

Francesca Guma

Becoming Better Moral Agents by Strengthening Free Will. A Possible Prospect?

1. *The limits of judgment*

Several theories and experimental proposals in contemporary ethics show an intent to study various aspects of moral reasoning and a desire to investigate whether and how it is possible to improve and/or enhance the ability to make moral choices (Klenk & Sauer 2021; Songhorian, Guma, Bina & Reichlin 2022). This is a relevant issue, especially as a result of studies in cognitive psychology, neuroscientific research, and reflections on self-control that psychologists, philosophers, and decision theorists have made in recent years (Bermúdez 2018).

Referring primarily to Kahneman's studies (Kahneman 2011), many authors point out that decisions are often made in absence of awareness. Individuals frequently give automatic responses that do not result from thoughtful, informed reflection. This happens not only in tasks considered entirely solvable by resorting to fast and intuitive thought processes but also in those which seem important to appeal to slower but also more logical and reflective processes. Similar remarks have led to more inquiry into the origin – rational or emotional – of moral judgments. In an attempt to integrate empirical knowledge with philosophical knowledge, several vastly different theoretical proposals have arisen in ethics (Greene 2013; Nichols 2004; Sauer 2017). Regardless of the position considered, the presence of automatic and unconscious reasoning, even in the elaboration of moral judgments leads to problematizing the subject's capacity to make conscious ethical choices. Similar remarks can also be made by taking psychological theories other than Kahneman's model as a reference. Considering, for example, a psychodynamic approach that takes the psyche in its

topical, dynamic, and economic relationships into account, the difficulties related to the agent's control and awareness remain evident: a thoughtful, rational, conscious choice is an intricate and complex matter. Stressing that the human will cannot be understood in its totality if confined to the level of consciousness, these theories show that it is easy to find both situations in which an unconscious will can be traced and situations in which conscious psychic acts experienced as willed and free although, in reality, are not (Guma 2021).

For these reasons, regardless of the theoretical frame of reference, it is important to ask how much control of action and choice the agent really has, to what extent it is possible to make conscious and thoughtful moral choices and if so, whether one can call upon greater commitment and concentration to arrive at more accurate reflections. These questions are important because, especially in ethics, the will must depend on us and must be in our power. Generally, agents are exempt from responsibility for effects that are not caused by them and that they cannot avoid. Recognizing the contribution scientific research has offered, it becomes difficult to evaluate an individual who acts driven by internal influences that she does not know and cannot control. This problem emerges, for example, in actions caused by implicit biases, which undermine the subject's agency. Several authors question whether – and how – we can hold ourselves accountable for our implicit biases, or how we should structure society to counterbalance them (Beeghly & Madva 2020, part 3). As we recognize the presence of automatic and unconscious modes of functioning, it becomes crucial to ask how much control we can actually exercise over the choices and moral judgments we make.

2. Free will and moral judgment

Theories developed from the observation of the limits of rationality and conscious control that human beings exercise over their choices, decisions, and actions are closely related to the question of free will. To understand this relationship, it is essential to start from a notion of free will that can dialogue with empirical research. Free will can, thus, be identified with the opportunity and capacity to will otherwise. To do so, it is necessary to give an empirical interpretation to the two conditions deemed necessary to define free will: the existence of alternative possibilities and the agent's conscious control of their will. In this view, freedom of will on the one hand

requires the existence of actual alternative possibilities to be understood in both a negative and a positive sense: to choose and act freely, a subject must suffer neither the effect of exogenous forces that constrain from outside (she must have the actual opportunity to act and will), nor the effect of endogenous forces that limit or prevent her from being able to act otherwise (she must have the actual ability to act and will). On the other hand, to speak of a free choice or action, it becomes essential to ascertain the presence of actual conscious control of the will by the subject, that it is crucial to assess the subject's agency: the individual must subjectively perceive that she is in control of her behavior, must feel endowed with the opportunity and ability to act and will (Magni 2019).

This naturalistic conception of free will reveals the connections between the considerations seen in §1 and freedom of will: as List points out, the subjects' sense of agency is intimately connected to the idea of being able to make decisions independently; individuals want to be their own masters, to achieve their own ends, to act and choose consciously (List 2019); however, empirical findings undermine or even disprove the very agency of the subject (Soon, He, Bode & Haynes 2013). Considering psychic determinism and the related difficulties an individual may face when confronted with her own mental resistances, cognitive deficits and motivational constraints coincide with admitting humans' concrete difficulties to cope with endogenous and exogenous forces that hinder them in effectively choosing freely.

If judgments are believed to be at least in part «a form of measurement in which the instrument is a human mind» (Kahneman, Sibony & Sunstein 2021, p. 361) and the results of experimental research are noted, it becomes important to ask whether and how it is possible to develop strategies that can make moral judgments more competent, more rational and more solid. The possibility of becoming better moral agents is related to the possibility of increasing the effective opportunity and ability to will otherwise. Reflecting on strategies that can improve and/or enhance the capacity to make moral choices can be seen as reflecting on strategies that can improve and/or enhance free will. Assuming that for human beings nothing is truly neutral and that, therefore, depending on the characteristics of the observed objects one is pushed to produce unconscious/automatic inferences that condition the results of reasoning, some proposals developed to improve moral reasoning can be considered attempts born out of reflection on the possibility of expanding the subject's positive freedom, of increasing control over action and choice to make her less exposed to automatisms.

But is it possible to strengthen free will to become better moral agents? If so, how?

In the next two paragraphs, I will present two possible ways to achieve an improvement in individual moral action, showing how, while both acknowledging free will natural and empirical limits, these proposals arrive at extremely different solutions.

3. Nudge and suggestion

Assuming the ineradicable presence of endogenous and exogenous conditions that make it difficult for the agent to control her action and choice, those advocating for nudges and suggestion believe it is impossible to increase the subject's agency. Thus, the only possibility to achieve individual moral improvement, lies in introducing specific exogenous conditions capable of directing individual's actions and choices: to improve people's moral judgments it is appropriate to induce them, through more or less gentle nudges and suggestion, to make objectively better choices. Thaler and Sunstein's proposal (2008) is one of the most famous examples of such an approach. The authors believe choice architects must influence individuals' behavior to improve their lives. Subjects are left free to choose, meaning that the existence of actual alternative possibilities is not affected. However, because they are judged to be fallible in making their own decisions, individuals are prodded through information disclosure, warning, and making rules about default situations. Clearly, from the negative aspect of freedom point of view, there are no restrictions: choices are not prevented, blocked, or made overly burdensome; no constraints or prohibitions are placed, and the subject can choose among possible alternatives. On the positive aspect of freedom, although the authors start with the idea of providing measures that protect or increase freedom of choice (Thaler & Sunstein 2008, p. 5), the proposed interventions are not designed to maintain or increase self-actualization and non-hetero-directed action.

Such an approach offers, at least in the first instance, guidelines that appear attractive also to be applied to the increasing moral capacities context. This theory not only considers the objective and incontrovertible evidence that humans will never be fully rational but also emphasizes that they will always and in any case be conditioned in their decisions. Starting from this acknowledgment, it suggests practical interventions that produce

concrete results, allowing agents to make choices that are deemed objectively better (John, Smith & Stoker 2009; Bhargava & Loewenstein 2015). Extending the theory of nudging to the field of reflection interested in improving and/or enhancing moral capacities is quite straightforward, especially since nudging is also applied in contexts that require ethical choices and reflections. Consider, for example, *Green Nudge*, which was created precisely to encourage individuals to engage in environmentally responsible behavior (Schubert, 2017).

However, applying these conceptions focused on nudges to the enhancement of moral capacities leads to some difficulties. Firstly, approaches based on nudging (or suggestive interventions) are not born to achieve an individual's true improvement. As much as they assume that humans would be better moral agents if they better control their moral reasoning and become more aware of it, they conclude that given the impossibility or difficulty in achieving such improvements, it is essential to develop a way to achieve objectively better, concrete results. To do that: they define (more or less explicitly) a scale of values that they consider preferable to the one that an individual might have; they devise interventions that play on what they have identified as cognitive weaknesses; and finally, they nudge/suggest the individual to make a certain type of judgment. Such interventions cannot achieve effective individual moral improvement because they are not aimed at increasing the subject's capacity. By choosing the path that leverages the agent's weaknesses, they aim only for outward behavior modification. The person undergoing such an intervention is not stimulated to produce an improvement in moral reasoning, but rather to give a particular response. The new way of acting or judging may be better considering the social context, but it cannot be considered an actual improvement of the agent. How much can a change achieved by this route be worth and how long can it last? Being the result of a suggestion, the subject has not acquired any ability, and the effect remains linked to the strength of the input given from outside.

Secondly, they do not seem to safeguard a morally relevant characteristic: the subject's agency. The moral actions we consider authentic are those that involve a strong sense of agency: our judgments of responsibility, praise, and blame are stronger if we can attribute agency to the individual. These approaches inserting external elements with the purpose of piloting judgments and decisions do not recognize the value of agency. The possibilities for moral improvement through the development of interventions that influence behavior, attitudes, dispositions, and motivation raise

important questions of freedom and responsibility that not only affect our sense of who and what we are, but also whether we are, or can remain, creators and masters of our decisions and actions (Harris 2016). Such proposals seem to imply that it is permissible to subjugate the autonomy of an individual who does not appear to behave rationally. Taking up a critique developed by Quong concerning paternalism, it is possible to note that starting from people's psychological deficits to direct their choices does not seem a good way to develop their individual moral capacities, but neither is it a good way to respect them as persons. Underneath these settings lies a «judgemental definition»: agent A attempts to increase agent B's values toward a particular decision or situation that B faces; A's action is motivated by a negative judgment regarding B's capabilities to make the right decision, or to handle the particular situation in such a way that she can effectively increase her own values. In making a negative judgment, A may have considered and examined three different abilities of B: practical reasoning, willpower, and emotional management. In identifying these three capacities as relevant, Quong rules out the possibility of considering actions as paternalistic solely because they aim to make up for physical or informational deficits. The heart of paternalistic action is always in negative judgment. The one who performs the paternalistic act, in each case, believes that she knows better than the other how the latter should act; she is convinced that the other does not possess the necessary level of rationality, willpower, or emotional management to accomplish what is best for her. To treat a person paternalistically is yes to treat her as a child, but in a specific sense: it is an attempt to act in her best interest because it is believed that such a person lacks the ability to do that for herself (Quong 2010). Considering this analysis, it is possible to say that choosing this first way to achieve moral improvement in the individual means judging one's mental abilities negatively, disbelieving her deliberative capacities, deeming her inferior in the faculty of decision-making and/or choice, and arrogating the right to direct her, more or less kindly, to the option that is deemed best.

Thirdly, these proposals risk generating coercive fallout in the social freedom area. Assuming the impossibility of increasing conscious control of the agent's will, the architects of choice, essentially, pose as directors of conscience: it is true that they always leave the possibility of alternative choices, but it is also true that they assume that individuals will be more likely to go toward what is suggested to them. Such a scenario does not seem very different from the one described by Berlin in *Two Concepts of*

Liberty (1969). Reflecting in the political sphere on conceptions that identify the notion of freedom with self-determination, Berlin asserted that it is possible to embark on a slippery slope capable of leading to a conclusion that is at least singular: the total denial of the freedom that, on the contrary, was intended. Following his argument, asserting the existence of an ego split into a rational part of an elevated nature (the true Ego) and an irrational, desiring part of a lower nature (the empirical Ego) can lead to the claim that a person is free if and only if she acts rationally, following her true Ego. Since it is commonly agreed that some people are more rational than others, if one or more people were to convince themselves that they know the true end to which actions should tend, they could construct a good argument for coercing other less rational individuals. The reasoning would be based on the consideration that the desires of the less rational subjects would be equal to those of the more rational ones if they were not distracted by the empirical Ego. From here the step to a totalitarian state would be a short one: leaning on the consideration that the many would desire what they are forced to if they were not at the mercy of their lower natures, such oppression would come to take the form of liberation. This perspective, however exaggerated, does not seem so surreal for positions that favor nudges and suggestions.

In conclusion, I agree with the starting point of these accounts and believe that it is important to consider the exogenous forces that every human is subjected to daily. Indeed, it is interesting to point out that such conceptions highlight that a good choice architecture system could help make information more comprehensible and could allow individuals to refine their capacity to map decisions. Noting that subjects are always inevitably subjected to events and rules that influence their judgments can help develop strategies that, by calculating exogenous forces, stimulate agents to behave more consciously. In this sense, nudges could be thought of as tools aimed at increasing the subject's positive freedom, or as tools that respect the decision-making autonomy of the individual and enhance reflective decision-making (Baldwin 2014). However, these would be different interventions because they are not designed to obtain a specific response or behavior from the agent. For these reasons, I believe it is appropriate to seek an alternative way to achieve an improvement in individual moral action.

4. *Increasing agency*

Always starting from the observation that individuals defect in moral capacities due to a lack of conscious control of will, I argue for the possibility of identifying ways to increase the subject's agency, focusing, for example, on knowing one's own psychic dynamics, augmenting or elucidating information, and providing spaces to reflect on one's logical and argumentation capacities. Acknowledging humans' natural, automatic, and unconscious component does not necessarily lead to the conclusion that the subject has, as an individual, no margin for improvement: the individual certainly has a limited capacity to consciously control her will, but this does not eliminate the possibility of considering her capable of increasing it. If one considers agency a crucial aspect of ethics, then it seems essential to develop interventions that aim to track ways to increase the effective possession of the opportunity and capacity to want otherwise. Indeed, this approach does not have the advantage of achieving definite and obvious concrete results, but it appears preferable for at least two reasons.

Firstly, it aims for effective and stable individual moral improvement because it preserves the interest in identifying ways that can effectively expand the subject's possibilities for conscious choice and decision-making. Accepting that humans would be better moral agents if they could better control their moral reasoning and acquire greater awareness of it, this approach aims to increase the subject's effective possession of the opportunity and capacity for conscious control. The focus is not on the content of moral judgments, but on the ability to develop them, be aware of them, justify them, recognize them, argue for them, and provide good reasons in their defense. Maintaining the goal of increasing agency, one seeks not a change in outward behavior (in the output), but a change in the procedure underlying the capacity for moral reasoning, an increase in the awareness of the reasons for one's moral judgments. Although this approach recognizes the impossibility of generating automatons with perfect morality, it allows effective individual moral improvement. Developing procedures that can strengthen the subject's free will indeed makes it possible to think of genuine and stable moral improvements because there would be changes and improvements not in any specific outward behaviors but in the individual's general moral attitude. This view also avoids possible coercive effects on social freedom: by focusing not on the content of judgments, but rather on how they are made, this approach is not committed to a specific nor-

mative framework, nor does it presume to list what can be judged as good or right, leaving the agent free to work out the judgment she deems most appropriate.

Secondly, keeping the goal of increasing agency, safeguards the morally relevant characteristic that previous positions do not value. As much as the literature on moral responsibility provides different perspectives about what makes a subject responsible for an action, it is fairly common to believe that being morally responsible is deeply connected to what it takes for that action to be an expression of the agent's will. Thinking of ways to increase people's agency, thus seems a good way both to respect and to increase their moral capacities (Reichlin 2017). The project is certainly ambitious, but some studies suggest it is feasible. For example, recent studies show that in some contexts our implicit biases can be changed easily (Beeghly & Madva 2020, part 1), while some authors suggest developing strategies enabling indirect control over such biases, such as through the development of a set of long-term habits or certain social policies (Beeghly & Madva 2020, part 3). In view of this, it is also useful to consider the problem of adaptive preferences highlighted by Elster and Sen: individual preferences are influenced by the social and environmental conditions in which humans are embedded, which is why their choices may often not be the ones they would make if they were more aware of their situation. Sometimes, by increasing information, the individual shows that she acts differently than she would have done by ignoring certain factors. Another interesting proposal, and not far from these considerations, comes from Gigerenzer, who stresses the possibility and importance of educating individuals to make the best possible decisions for themselves. Reconfirming the impossibility of Olympic rationality, Gigerenzer points out that intuitive, quick, and immediate mental processes are useful, often necessary, and capable of leading to optimal decisions if one has the right tools and knowledge to avoid falling victim to bias and to the way information is presented. Indeed, for the author, it is not only heuristics that lead us to erroneous conclusions and limit our evaluative ability, but also poor statistical education. Gigerenzer's proposal can be read from the perspective of developing interventions aimed at increasing subjective agency: the author creates teaching methods that enable even elementary school children to learn how to recognize and solve some Bayesian statistical problems that often underlie bad decisions (Gigerenzer 2008). In this framework, the goal is not to steer the mind, but to empower its cognitive and deliberative tools (Hertwig & Grüne-Yanoff 2017). Strengthening the subject's capacity for

conscious control enables her to consider her moral choices and actions in a more authentic dimension, reflexively increasing judgments of responsibility, praise, and blame. In this sense, interventions that act not only on moral behavior constitute true moral enhancements, because they improve either the individual's capacity for moral insight or his or her ability to weigh the reasons for and against a certain course of action and decide accordingly.

5. Conclusion

Considering a naturalistic conception of free will in relation to a theory of individual moral improvement makes one think about interventions that help individuals increase their opportunity to act, will, and choose consciously when exercising their moral capacities. It also highlights the importance of the agent reinforcing her own power or ability to judge something good or right, without having external judges deciding for her, regardless of her inclinations. There are great differences among people, which is why it does not seem attractive to advocate a position aimed at providing a guide that applies to everyone: such a guide would destroy some of the conditions necessary for freedom and, ultimately, for achieving real individual moral improvement. Moreover, such a position would leave open the question as to who might be entitled to establish what is right or wrong, good or bad.

Making information more comprehensible, educating individuals to recognize their own limitations and mistakes, and helping them improve their ability to map decisions, can be considered some of the ways to increase the positive aspect of the individuals' free will. As this is related to being able to control one's own choices and will, giving information, increasing the opportunity to receive feedback, providing spaces for discussion, and developing *ad hoc* education programs can be interventions aimed at helping the individual make more informed decisions. In this sense, it seems that «the best, more promising methods we have of moral enhancement are [...] traditional ones: education, parental and peer group guidance, social and personal example, and indeed reflection on what's rights, namely ethics» (Harris 2016, p. 117).

Focusing on the agency does not exclude the possibility of educating subjects, even wanting to try to persuade them to make some choices rather than others; however, it leads one to reflect on what to consider moral

enhancers. Viewing moral empowerment as indirect (Schaefer 2015), from a formal, procedural rather than a substantive perspective (Songhorian, Guma, Bina & Reichlin 2022), concerned with the capacities of individuals can lead to defining such empowerment in terms of an enhancement of free will. This would not wish to see situations realized in which individuals would be incapable of doing evil, but rather it would become possible to observe better moral agents because they are more capable of consciously choosing their judgments and actions.

References

- Baldwin, R. 2014, "From regulation to behaviour change: giving nudge the third degree", *The Modern Law Review*, vol. 77, n. 6, pp. 831-857.
- Bhargava, S., Loewenstein G.A. 2015, "Behavioral Economics and Public Policy 102: Beyond Nudging", *American Economic Review, Papers & Proceedings*, vol. 105, n. 5, pp. 396-401.
- Beeghly, E., Madva, A. (Ed.) 2020, *An Introduction to Implicit Bias: Knowledge, Justice, and the Social Mind*, Routledge, New York.
- Bermúdez, J. (Ed.) 2018, *Self-Control, Decision Theory, and Rationality: New Essays*, Cambridge University Press, Cambridge.
- Berlin, I. 1969, *Two Concepts of Liberty*, in I. Berlin, *Four Essays on Liberty*, Oxford University Press, London.
- Gigerenzer, G. 2008, *Rationality for Mortals: How People Cope with Uncertainty*, Oxford University Press, New York.
- Greene, J.D. 2013, *Moral Tribes: Emotion, Reason, and the Gap between Us and Them*, Penguin, New York.
- Guma, F. 2021, "Determinato ma possibilmente libero. La libertà del volere nella teoria psicanalitica", *Metapsychologica – Rivista di psicanalisi freudiana*, vol. 1, pp. 95-124.
- John, P., Smith, G., Stoker, G. 2009, "Nudge Nudge, Think Think: Two Strategies for Changing Civic Behaviour", *The Political Quarterly*, vol. 80, n. 3, pp. 361-370.
- Harris, J. 2016, *How to be Good: The Possibility of Moral Enhancement*, Oxford University Press, New York.
- Hertwig, R., Grüne-Yanoff, T. 2017, "Nudging and Boosting: Steering or Empowering Good Decisions", *Perspectives on Psychological Science*, vol. 12, n. 6, pp. 1-14.
- Kahneman, D. 2011, *Thinking, Fast and Slow*, Penguin Books, London.

- Kahneman, D., Sibony, O., Sunstein, C.R. 2021, *Noise. A Flaw in Human Judgment*, Little Brown Spark, 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.
- List, C. 2019, *Why Free Will Is Real*, Harvard University Press, Cambridge.
- Magni, S.F. 2019, *L'etica tra genetica e neuroscienze: libero arbitrio, responsabilità, generazione*, Carocci, Roma.
- Nichols, S. 2004, *Sentimental Rules: On the Natural Foundations of Moral Judgment*, Oxford University Press, New York.
- Quong, J. 2010, *Liberalism without Perfection*, Oxford University Press, Oxford.
- Reichlin, M. 2017, "The Moral Agency Argument Against Moral Bioenhancement", *Topoi*, vol. 38, n. 1, pp. 53-62.
- Sauer, H. 2017, *Moral Judgments as Educated Intuitions*, MIT Press, Cambridge (MA).
- Schaefer, G.O. 2015. "Direct vs. Indirect Moral Enhancement", *Kennedy Institute of Ethics Journal*, vol. 25, n. 3, pp. 261-289.
- Schubert, C. 2017, "Green Nudges: Do They Work? Are They Ethical?", *Ecological Economics*, vol. 132, pp. 329-342.
- Songhorian, S., Guma, F., Bina, F., Reichlin, M. 2022, "Moral Progress: Just a Matter of Behavior?", *Teoria. Rivista di Filosofia*, vol. 2, pp. 175-187.
- Soon, C.S., He, A.H., Bode, S., Haynes, J.D. 2013, "Predicting free choices for abstract intentions", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, n. 15, pp. 6217-6222.
- Thaler, R.H., Sunstein, C.R. 2008, *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Yale University Press, Yale.

Abstract

A relevant challenge in contemporary ethics is to understand whether and how individual moral improvement is feasible. Assuming the ineradicable presence of endogenous and exogenous conditions that make it difficult for the agent to control her action and choice (§1), I argue that theories developed from the observation of the limits of rationality and conscious control that human beings exercise over their decisions and actions are closely related to the question of free will (§2). I present two possible approaches to achieve individual moral improvement, showing their strengths and weaknesses. One proposal advocates nudges and suggestions to enhance people's

moral judgments (§3), whereas the other identifies ways to increase the subject's agency (§4). I conclude by arguing that developing procedures that can strengthen the subject's free will makes it possible to think of genuine and stable moral improvements because it generates would be enhancements not in any specific outward behaviors but in the individual's general moral attitude (§5).

Keywords: moral improvement; moral enhancement; free will; agency; nudging.

Francesca Guma
Vita-Salute San Raffaele University
guma.francesca@hsr.it