

Sven Nyholm, *Revisiting Kant's Universal Law and Humanity Formulas*, De Gruyter, Berlin/Boston, 2015

Jelena Govedarica<sup>1</sup>

At the beginning of his book *Revisiting Kant's Universal Law and Humanity Formulas* Sven Nyholm asks: "Why Yet Another Book on How to Interpret Kant's Ethical Theory?" This may seem like an odd question to begin with. However, since there are countless books on Kant's ethics, many written by distinguished Kant scholars, the author's desire to lay out his reasons for writing a comprehensive analysis on this subject is quite understandable. He offers two main reasons: firstly, Kant's theory is widely misunderstood in contemporary Anglophone moral philosophy and that is why many objections to his theory fail; secondly, contemporary ethics is heavily influenced by "methodological intuitionism", often leading to moral relativism or skepticism, and Kant's theory can offer an alternative stance, which he calls "moral constitutivism". In answering this question, Nyholm also wants to emphasize his opinion that the majority of influential interpretations of Kant's ethics are in fact distortions of his theory; if we truly want to understand Kant's thoughts, we must read his words carefully and accept that nothing in his theory is redundant.

699

The book itself is divided into four main parts, of which the first one is the introduction. In the introductory chapter, Nyholm explains in what sense Kant's ethics is a "kind of constitutivism", building his view on Christine Korsgaard's "Constitutional Model" of human agency and its relation to the principles of morality (p. 12). Korsgaard offered two models for understanding human agency: "combat" and "constitutional" model. While the former model explains human agency as the outcome of a struggle between different forces within the agent, such as passions, desires, impulses, etc. (the paradigm example for this model is Hobbes's moral theory), the latter pictures human agency as an act of self-constitution. Korsgaard thinks that Plato and Kant are the best representatives of "moral constitutivism". Plato gave the first hint by claiming that the unity required for human agency is the same kind of unity that the republic with just constitution has, and that virtue is the outcome of the inner harmony among different parts of our soul with reason acting as a ruler, helping people achieve inner peace. On the other hand, Kant claimed that the distinction between agency and a mere behavior or reaction to impulses consists in the fact that it is authored and that the unity necessary for an action cannot be achieved without commitment to morality because we constitute ourselves by making laws for ourselves. Furthermore, Korsgaard claims, the only way to be well-constituted is "by

<sup>1</sup> PhD student, Faculty of Philosophy, University of Belgrade; govedarica\_jelena@yahoo.co.uk.

governing yourself in accordance with universal principles which you could will as laws for every rational being” (p. 13). Thus, agency is the result of a person’s constitution and it sets a criterion for judging the morality of an action.

Nyholm affirms that moral law is the constitutive principle of *particular kind* of agency and that Kant’s theory is a theory of moral constitutivism, but he also argues that Korsgaard’s view is much more stringent than Kant’s. Namely, while she claims that agency is *only* possible if we act according to the moral law we set for ourselves, Kant thought that agency is actually possible outside of the moral law and that is in fact why moral law is an *imperative* for human beings. Because of freedom that each human has, all of us exercise some form of agency: if, for example, someone chooses not to be guided by moral principles, but rather by maxims which promote their own happiness, they enjoy freedom of the will in the sense they are not absolutely determined by impulses (like animals are), and if someone acts in accordance with the categorical imperative they can be considered autonomous because their will is independent from determination through the sensual drives. Hence, the agent is any creature capable of exercising some degree of impulse-control by reason. “The agent who is *not* governed by the moral law is somewhere on a *continuum* between the following two extremes: full autonomy through the rule of her own practical reason, at the one extreme, and a complete lack of agency and freedom through the rule of irresistible impulses, at the other extreme” (p.14). Accepting that moral law is constitutive principle of fully autonomous agency (or positive freedom), Nyholm relates this idea to the humanity formula. Since we are beings with dual nature – rational, capable of autonomy, and animal, subject to needs and desires – we can fully realize our humanity only if our pursuit of happiness does not come before our autonomy. In other words, if our actions are always guided by principles which could serve as universal laws of autonomous human agency, without abandoning our personal pursuit of happiness, the preservation and full realization of humanity becomes the most general purpose of morality. Therefore, Nyholm concludes, Korsgaard is right in claiming that human beings constitute themselves through the principles they choose for their actions, adding that full realization of humanity is only possible if the sensual aspect of human being is governed by the rational one, that is, only if people subject themselves to maxims that can serve as universal laws.

Furthermore, Nyholm points out that Kant had accepted Plato’s analogy between just political constitution and just constitution of the soul, and this is an additional argument in favor of Kant’s moral theory being a kind of constitutivism. Namely, the author claims that Kant formed his view regarding the freedom of the will by employing Rousseau’s theory of civil liberty: just as people can only achieve civic liberty by subjecting themselves to laws they give to themselves as citizens, so too human will can only be free from external rules by subjecting itself to self-legislated laws. The analogy, however, ends here. When explaining the nature of moral laws autonomous people set for themselves as guiding principles for their actions, Kant uses the analogy with laws of nature, as defined in his critical philosophy. Laws of nature serve as a model on which we form our moral laws, and that is why Kant had argued that the universal law formula is also equivalent to the law of nature formula. In the formal sense the term “nature” refers to a particular constitution or a type of being that is possible in accordance with some universal law. The term “reason-endowed nature” refers to beings who can be governed by their own practical reason, and whose existence is possible in accordance with maxims such beings give to themselves that are fit to serve as universal laws. Consequently, when reason-endowed beings act according to such principles, they are making the existence and realization of their nature into the most general end or purpose around which all other actions are organized. On the other hand, human beings are also animals with needs and desires, from which their general wish for happiness springs. So, when humans make their twofold nature and full realization of humanity into the most general purpose of their action, they never treat other human beings only as means for some other purpose, but at the same time as a purpose in itself. Thus, the author concludes that the principle of humanity is the constitutive principle of human nature. Extensively explaining these arguments and contrasting them with various objections all throughout the book, he will offer a plausible argumentation in favor of Kant’s claim that the humanity and the universal law formulas are equivalent.

The second chapter is dedicated to a detailed analysis of the universal law formula. Nyholm firstly presents the main features and characteristics of prevailing understanding, or as he calls it “the standard reading”, of Kant’s universal law formula, in order to explain the background

assumptions leading to “the various seemingly devastating objections that have been raised against” it (p. 24). These objections are summarized in Derek Parfit’s book *On What Matters*, but they are also previously put forward by authors such as Barbara Herman, Onora O’Neill, Thomas Hill, Christine Korsgaard, and Allen Wood. Nyholm points out five features of this „standard reading“, showing that mentioned authors base their objections on some peculiar understanding of the meaning of Kant’s term „maxim“, or on an insertion of the notion of rationality into his universal law formula, or on a quite common disregard for the wider context of his ethics and philosophy as a whole. Having done this, the author then presents seven objections against the universal law formula, designed to illustrate that it has completely absurd and outrageous implications. Going through these objections one will surely understand better Nyholm’s motives to write “yet another book on how to interpret Kant’s theory”, as the extent of distortions, interpretative mistakes and overlooked fundamental features of Kant’s theory had led some philosophers to draw conclusions “that we are morally permitted to engage in convenience-killing, mass-murder, deception when it benefits us etc., while at the same time forbidding us to save drowning children if we are egoists and requiring us not to devote our lives to philosophy or to move to Iceland, etc.” (p. 30).

Moreover, Nyholm stresses, most of these interpretations show a remarkable misunderstanding of the claim Kant repeatedly makes that the universal law and humanity formula are two formulations of one law, or in other words that if we subject ourselves to maxims that could hold as universal laws, we are in fact treating the humanity in each person as a purpose in itself. Being unable to understand the logic behind this so called “equivalence claim”, some of these authors simply ignore it, and some analyze two formulations as different laws, attempting to prove that Kant was mistaken as the “equivalence claim” is false. Because of this, Nyholm offers “eight steps towards new understanding of the universal law formula” (p. 32) devised to help the readers form a better understanding of the universal law formulation of the categorical imperative, to show in detail why Kant’s “equivalence claim” is undeniably true and why all mentioned objections must fail. Indeed, if the readers had any doubts or confusions about the essential features of the categorical imperative, Nyholm’s “steps” would help them understand every aspect of Kant’s thoughts by indicating

crucial terms and their meaning, underlining Kant’s examples and clarifications, as well as notions which have to be considered within wider context of his entire philosophy.

In the third chapter Nyholm thoroughly investigates the humanity formula with respects to its interpretations offered by influential contemporary scholars who teach in the American universities – the members of “the American school of Kant’s interpretation” as he calls them (p. 46). What they all have in common is the conviction that the humanity formula does not have the same content as the universal law formula, because the former, as opposed to the latter, affirms one substantive value – the absolute value of all human beings. The point of dispute here is Kant’s reasoning leading up to articulation of the humanity formula. The American school believes that, on the one hand, he claimed that the humanity in each person has an absolute value, and that, on the other hand, he did not defend this claim with the arguments having to do with the universal law formula being the constitutive principle of human agency, but rather on some other grounds (upon which they disagree amongst themselves). The second controversial issue is their understanding of the relation between what has an absolute value – a good will and what is an end in itself – the humanity within each person. Nyholm, however, argues that the only thing Kant had believed to have an absolute value is a good will, and humanity „has absolute value only in a derivative or conditional way, only insofar as it is related to the capacity for morality (i. e. the possession of a good will)“ (p. 47).

In order to show why the American school’s understanding of the humanity formula is unjustified, the author locates key sections in the *Groundwork for the Metaphysics of Morals* and main claims Kant himself had made before he put forward the humanity formula. At this point he is letting Kant speak for himself, cautiously clarifying his words and intentions to the reader, trying to hold up his own comments. After Kant’s words, Nyholm introduces arguments held by the members of the American school, most notably Wood and Korsgaard. When reconstructed attentively, their interpretation of the humanity formula implies that the humanity and the universal law formula are not equivalent and that Kant attributes the absolute value to the same thing he claims to exist and should be treated as an end in itself. The author suggests a simple test for this reading: if these implications were correct, there should be a match between the

reasons they think Kant had offered in favor of humanity being an end in itself in the *Groundwork*, and those he explicitly gave in other works such as the *Critique of Practical Reason* and *The Metaphysics of Morals*. However, bearing in mind previously cited passages, Nyholm concludes that Kant's own words do not confirm this reading and the only conclusion possible, other than Kant was completely inconsistent, is that the American school's understanding of the humanity formula is not accurate. This conclusion leads to Nyholm's reconstruction of Kant's reasoning, which passes the test of consistency with Kant's other works and, therefore, proves to be superior to those offered by highly praised Kant scholars. In addition, he writes: "In offering us a way of approaching the basic principles of ethics that derives these from the basic preconditions for the possibility of fully responsible self-governing agency, while at the same time recognizing that the human being also has (if you will) a "non-rational" side (which means that many of our personal values tied to our individual or communal conceptions of happiness are based in our "particularly constituted faculties of desire" or our personal sensibilities), Kant offers us an important alternative to the intuitionist approach to moral theorizing that plays such a dominant role in much contemporary moral philosophy." (p. 74)

The last chapter of this polemical book explores Kant's views on virtue, human flourishing, the highest good and permissibility in relation to the humanity formula, by refuting some further objections against his theory. The major source of misunderstanding on these subjects, Nyholm thinks, is the inability to detect two important distinctions Kant rigorously makes. The first one is between morality and legality: moral action is done out of respect for the moral law and legal action is merely in accordance with it. Furthermore, the action is permissible if it has the property of legality, but, as some commentators fail to notice, we cannot infer anything about its morality. The second distinction is between holiness and virtue: while it is in the nature of the holy will that it necessarily acts in accordance with the moral law, human beings are not capable of holiness; because of their twofold nature, people are only capable of virtuousness. Another recurrent source of misunderstanding is related to Kant's view of morally ideal human life and the content of his notion "the highest good" with respect to the humanity formula. Nyholm argues that, although concepts of permissibility and holiness do have a role in Kant's moral philosophy,

in the analysis of the humanity formula virtue is the key concept. However, many authors understood Kant to be claiming that holiness is the highest moral ideal (which is false because holiness is not even possible for human beings), and that the humanity formula's chief function is to determine permissibility of actions (which is also not the case because Kant's discussion is mostly concerned with virtue and, as already mentioned, there is no inference to be drawn from permissibility to virtuousness). In order to clarify Kant's thoughts on these subjects, Nyholm chooses the objections to the humanity formula being a test for permissibility made by T.M. Scanlon and Derek Parfit, and an interesting (albeit unjustified) analysis of the distinction between virtue and holiness in relation to the humanity formula in Rae Langton's discussion of Kant's correspondence with his admirer Maria von Herbert.

After careful analysis of Scanlon and Parfit's interpretation of Kant's examples, the author detects their basic assumption: that *permissibility* of person's action depends on what attitudes, motives, etc. they would be acting on in performing these actions (p. 82), which means that Kant is mainly interested in the legality of actions when he discusses the humanity formula. But, if Kant had been exploring the legality of actions, then he would have considered only whether the action is *compatible* with treating the humanity in each person as a purpose in itself, without going into that person's motives for acting in a certain way. Alternatively, if Kant had also been interested in morality of actions, then the question of whether actions are performed on the basis of, or out of respect for, the moral law would have become the crucial one. In order to test Scanlon and Parfit's objections, Nyholm presents Kant's own words in the suicide example, the example of leaving one's talents and capacities uncultivated, the example of being helpful to others and the example of the lying promise. He concludes that, firstly, Kant was also interested in morality of the actions and not just their legality, and secondly, permissibility of actions does not depend on a person's attitudes, motives, etc. Therefore, their objections fail because they overlook Kant's crucial distinction between legality and morality.

Having said quite a lot about Kant's theory about one aspect of human nature – the rational one – and its capacity for inner freedom (with regard to which human flourishing consists in acting in accordance with the categorical imperative), Nyholm calls our attention to the sensual or

animal aspect of human beings. Arguing against Langton's interpretation that morally ideal human life includes apathy understood as a state of indifference, he claims that human flourishing from the sensual aspect simply consists in happiness, viewed as "the achievement of a state where we can be content with our lives since everything goes in accordance with our desires and wishes" (p. 91). Besides morality, human beings also seek happiness, and because of that making the humanity in each person into an end of ours also involves making the happiness of every other person an important part of that goal. Therefore, Kant's term "the highest good" relates to both sides of our nature, and it must be understood as the best possible state of affairs we could imagine from the point of view of practical reason – achievement of both happiness and virtue. The highest good is possible "if we subordinate our pursuit of happiness to the condition that we always act in according with and on the basis of self-adopted basic maxims that are fit to serve as universal laws in accordance with which all human beings can fully realize their nature as the particular kind of reason-endowed beings they are." (p. 385)

Returning to the starting point of his book, in the concluding passages Nyholm defies Langton, O'Neill and Korsgaard's understanding of the humanity formula, as they claimed its main purpose is to enable everyone exercise agency. This

reading captures Kant's intentions only partially, because it does not take into account the sensual part of person's humanity and, consequently, happiness of an agent. Therefore, according to this reading, the highest good would not constitute the world "of virtue (which indeed involves the exercise of fully autonomous agency) in which everyone *also* gets all the happiness they desire and their virtue makes them worthy of" (p. 94). Making the full realization of the humanity in each person into an end of ours involves the flourishing of the whole human being – with respect to their capacity for autonomous agency, as well as their desire for happiness. Thus, Kant's suggestion to use the laws of nature as a type when choosing maxims in accordance with which we will act, means that we should choose the most general principles in accordance with which human beings can exist and fully realize their nature, and that both the capacity for agency and fulfillment of needs and desires have to be the objects of the moral law.

Overall, this study presents a systematic, carefully conducted evaluation and refutation of major objections against Kant's "universal law" and "humanity" formulations of the categorical imperative discernible in contemporary interpretations of his ethics, aiming to show that Kant's insistence on the fact that these formulations express the same law is not to be neglected or considered false.