

Kantian moral universalism, the “Enlightenment Project” and experimental ethics

Vadim Yu. Perov^{1*}

¹Saint Petersburg State University, Institute of Philosophy, 7/9 University Embankment, 199034 Saint Petersburg, Russia

Abstract. The main ideas of Kant’s moral philosophy were embodied in what can be called the “Enlightenment Project”. Kant’s calls for freedom, anti-paternalism and the requirement of abandonment of concern for moral behaviour are associated with ethical ideas of autonomy, the categorical imperative, negative and positive freedom, the universality of morality in relation to the human being as a rational actor etc. These ideas are consistent with the ideals of the “Enlightenment Project”. The following ideals can be distinguished: 1) the ideal of classical scientific rationality; 2) the idea of “pure reason”; 3) free, equal and autonomous individuals; 4) moral universalism; 5) the creation of a united human civilisation; 6) the ideals of moral progress and universal happiness. In contemporary philosophy and ethics the main objects of criticism are the ideals of a free and autonomous individual, as well as ideas about the possibility of the existence of universal morality. In the context of the discussion of the possibility of the existence of rationally justified universal morality and the existing moral pluralism, the results of the experimental on-line research of the Massachusetts Institute of Technology (MIT) – “Moral Machine” – are interesting. The analysis of the results of this study shows the possibility of the existence of different types of rational universality (scientific-epistemological and moral). But while scientific universality requires theoretical unity of empirical data, moral universality firstly allows pluralism of norms and values and, secondly, requires a generalised normative regulation of empirical moral diversity.

Keywords: Enlightenment Project, moral universalism, moral machine, experimental ethics

1 Introduction

With his short essay *An Answer to the Question: What is Enlightenment?* I. Kant to a large extent revealed the value and normative dominants of what is commonly called Modernity. It is also important that these ideas were based on his moral philosophy. The “motto” of the Enlightenment formulated by him, “Have courage to make use of your own understanding!”, demands for personal freedom and ethical anti-paternalism are correlated with the ideas of

* Corresponding author: vadimperov@gmail.com

ORCID: 0000-0002-9925-9527

autonomy and the universality of human morality as characteristics of a rational being, etc. The subsequent development not only of philosophy, ethics and political thought, but also historical events and processes of social change have largely confirmed the relevance of these Kantian thoughts. But in the second half of the 20th century critical views emerge that cast doubt on the moral and socio-practical correctness of the ideas of the Enlightenment, including their Kantian interpretation. Discussions about what has received the general name “The Enlightenment Project” are becoming widespread, the main questions being not only “what is the Enlightenment?” but also how correct it is both in itself and in the context of its historical implementation. Directly opposite theoretical positions can be noted. On the one hand, there is the point of view of A. MacIntyre: He denies the ideas of ethical universalism, defends the position of communitarianism and asserts that the Enlightenment Project was doomed to failure, due to the presence of a number of theoretical and methodological contradictions in it. On the other hand, J. Habermas believes that humanity has not even begun or is at the very beginning of the implementation of the ideas of the Enlightenment (modern society) in reality. This opposition of points of view requires a new look at the question “What is the Enlightenment?” in terms of its moral meaning. This question becomes more relevant with the advent of experimental philosophy, which challenges the moral philosophy of recent centuries.

2 The “Enlightenment Project”: Moral ideals and imperatives

It is extremely difficult to investigate the “pure” moral content of the “Enlightenment Project”. Most ethical ideas almost inevitably become social, and even more so, political reasoning. This can be seen in some of the works of Kant himself, in which the ideas of autonomy and *a priori* morality “dissolve” into socio-political reflections on the state, law, civil society, etc. It should be assumed that the subject in question itself sets a perspective that cannot be avoided. Nevertheless, it is worth making a reservation that the provisions formulated below will focus precisely on the ethical components of the “Enlightenment Project”, leaving aside a number of other important aspects.

It should be borne in mind that the “Enlightenment Project” is not some chronologically and formally defined event with a fixed “membership” of participants. The Enlightenment is a set of worldview positions based on close, but far from identical, scientific, social, political, cultural, philosophical and moral ideas and values, which have largely determined the development of European and then all human civilisation over the past 200-300 years.

As a basis for understanding, we can take the almost classical definition of the Enlightenment (“Modernity”) proposed by J. Habermas and S. Ben-Habib (1981, p. 9) in their work “Modernity versus Postmodernity”: “The Project of modernity formulated in the eighteenth century by the philosophers of the Enlightenment consisted in their efforts to develop objective science, universal morality and law, and autonomous art, according to their inner logic. At the same time, this Project intended to release the cognitive potentials of each of these domains to set them free from their esoteric forms. The Enlightenment philosophers wanted to utilize this accumulation of specialized culture for the enrichment of everyday life, that is to say, for the rational organization of everyday social life. Enlightenment thinkers of the cast of mind of Condorcet still had the extravagant expectation that the arts and the sciences would promote not only the control of natural forces, but would also further understanding of the world and of the self, would promote moral progress, the justice of institutions, and even the happiness of human beings. The 20th century has shattered this optimism.”

Concretising and supplementing this definition, it is possible to formulate the following main moral and ideological features of the “Enlightenment Project” which, to a greater or

lesser extent and with different accents, are reflected in the majority of its researchers and critics.

1) The formed “classical ideal of rationality” of the natural sciences began to be identified with rationality itself. At the same time, the prevailing point of view was formed that it is precisely this scientific and epistemological rationality that is the essential characteristic of human nature – homo sapiens. It is important that this rationality is precisely the “ideal” with properties corresponding to any ideal.

First, this “ideal of rationality” can be interpreted as an “ideal object” of any scientific experimental theory. And like any “ideal object” it is a theoretical abstraction and the result of mental work with empirical data, and, in a strict sense, it does not exist in reality. The “classical ideal of scientific rationality” is as real as the “point particle” or “ideal particle” of Newton’s classical mechanics. But without such “ideal objects” true experimentally and logically verifiable scientific knowledge is impossible.

Secondly, the “classical ideal of rationality” as an ideal performs not only epistemological, but also value and normative functions. It is a kind of moral imperative that requires a person to be reasonable. This determines the ethical perspective of the “Enlightenment Project”: a person should be enlightened, it is his moral duty. This means being guided by one’s own reason, which must be based on reliable and verifiable scientific knowledge. The moral obligation of a person is actually charged with the requirement to take care of his own mind, its enlightenment and education.

2) “Pure mind”. Only a subject whose mind is free from everything that interferes with the activity of the mind can achieve true knowledge. This understanding is historically embodied in what can be called the “programme of cleansing the mind”, which is formulated in the philosophy of F. Bacon and his doctrine of “idols of the mind”. Different versions of this “programme” can be found in the theory of the “clear and distinct” ideas, including moral ones, by R. Descartes, the concept of *tabula rasa* by J. Locke, in the rationalism of T. Hobbes and B. Spinoza and, of course, in Kantian apriorism of pure reason and pure practical reason. A person should not look for the reliability of knowledge and his moral standards and values in something external, be it historical traditions, past experience or religious dogmas. The latter circumstance served as an essential basis for worldview secularisation. Not only scientific knowledge refused to be based on religious ideas, but also the ethical provisions of religion had to be cast into “fundamental doubt”, pass a test for rationality and be accepted only because of their rationality (in this context, the title of Kant’s work *Religion within the Bounds of Bare Reason* is an illustration). Only one’s own enlightened mind, which has been cleaned of all unnecessary things, can and should be the only support for a person not only in the epistemological, but also in the moral sense.

3) “Free, equal and autonomous individuals”. The concept of equal epistemological subjects was almost completely extended to the “moral agent”. People acting within the framework of moral relations within the framework of the “Enlightenment Project” began to be understood not just as morally free (independent, self-determining, autonomous, etc.), but also as individuals equal in this moral freedom. This denies the idea, prevalent in previous history, that some people are more moral in nature due to their birth, possession of sacred knowledge, etc. It is this idea of the existence of “morally privileged” people that is the basis for “moral paternalism”. This “moral paternalism” assumes that most people are morally weak, “unreasonable children” who are unable to act morally correctly on their own. Therefore they need not only paternal care, but also direct moral guidance from the “adults” who can even force them to a “morally correct life”. The rejection of paternalism and the call for the moral self-determination of free and equal individuals who independently decide what their right life consists of can be considered as the central ethical idea of the “Enlightenment Project”, which was itself considered as a criterion for social and political institutions (state,

law, civil society etc.). “Thus the civil condition, regarded merely as a rightful condition, is based *a priori* on the following principles:

1. The *freedom* of every member of the society as a human being.
2. His *equality* with every other as a *subject*.
3. The *independence* of every member of a commonwealth as a *citizen*.

These principles are not so much laws given by a state already established as rather principles in accordance with which alone the establishment of a state is possible in conformity with pure rational principles of external human right” (*TP*, AA 08: 290; Kant, 1996, p. 291).

4) “Moral universalism”. Reliable scientific knowledge, which is individual in the way it is obtained, becomes true knowledge only by virtue of its universal nature. The same is true of moral knowledge, which is knowledge of universal moral laws and is rationally testable “in the perspective of humanity as a whole”. As MacIntyre (2007, pp. 43-44) notes: “Central to Kant’s moral philosophy are two deceptively simple theses: if the rules of morality are rational, they must be the same for all rational beings, in just the way that the rules of arithmetic are; and if the rules of morality are binding on all rational beings, then the contingent ability of such beings to carry them out must be unimportant - what is important is their will to carry them out. The Project of discovering a rational justification of morality therefore simply is the Project of discovering a rational test which will discriminate those maxims which are a genuine expression of the moral law when they determine the will from those maxims which are not such an expression.”

5) The specified characteristics of a person as a moral agent were understood as the basis for the theoretical substantiation of the socio-political reorganisation of human life. In an extremely simplified form, the views that developed during the Enlightenment can be represented as follows: There are intelligent and enlightened people who know what exactly and why they are doing something. Their knowledge is reliable since it is based on knowledge of the objective laws of nature and morality, therefore they can critically test their own and other people’s knowledge and moral convictions. Political, religious, moral and other disagreements between people are the result of mistakes or limited knowledge. They can be successfully overcome in the same way as scientific disagreements, i.e. through more thorough research, education and discussion. Scientific disputes are resolved by rational arguments, not by armed conflict. Therefore in the coming “enlightened world”, instead of waging ruinous and painful wars, people and governments will conclude treaties. The idea of the contractual nature of morality was a key motive of the social and political thought of the seventeenth and eighteenth centuries as formulated by T. Hobbes, J. Locke, J.-J. Rousseau, etc. And just as scientific debate leads not only to agreement, but also to better knowledge, social and political discussions on the reasonable moral foundation will necessarily lead to the best society (*cf.* Kant’s *Perpetual Peace: A Philosophical Sketch*).

6) “Happiness.” It is known that Kant’s understanding of the “ethics of happiness” is ambiguous. On the one hand, happiness cannot be the goal of moral behavior. On the other hand, Kant repeatedly emphasises that the categorical imperative does not require people to give up being happy. In addition, from his point of view, behaviour based on moral duty makes a person worthy of happiness. Without going into the details of Kant’s understanding of happiness, it should be noted that it corresponds to the previously described ethical guidelines of the “Enlightenment Project”: “As for the *freedom* [of every member of a state] as a human being I express its principle for the constitution of a commonwealth in the following formula: No one can coerce me to be happy in his way (as he thinks of the welfare of other human beings); instead, each may seek his happiness in the way that seems good to him, provided he does not infringe upon that freedom of others to strive for a like end which can coexist with the freedom of everyone in accordance with a possible universal law (i.e., does not infringe upon this right of another)” (*TP*, AA 08: 290; Kant, 1996, p. 291).

The “moral image of the world”, which is formulated on such principles, looks very attractive and, at first glance, does not raise any particular objections. At least as an ideal, which may be unattainable, but the pursuit of which is a moral duty and evidence of moral progress. The thinkers of the Enlightenment saw this progress in the form of a society in which free and equal people independently and without guidance from others decide how to live. In this world, all important issues including moral ones are the subject of discussion, and potential conflicts are resolved through negotiations, and everyone has a “right to happiness”. The moral component of such a world seems so obvious that Kant uses the specific argument that it should be obvious even to a child how to act morally correctly (*TP*, AA 08: 286; Kant, 1996, p. 288). Today, such optimism looks naive and utopian, although a possible negative assessment of the results of the practical implementation of the “Enlightenment Project” is not so unambiguous. On the one hand, the supposed “happy world” did not arise. On the other hand, the people of Europe, North America and many other countries today are living much better than 200 years ago, and this confirms progress. But we must admit that this progress has historically been largely achieved at the cost of numerous victims and extremely morally dubious means (world wars, colonialism, neo-colonialism, exploitation of people and nature, Nazism, genocide, etc.). Few doubt the existence of scientific and technological advances that have improved people's lives and expanded their abilities, but this is not enough to recognise whether people's lives have become happier. In addition, advances in science and technology have given rise to new problems, including the threat to the existence of the world. This list of real and illusory contradictions can be continued.

It should be noted that the numerous and varied criticisms are very similar in argumentation. Reasoning is usually aimed at understanding, justifying and analysing the “Enlightenment Project” that exists in philosophical theories, and the denial of the “free, equal and autonomous individual” and “moral universalism” is justified by such statements as “but in reality everything exists differently”. As J. Schmidt (2000, pp. 737-738) correctly notes: “Critics of the Enlightenment typically begin either by noting a suitably appalling current practice, which is then linked to what is alleged to be a questionable principle [...] or by examining a questionable principle, which the critic then illustrates with a particularly grating example [...] Once the link between dubious principle and appalling example has been made, the critic typically proceeds to argue that this principle is the legacy of something called “the Enlightenment Project”: a set of intentions, originating in the eighteenth century, that still work mischief two centuries later [...]” Even if we ignore Schmidt's conviction that critics often choose dubious principles, he correctly reflected the widespread attempt to refute or substantiate the provisions of moral-philosophical theoretical constructions by referring to empirical facts.

3 The “Enlightenment Project” and experimental ethics

Much of the criticism of the “Enlightenment Project” focuses on the moral ideas of “free, equal and autonomous individuals” and “moral universalism”. The main conclusions of this approach are: a) people are not completely autonomous and rational in their moral behavior; b) universal rational morality does not exist in principle. Morality exists in local and even individualistic morals. The most essential is the problem of pluralism of disparate moral norms and values, which are the source of conflicts that do not have a universal rational method of resolution.

On the one hand, refuting theoretical positions by simply pointing out what people think about moral behaviour and how they act morally is directly contrary to the ideas of Kant's ethics. On the other hand, in contemporary philosophy, there are studies that are called experimental philosophy (X-Phi), including experimental moral philosophy. One of the

central ideas of experimental philosophy is the idea of the fundamental diversity of concepts both in philosophy and among people. “Work in experimental philosophy suggests that there is diversity even in the most basic concepts we deploy in Western philosophy” (Knobe and Nichols, 2007, p. 11). The results obtained from both mental and real experiments provoke very ambiguous reactions and assessments, but at the same time they are the subject of numerous discussions and the basis for further research. The focus of experimental moral philosophy is the study of moral relativism based on the opinion of a majority of people, which is called “folk moral relativism” (Sarkissian *et al.*, 2011). Although experimental philosophers direct their criticism primarily against analytic philosophy, this criticism can be extended to Kantian philosophy, which can rightfully be addressed as “armchair philosophy” (it is significant that the unofficial emblem of experimental philosophy is the image of a burning armchair).

In the context of discussing the possibilities for the existence of rational theoretical universal morality and empirical moral pluralism, the results of an experimental online study of the Massachusetts Institute of Technology (MIT) called “The Moral Machine” (the Project began in 2016) are interesting. As written in the site description: “This website aims to take the discussion further, by providing a platform for 1) building a crowd-sourced picture of human opinion on how machines should make decisions when faced with moral dilemmas, and 2) crowd-sourcing assembly and discussion of potential scenarios of moral consequence” (Massachusetts Institute of Technology, 2016).

On the website the researchers formulated for discussions and decisions the different situations based on the well-known “trolley problem”, which was formulated by P. Foot (1967) and became known through J. Thomson (1976) who started a tradition of considering variable scenarios. In addition, any participant in the experiment can offer their own scenarios, which will be further considered by others. This platform gathered 40 million decisions in ten languages from millions of people in 233 countries and territories. As a result, “the four most spared characters are the baby, the little girl, the little boy, and the pregnant woman” (Awad *et al.*, 2018, p. 60).

During the experiment some differences in preferred actions were identified, depending on the country and region of residence of the participants. But the main question arises: Can these individual and cultural differences be made part of the “moral machines”, programmed to act in accordance with the moral preferences of the majority? Is it ethically correct to adjust machine algorithms to the peculiarities of the moral preferences of residents of different countries and regions? The answer of the authors of the study is unequivocal: “In summary, the individual variations that we observe are theoretically important, but not essential information for policymakers” (*ibid.*, p. 61). True, they offer no explanation for this conclusion. Not only do they fail to offer detailed explanations but they limit themselves to pointing to the normative document “Ethical Rules for Automated and Connected Vehicular Traffic” (Germany). These rules specifically emphasise that existing differences between people cannot be the basis for making decisions. “9. In the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited. It is also prohibited to offset victims against one another. General programming to reduce the number of personal injuries may be justifiable. Those parties involved in the generation of mobility risks must not sacrifice non-involved parties” (Ethics Commission Automated and Connected Driving, 2017, p. 11). This means that these rules require the abandonment of empirically existing differences. Thus, a significant difference arises between the existing preferences of people’s moral choice revealed in the course of experimental research and the requirements to disregard them in the course of making moral decisions (the famous “is-ought problem” can be mentioned here). To solve this problem, it should be remembered that within the framework of the “Enlightenment Project” rationality and other characteristics of a person (freedom, autonomy, etc.) are

considered not only as a reality, but also as a requirement. The formulated ideals of the Enlightenment are not only scientific facts, but also peculiar moral imperatives. One of the consequences is the realisation that different types of rationality are possible (Kant's theoretical and practical reason, Habermas's instrumental and communicative actions, etc.) and imply different universalities. And if scientific universality inevitably strives for uniformity of theoretical and empirical verification, then a moral one allows, firstly, a pluralism of norms and values, and secondly, the possibility of normative regulation of the empirical reality of human life and society. But this is only possible if there is a "basic framework normative order" based on the ideals of the "Enlightenment Project".

4 Conclusion

The article highlights the main features of the "Enlightenment Project" which are of moral importance. The experimental philosophy that has emerged in recent years challenges the basic moral ideals of the Enlightenment, above all the ideals of an autonomous, free and rational individual. Based on the analysis of the experiment "Moral Machine" of the Massachusetts Institute of Technology (MIT), it is shown that the existence of diverse moral preferences of people does not cancel the possibility and necessity of moral universalism. In this regard, the moral ideals of the Enlightenment, including the principles of the ethics of Kant, are relevant moral ideas.

Acknowledgements. This research was supported by Russian Foundation for Basic Research, Project 19-011-00234 "Moral normality in professional ethical codes in the conditions of formation of a digital society".

References

- Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F. and Rahwan, I., 2018. The Moral Machine Experiment. *Nature*, 563, pp. 59-64.
- Ethics Commission Automated and Connected Driving, 2017. *Ethical Rules for Automated and Connected Vehicular Traffic. Appointed by the Federal Minister of Transport and Digital Infrastructure*. Federal Republic of Germany, pp. 1-33.
- Foot P., 1967. The Problem of Abortion and the Doctrine of the Double Effect. *Oxford Review*, 5, pp. 5-15.
- Habermas, J. and Ben-Habib, S., 1981. Modernity Versus Postmodernity. *New German Critique*, 22, pp. 3-14.
- Kant, I., 1996. On the Common Saying: That May be Correct in Theory, but it is of no Use in Practice (1793). In: I. Kant, 1996. *Practical Philosophy*. Edited by M.J. Gregor, translated by A.W. Wood. Cambridge: Cambridge University Press, pp. 277-309.
- Knobe, J., and Nichols, S., 2007. An Experimental Philosophy Manifesto. In: J. Knobe and S. Nichols, eds. 2007. *Experimental Philosophy*. Oxford: Oxford University Press, pp. 3-14.
- MacIntyre, A., 2007. *After Virtue: A Study in Moral Theory. Third Edition*. Notre Dame: University of Notre Dame Press.
- Massachusetts Institute of Technology, 2016. *Moral Machine*, [online] Available at: <<https://www.moralmachine.net/>> [Accessed 15 October 2020].
- Sarkissian, H., Park, J., Tien, D., Wright, J.C. and Knobe, J., 2011. Folk Moral Relativism. *Mind & Language*, 264, pp. 82-505.
- Schmidt, J., 2000. What Enlightenment Project? *Political Theory*, 28(6), pp. 734-757.
- Thomson, J., 1976. Killing, Letting Die, and the Trolley Problem. *The Monist*, 59, pp. 204-217.