Conclusion*

The aim of this paper was to argue in favor of the need – within the empirically informed debate on moral progress – to focus on an individual procedural moral improvement. We have argued that moral improvement is an essential condition for moral progress and that it should be understood not in substantive terms, but rather in procedural ones. A change in behavior or in moral judgment, while being an indicator of moral improvement, should not be considered the only possible one, nor the most indicative.

For this reason, we have gestured towards a procedural account of the abilities required to reason and to justify one's actions and beliefs as the

first necessary step to truly understand the contribution made by individual moral improvement to the debate on moral progress.

Finally, we have considered a challenging objection to our account – i.e., whether the abilities a procedural account proposes to improve allow us to distinguish appropriate moral justifications from mere post-hoc confabulations; we have argued that such a distinction can in fact be drawn and that not any reason-giving account counts as a proper form of moral justification.

References

Buchanan, A., Powell, R. 2018, *The Evolution of Moral Progress: A Biocultural Theory*, Oxford University Press, Oxford.

Campbell, R., Kumar, V. 2012, "Moral Reasoning on The Ground", Ethics, vol. 122, n. 2, pp. 273-312.

Cushman, F. 2020, "Rationalization is Rational", Behavioral and Brain Sciences, vol. 43, pp. 1-16.

Festinger, L. 1957, *A Theory of Cognitive Dissonance*, Stanford University Press, Stanford.

Greene, J.D. 2008. "The Secret Joke of Kant's Soul". In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Vol. 3. The Neuroscience of Morality: Emotion, Brain Disorders, and Development* (pp. 35-80). MIT Press, Cambridge.

Haidt, J. 2001, "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment", Psychological Review, vol. 108, pp. 814-834.

Haidt, J. 2012, *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. Pantheon Books, New York.

Klenk, M., Sauer, H. 2021, "Moral Judgement and Moral Progress: The Problem of Cognitive Control", Philosophical Psychology, vol. 34, n. 7, pp. 938-961.

Kunda, Z. 1990, "The Case for Motivated Reasoning", Psychological Bulletin, vol. 108, n. 3, pp. 480-498.

Railton, P. 2014, "The Affective Dog and Its Rational Tale: Intuition and Attunement", Ethics, vol. 124, n. 4, pp. 813-859.

Rawls, J. 1951, "Outline of a Decision Procedure for Ethics", The Philosophical Review, vol. 60, n. 2, pp. 177-197.

Sauer, H. 2017, *Moral Judgments as Educated Intuitions*. MIT Press, Cambridge (MA).

Sauer, H. 2019, "Butchering Benevolence Moral Progress beyond the Expanding Circle", Ethical Theory and Moral Practice, vol. 22, n. 1, pp. 153-167.

Sauer, H., Blunden, C., Eriksen, C., and Rehren, P. 2021, "Moral progress: Recent developments", Philosophy Compass, 16(10), e12769.

Schaefer, G.O. 2015. "Direct vs. Indirect Moral Enhancement", Kennedy Institute of Ethics Journal, vol. 25, n. 3, pp. 261-289.

Schaefer, G.O., Savulescu, J. 2019, "Procedural Moral Enhancement", Neuroethics, vol. 12, n. 1, pp. 73-84.

Schinkel, A., de Ruyter D.J. 2017, "Individual Moral Development and Moral Progress", Ethical Theory and Moral Practice, vol. 20, n. 1, pp. 121-136.

Schwitzgebel, E., Cokelet, B., and Singer, P. 2020, "Do Ethics Classes Influence Student Behavior? Case Study: Teaching the Ethics of Eating Meat", Cognition, vol. 203, p. 104397.

Schwitzgebel, E., Cokelet, B. and Singer, P. 2021, "Students Eat Less Meat After Studying Meat Ethics", Review of Philosophy and Psychology. https://doi.org/10.1007/s13164-021-00583-0

Smith, A. (1774), *The Theory of Moral Sentiments: An Essay Towards an Analysis of the Principles by which Men Naturally Judge Concerning the Conduct and Character, First of their Neighbours, and Afterwards of Themselves*. The Fourth Edition, Millar, London; Kincaid and Bell, Edinburgh.

Abstract

The aim of this paper is to argue in favor of the need – within the empirically informed debate on moral progress – to focus on an individual procedural moral improvement. We argue that moral improvement is a prerequisite for moral progress and that it should be understood in procedural (rather than substantive) terms.

Thus, we gesture towards a procedural account of the abilities required to reason and to justify one's actions and beliefs as the first necessary step to understanding the contribution individual moral improvement makes to the debate on moral progress.

Finally, we consider a challenging objection to our account – i.e., whether the abilities a procedural account proposes to improve allow us to distinguish appropriate moral justifications from mere post-hoc confabulations – arguing that not any reason-giving account counts as a proper form of moral justification.

Keywords: moral progress; individual moral improvement; procedural improvement.