

Jovan Babić

The Golden Rule and the Categorical Imperative

Jovan Babić – Professor of Ethics. University of Belgrade. 11000, Čika Ljubina 18–20, Belgrad, Serbia.
ORCID: 0000-0002-7591-4902
e-mail: jovanpbabic@gmail.com

The Golden Rule – “do unto others as you would have them do unto you,” or in its negative formulation “do not do unto others what you would not have done unto you” – is one of the most ancient formulations of moral criteria. We encounter it in some form in the canonical texts of practically all religions that aspire to any kind of universality. At the first glance, the categorical imperative states and demands the same as the Golden Rule. This presents the question of whether the categorical imperative is a variation of the Golden Rule, or whether, inversely, the Golden Rule is a different way of stating the demand that morality be one and the same the world over? In other words, can these two “rules” be reduced to each other? If this is indeed the case, it would represent an exceptional example of correlation between practice and theory, sophisticatedly elaborated in Kant’s moral philosophy. Alas, if the argument that follows in this text is correct, it will show that the categorical imperative is not a variation of the Golden Rule, nor is the Golden Rule a popular form of the demand placed by the categorical imperative. Furthermore, it will show that the Golden Rule does not present a deontological standpoint at all, and that it fails to guard against arbitrariness; it does not have the capacity to be a criterion of the particular kind of evaluation we call moral or issue the cardinal demand of universal objectivity and impartiality. None of which can be denied the categorical imperative, due to its precious property Kant labels as “formalism”.

Keywords: Golden Rule, positive and negative formulation of GR, categorical imperative, Immanuel Kant, formalism of categorical imperative, universalization, impartiality

The Golden Rule – “do unto others as you would have them do unto you,” [Mt., 7:12; Lk., 6:31] or in its negative formulation “do not do unto others what you would not have done unto you” (cited in: [Rost, 1986, 7]) – is one of the most ancient formulations of moral criterion. We encounter it in some form or other

in the canonical texts of practically all religions that aspire to any kind of universality. Seemingly opposed to it is what the Golden Rule declaratively confronts and is supposed to regulate – selfishness, in all guises.

On a theoretical, philosophical plane, egoism is the basis for the articulation of various *teleological* doctrines, such as eudemonism or utilitarianism. In opposition to such teleological standpoints is the *deontological*: a standpoint that does not start from any given purpose or characteristics of such purposes; instead, it begins with those determinations that ought be contained in a strong value *criterion* (as the moral one should be) – determinations of objectivity and impartiality. However, the Golden Rule was not *constructed* on the basis of these principles; within the history of ethics, they appear as foundational in formulation of Kant's approach. In his moral philosophy, the form of the general good is articulated within the world of universal respect: this is a world wherein respect takes the place that happiness, love, or servility held in other formulations of the criterion of value, and is known as the *categorical imperative*.

This imperative, in its first formulation, states: “Act only in accordance with that maxim through which you can at the same time will that it become a universal law” [Kant, 1996, 73]. At the first glance, the categorical imperative declares and demands the same as the Golden Rule, presenting the question whether the categorical imperative is a variation of the Golden Rule or whether, inversely, the Golden Rule is a different way of stating this demand that morality (whose universality assumes singularity) be one and the same the world over. Is Kant only presenting the Golden Rule in a philosophically precise way (as opposed to its more popularly understood form)? In other words, can these two “rules” be reduced one to the other, either as the categorical imperative being a variation of the Golden Rule, or the Golden Rule a popular formulation of the categorical imperative? Both principles assume a form of universalism, and what is more, both imply the principle of impartiality.

If they are indeed different ways of saying the same thing, it would represent an exceptional example of correlation between practice (embodied in the application, or at least its pretense, of a practical rule) and theory, sophisticatedly elaborated in Kant's moral philosophy. Alas, if the argument that follows in this text is correct, it will show that the categorical imperative is not a variation of the Golden Rule, nor is the Golden Rule a popular form of the demand placed by the categorical imperative. Furthermore, it will show that the Golden Rule does not present a deontological standpoint at all, and that it fails to guard against arbitrariness; it does not have the capacity to be a criterion of the particular kind of evaluation we call moral or issue the cardinal demand of universal objectivity and impartiality. None of which can be denied the categorical imperative, due to its precious property Kant labels as “formalism.”

* * *

In Kant's elaboration of the categorical imperative, we often encounter the argument that not only includes the principle of the Golden Rule but bumps up against the edge of a utilitarian explanation, in particular when considering duty to help others in trouble [Ibid., 75]. Still, there are significant differences. An analysis of these

differences will show that only one of these two principles (the categorical imperative) can be considered an authentic moral principle. The other, the Golden Rule, in contrast to its appearance, is only capable of producing various forms of long-term and relatively well guarded *self-love* (as Kant would call it); that is, everything which (broadly speaking) falls into the domain of human happiness, including naive egotism. Further, it will turn out that the most important, but also morally most dangerous, feature of the Golden Rule is that it provides the basis on which, under certain conditions and given context, one could justify all kinds of heteronomous ideals. This ensures not only an easy and efficient conformity to a given social, political, or religious situation, but provides justification for all manner of moral (or other) violence. Thus, the Golden Rule could offer justification of acts that could only be performed and justified under the cloak of such strong explanations, capable of producing strong convictions and readiness to reject as irrelevant all that does not fit into the framework of such explanations.

When asked about the potential equivalence of Golden Rule and the Categorical Imperative¹, Kant himself answered the question negatively, explaining that the necessary duty to the other, contained for example in the duty to truthfulness, cannot be reduced to “the trivial *quod tibi non vis fieri* etc.” because “it contains the ground neither of duties to oneself nor of duties of love to others (for many a man would gladly agree that others should not benefit him if only he might be excused from showing them beneficence), and finally it does not contain the ground of duties owed to others; for a criminal would argue on this ground against the judge punishing him, and so forth” [Kant, 1996, 80]². Kant’s well-nigh casual remark has caused continental philosophy to nearly completely neglect all theoretical considerations of the Golden Rule, either as a moral or rational principle of acting, even though it has remained in widespread use in both daily life and the religious relationship to the world. In Anglo-Saxon philosophy, however, the principle was much more often the object of attention, in both the past [Sidgwick, 1893, 380–389] and more recently (see: [Gewirth, 1978; Singer, 1963; Hare, 1963]). Which makes sense, as rational egoism has a much more central place there; moreover, one might argue that it could be shown that the principle of universal benevolence, which rests at the heart of utilitarianism, represents a variation (and one of the better ones at that!) of the Golden Rule.

In the very demand for elementary consistency, which precedes any mutuality and what we usually understand by universalization, requires that we behave the same (or similarly) in the same situations (or those similar in relevant ways). Only once this demand is fulfilled is there room to give a reason for some act; and if it is not satisfied, it implies complete arbitrariness and absence of justification. Indeed, if I am willing to neglect the demand of consistence, to ignore whether two situations are in a relevant sense *same*, it relieves me of the ability to give any reason for my action: any reason given in one such situation would as such, as *reason*, be

¹ There are, of course, other, opposite opinions. R.B. Brandt, thus, thinks that the categorical imperative is like any other maxim, and that to be universalizable, it must be compatible with the Golden Rule (cf.: [Brandt, 1959, 29]).

² The “trivial” is precisely the Golden Rule: “*Quod tibi non vis fieri, alteri ne feceris*”.

applicable in any similar situation. The reason itself does not necessarily contain reference to another subject, its generality is more abstract and demands the similarity be admitted and respected always, regardless of other circumstances. In that sense, the demand pledged in the reason can always be expressed in the form of a *hypothetical* imperative, which is always an indicative statement of the relation of some cause and effect: 'if I *want* to arrive faster, I *must* hasten'. On the other hand, its flouting is a fatal defect, as inconsistency abolishes the very possibility of justification, a rather broad position that refers not only to ethical standpoints, but the entire field of human practice. Inconsistency is a sure sign of absence of any possible reason for any given act. This means that such an act does not have a reason for its existence (which is why the defect is fatal). Of course, we could say that inconsistent acts are 'singular', that they cannot fit into any broader whole, nor any life plan, that they serve no purpose. Which is all true. Yet, what makes it important here is that preceding every moral principle, which in one way or another interprets and makes concrete the abstract universality built into the principle of consistency, is precisely this principle and the assumption of at least this minimal universality. For example, if we say that the difference between the Golden Rule and the categorical imperative is in the latter being based on *universalism*, while the former, in contradistinction, the emphasis is placed on *mutuality*, it cannot mean that the elementary demand of consistency must not be satisfied in both, that is, that both do not contain an at least minimal universalization required for this demand.

However, consistency, although necessary, is not sufficient in questions of morality. Another universal dimension is needed – according to which if I do something because I *ought* to (because it is morally *right*), or the other way round, if I do not do something because I *ought not* (because it is wrong) – that is, not merely because I would (not) like to or because I have (no) interest – then I must be ready to accept not only that I *ought* (not) do so in every such situation, but that this is true for *everyone else* in such a situation. Moreover, when I am considering whether something *ought* (not) to be done, if something that needs to be done to another person, then what must be taken into account is not only what it means for someone to do something like this, but also what it means to *them*, to that person. And since I am also only a person among persons, I am, consequently, *one of them*. When this is generalized, we arrive that it cannot be right for A to treat B in a way that would be wrong for B to treat A, *merely* for the reason of A and B being two different individuals, without any justification in difference or circumstance (or some other relevant and substantive difference).

We are now presented with the question of value of this principle. Abstractly put, as it is here, all it confirms is that moral judgments must contain a second dimension of universality. The demand of consistency established the need for the most basic universality, common to all universal moral norms and universal statements (judgments) we encounter, for example, in natural sciences. This is the minimal demand that is the necessary condition for an attempt at objectivity, and, it could be added, the basic demand of seriousness that ought to be the barrier against complete arbitrariness and chaotic relativism.

But the new dimension of universality introduces a new morally important dimension: mutuality. Mutuality places us into a world which contains, in addition

to valid logical and natural laws, an indeterminately large multiplicity of rational beings. The only instruction mutuality provides within this world is the prohibition on making exceptions to one's benefit. Thus, Henry Sidgwick says: "Such a principle manifestly does not give complete guidance – indeed its effect, strictly speaking, is merely to through a definite *onus probandi* on the man who applies to another a treatment of which he would complain if applied to himself" [Sidgwick, 1893, 380]. It is up to one who acts in this way to justify "the difference." Which puts anyone who does not immediately see such a difference in the position of seeking or expecting such justification; and yet, since not every reason is morally relevant, there is always the possibility that there is no difference. But even when there is no difference, it might not be significant to provide convincing justification. Therefore, the principle of impartiality contained in the principle, although it does not give "full instructions," has an easy applicability and great practical importance. It is incumbent upon one who acts to give reasons for those actions, which, by their nature, cannot be considered reasons at all unless impartiality is respected in the use of terms such as 'right', 'ought', 'good', or 'bad'.

For this reason, the demand of impartiality is also called "the rule of equity." Thus, for example, Sidgwick cites Samuel Clarke's rule of equity: "Whatever I judge reasonable or unreasonable that another should do for me: that by the same judgment I declare reasonable or unreasonable that I should *in the like case* do for him – which is of course the 'Golden Rule' precisely stated" [Ibid., 385]. Sidgwick, however, charges Clarke with tautology, because 'right' is defined as 'that which is reasonable to do', which leads to a certain circularity; nevertheless, Sidgwick thinks that an escape can be found in the empirical determination that as a rational being, every person is capable of accepting what is right to a lesser or greater extent, and that then the rational aim of any reasonable being is the *greater* measure of everyone's (or anyone's) good. In other words, universal benevolence is the best way of achieving universal good. The charge of tautology *in this sense* will also be leveled, as we will see, at other articulations of the Golden Rule. Indeed, ultimately, they all face a luminal determination: any moral principle determined through content, yet seeking to avoid the trap of tautology hidden in calling on any *definition* of 'right', will always imply a, perhaps also hidden, *ad hominem* argument.

And yet, it could be said that the Golden Rule contains no other demand other than a type of consistency, and that therefore its scope is insufficient. Alan Gewirth (cf.: [Gewirth, 1978, 162–172]), thus differentiates between so-called *generic consistency*, which assumes the use of the same rule in similar situations, that is, a general impartiality in application of the rule considered correct; and the "*appetitive-reciprocal consistency*", which demands others be treated according to the same rules applied to oneself, which encompasses the Golden Rule. But according to him, the Golden Rule has rational justification of the kind that would provide objectivity of moral judgment (which he thinks only the principle of generic consistency has), because the Golden Rule leaves the actor to determine on their own the level of generality at which they will describe their acts. They could, thus, describing them truthfully, only take into consideration their appetites or their arbitrary whims, on which they could then build the 'universal' rule about how to treat others.

The Golden Rule demands that the equal be treated equally and the unequal unequally. But the basic problem is the determination of the relevant measure of equality. The simplest articulation of this measure is to take the Golden Rule as one of egotistical lucidity: ensuring good by ‘obligating’ the other in advance, thus ‘obviating’ them treating us cruelly or without considering our preferences. This form of egoism cannot at all be called naïve; but the *mutuality* contained in this standpoint stems from the so-called *positive* formulation of the Golden Rule: ‘do unto others as you would have them do unto you’. Notwithstanding opinions that this positive formulation is not significantly different to its more popular, *negative* formulation (cf.: [Bruelisauer, 1980]), it is immediately evident that the positive formulation has serious drawbacks. The positive formulation advocates, so to speak, sinning together: it tells the masochist to become a sadist: ‘do unto others as you would have them do unto you’. Only from the standpoint of a maximalist, let us even call it saintly, morality could such moral demands be presented. This is the morality of the Sermon on the Mount – nor is it a coincidence that the sermon contains a positive formulation of the Golden Rule [Mt., 7:12; Lk., 6:31]. This is either a morality of renunciation – “Judge not, that ye be not judged” [Mt., 7:1] – or the maximalist morality of turning the other cheek [Lk., 6:32–33]. It must also be said, however, that the principle of impartiality is here expressed strongly indeed [Mt., 6: 2–6]. But what follows from it is that impartiality itself is a kind of guard against hypocrisy, a basis for the constitution of what we call conscience – which, however, in no way ensures moral autonomy or the value of freedom.

For its part, the *negative* formulation of the Golden Rule, ‘do not do unto others what you do not wish done to you’, has a few advantages. First, people are more inclined to agree about what they do not wish (done to them) than what they do; there is a certain axiological asymmetry³ between good and evil, happiness and suffering, just as there is asymmetry in the attitude towards opposing values: good and happiness, especially in others (precisely where the Golden Rule is relevant), do not elicit anywhere near the intensity of empathy that evil and suffering do. Second, the negative formulation avoids the masochist position paradox: others can have wishes and interests entirely different to our own. Further, it avoids the retributive implication of naïve egotism – ‘I do to you what you do to me!’ – which reduces the Golden Rule to merely a form of the well-known *lex talionis*, the principle of retribution⁴.

The negative formulation, however, is not without problems. One of the assumptions of the Golden Rule, and not at all implausible, is the understanding of people as uniform or single-minded in terms of emotions (which can here be understood as either empirical hypothesis or a postulate of value beliefs). Certainly, the Golden Rule rests on the assumption of *acceptance* that we all, generally, have the same capacity of emotion, not only in purpose, but in a practically even more important sense: that one who judges using the Golden Rule assumes that all others

³ This asymmetry exists at the foundation of many arguments in various theories, for example in the concept of “value strength” [Hartmann, 1935, 251–252, 544], in the negative formulation of utilitarianism (e.g.: [Smart, 1958]).

⁴ It is sometimes difficult to differentiate the Golden Rule from *lex talionis* in the Bible (cf.: [Gen., 9:6], but also: [Mt., 10:33; 1Cor., 3:7; 2Cor., 9:5]).

feel the same. Abstractly speaking, the understanding does not violate the demand of impartiality – all are indeed treated in the same way. The assumption that others feel *as I do* transforms into a normative thesis that all *should* be treated in the same way, etc. The general nature of this norm is grounded in that hypothesis. However, this has a completely unexpected relativist consequence: if I am, for example, *insensitive*, then the corresponding norm allows me to equally cause pain to others. And conversely, if I am hypersensitive, the care I will extend to others will be unnecessary, indeed excessive, and will be perceived not as care, but paternalism, violence even. Just as callousness will be *felt* in the case of insensitivity, without any real ill intention in either case.

In general, the question of practical applicability of the Golden Rule has never been as simple as it first appeared. Another difficulty in applying the rule ‘Do (not do) unto others as you would (not) have them do unto you’ issues from the problem of determining the exact meaning of ‘other’. To apply the rule in the context of competition (such as a sport) would seem to prohibit sporting contest itself. However, this is not the case. As such situations include institutional facts, and the institution is constituted and acquires its reality only through institutional rules in which competition is established as the very rule of behavior, it follows that the Golden Rule cannot be applied in this context. One who “enters the game” can wish to win or not be defeated; but by their very entry, they have *already* accepted the given rules, rendering the application of the Golden Rule void and meaningless. Such application would perhaps not be senseless *prior* to the game, but once the game is entered, it has nothing *to which* it could be applied. Similar argument applies to various other strictly defined human relations with a high degree of artificiality.

Things become less evident in relations that appear to us more ‘natural’, such as in the case of *property*. The institutional aspect resists the application of the Golden Rule, while the ‘natural’ seems to, at least partially, opens the door to its applicability and that often indirectly, through questions of justness the institution of property presents. In all these cases, the universality in the determination of the ‘other’ is taken as read, without much ambiguity.

However, in case of, for example, racial discrimination (or any discrimination for that matter), the determination of the other acquires crucial importance, as various possible interpretations open up. For a landowning slaveholder in the American South, a two-fold application of the Golden Rule is possible precisely due to different interpretations of the term ‘other’. The strength, or even the possibility, of the application of the Golden Rule depends on the assumption that these two persons, ‘I’ and ‘other’ are in a relevant sense similar⁵. Both as dual, and in their desired purpose (to ban their different treatment), the two persons are different. Each could be the referential framework that encloses the other. The determination

⁵ The *similarity* allows for the *other* to be *visible* in the first place, otherwise, they remain ‘invisible’, which is a problem for the articulation of the application of the Golden Rule (cf.: [Tarasenko-Struc, 2020]). A similar invisibility is manifested in our attitudes towards some animals: thus, a spider trying to escape and save itself will not elicit empathy, the use of the word ‘save’ remaining metagenic (‘trying to save itself’). There is a complete absence of the experience of *mutuality*, there is no ‘other’ that we could treat ‘inconsistently’.

of the 'I' can be universalized and expanded onto the 'other'; but equally, the determination of the 'other' can be designated as relevant and hypothetically put itself in the position of the initial 'I' ("if I were Black, primitive, uneducated, dirty, ill, a woman, on a low level of mental development, of low social status, etc."), and justify their position from there. The results of these two procedures will be different: if I say 'I will treat him, a Black man, the way I wish he (whoever he is) treat me, a white man, since I am white', it is entirely different to saying 'I will treat him, a Black man, the way he should treat me *if I were Black* (and he white)'. The implication that presents itself here is rather unpleasant: it means that the Golden Rule can not only fail to prohibit certain dubious moral practices, but just the opposite as well, it can justify them.

This does not exhaust problems that come from ambiguity. The rule formulation 'do not do unto others...' assumes an addition, perhaps redundant, but which could in its ambivalence pose significant problems. In the full formulation, the rule could be: 'What is right for one person is (should be) right for any other in a situation that is similar in a relevant way'. The problem emerges in the different interpretations of the word 'right': it can mean right for that person to do, but also the right way of treating them. The problem lies in that according to the latter, it might not be right (or could be downright bad) to treat the same person in the same situation in the same way if the description of the situation does not contain what is important for them in terms of character, ability, sensitivity, stamina, readiness, etc. Thus, in addition to the similarity of *situation*, we must consider the (relevant) similarity of the *person*⁶.

With all these issues in the background, the basic question in considering the Golden Rule is that of *universalization*. What does this mean? Can the Golden Rule be presented as an authentic moral principle? Does it guard from abuse, moral terror, heteronomy, attacks against moral integrity? Does it affirm or at least allow for freedom? Let us compare the following statements:

- 'Do (not do) unto others what you (do not) *wish* done unto you'
- 'Do (not do) unto others what they *should* (not) do unto you'

We can see that the universalization contained in the 'logic of meaning' of some words, such as 'should' is not the same as the 'logic of meaning' of some other words, such as 'wish'. If, for example, someone was to say 'I should act in this way, but no one else in a significantly similar situation should act in this way', we would have the impression that the speaker is abusing the word 'should'. Does that mean that (2) is an analytic statement? If it were, then its negation would be a contradiction. Opening the possibility for such questions provides a powerful means of examining not only the consistency of the statements that feature these terms, but also their practical applicability and validity. A differentiation needs to be made here, however: a practical principle can be invalid precisely because of the various possibilities of its use – for example, if it can be used for manipulation or covering and switching of value content⁷.

For the use of the word 'should' to achieve its purpose, it must be used in a universalizable way. Otherwise, one who uses principles that include "should" avoids

⁶ For more on this, see: [Singer, 1963, 15–17].

⁷ On the various aspects of the abuse of moral language, not only regarding the Golden Rule, but more broadly and specifically with what is called 'fanaticism', see: [Hare, 1963].

the duty that is inherent in their use, the duty to impartial and universal application of the criterion they contain. On the other hand, however, through their very use the speaker produces the appearance of acceptance of this duty. The merely simulated acceptance places the speaker in a position of a *moral judge*, making the evaluation made with these words a moral evaluation, a not an expression of a partial, factional, or selfish interest. This way, the partial interest is presented as universal, eliding potential resistance of those who do not have those interests, and more importantly of those who will be impacted by the actualization of this specific interest. This opens the possibility for a common and very powerful manipulation: those who stand opposed to the statement are accused that their resistance issues from the *fact* that they do not understand the *meaning* of the words used, and that all conflicts will be removed by *learning* the meanings of relevant words. Should they remain steadfast in negating the universality of this interest, they open themselves up to the charge of the very thing done through an improper use of universal moral terms and this parasitic and false universalization: the claim of *pars pro toto*, that partial, selfish goals have become general and moral – a clearly immoral position. Such manipulation in meaning of terms allows the manipulated party to ascribe ignorance where universal knowledge is assumed without demanding any particular expertise in recognizing the *meaning* of ordinary words, such as ‘good’ or ‘should’. Indeed, even if resistance nevertheless appears, the manipulation allows for any other meaning of these words, even the one most universal and most formal, to be declared of a given definition, getting away with grounding unacceptable and morally bad positions on the warped meaning of the words.

This manipulation should not be confused with hypocrisy. The manipulation is latently present in value discourse itself, that is, in the general characteristic of value words to be and act in a ‘propagandist’ manner; it is even contained as a tendency in any attempt at actualization of any ideal, due to the limit to universality imposed by its content, proffered as the ultimate criterion of value, which is thus ‘set apart’ from the subject being valued, ‘leaping over’ into the very measure of value. The measure of value and the object of value thus meld together into one whole, erasing the necessary internal border of value, resulting in a kind totalizing implication-tendency, contained in every ideal, to become the ultimate criterion of *all* value. Perhaps we would not be inclined to call all such content an ideal, but if we were to accept a minimal designation of an ideal as the content of a term or mental representation understood as value, then all such valuing, including along the lines of the Golden Rule, is in that sense ‘idealist’⁸.

⁸ The term ‘idealist’, in the ethical sense, should be elaborated here, but this requires a separate space. Fundamentalism, totalism, fanaticism are indeed latent designations of any such position, content aiming for universality. Such designations will always appear as practical implication of any such standpoint as soon as it is interpreted concretely in order to become actualized; in the process, the content becomes rather *rigid* – although there is always a tendency towards this – making the designations the dominant characteristics of this value standpoint. It seems that the only way to avoid this is to reject universalism, which is impossible in ethics; or else, in a Kantian manner, consistently avoid any positive and content-laden designation of our basic *criterion*, obviating thus the possibility for certain content, such as ideals, interests, desires, to become forces independent of us, taking our place in deciding what we want and what is right.

For the universalist thesis to have practical and moral relevance, it would have to carry with it a moral principle with substantive moral implications, rather than be a mere logical principle from which no moral consequence follows and is therefore useless. The logical strength of universalization is already contained in the demand of consistency, but if the logical strength of universalization is invested in the unconditionality and categorical nature of the moral or quasi-moral demand, it does not bear the possibility of a logical principle presenting itself *as* a moral one. Universalization, taken as mere logical generalizability, implies no contradiction of any practical judgment as such, because the question can always be posed *to what* or in relation to what is this judgment contradictory. For there to be a contradiction that obstructs universalization, we must have two judgments in conflict with one another. In general, this could also be said of moral rules. However, as a moral rule, the Golden Rule is only one general judgment that requires interpretation in application in order to be used. There should be the *possibility* of encountering an obstacle to universalization in the procedure of interpretation (such as in the case of testing the categorical imperative). This possibility is, however, completely empty, because the interpreted *content* cannot come into opposition with another judgment with which it is in relation, as this other judgment is itself, in both the positive version (“do unto others...”) and the negative (“do not do unto others...”), the formulation of the Golden Rule, the content of “that which” is strictly defined by the content of the terms “wish” and “do not wish.” It is the very subject of what is *wished* that is designated by the middle expression in the Golden Rule, of “that which.” Now, this expression does not truly appear twice in the Golden Rule, and it is no coincidence that it is so: it provides the appearance of the openness of the rule as a practical principle. It is enough for the Golden Rule to be only slightly reformulated to make this visible. This reformulated, synthesized form, the Golden Rule would be as follows:

That which you (do not) wish done to you, do (not do) *the same* to others.

Obviously, the phrase “that which” designates only that which is determined by the subject of the phrase “(do not) wish.” Meaning that, whatever is the subject of desire can, without obstacle, in this hypothetical sense, be universalized. It is enough for it to be *sincerely*, which is to say consistently, *truly* wished. The wish becomes the moral criterion. What is more, it becomes the source of conscience, at a cost that, is no trifle, and yet is insufficient to account for relativism and the absence of objectivity, and necessity of moral judgment. The cost can be described using the aforementioned term, *idealism*.

The cost of this idealism, and consequently the potential fanaticism that emerges from such *consistent sincerity*, is multiple and compounded, but so are the seeming gains: people act in accordance with what they (think they) wish and they are able to say that they are acting impartially and morally objectively. The *cost* of these ‘gains’ is abandoning a basic sense of the desire contained in the wish: even if I were someone else, and not myself, I would still want this, since I have universalized my wishes. Otherwise, I could wonder whether another, this other, perhaps wishes something other than what I wish, and then also wonder why I would not treat them in the way *they* wished to be treated, and not the way I wished they treated me, just as I *demand* that they do to me what I wish done to me, and not

have done what they wished I do to them. In this way, I have transformed the content of my wish into an ideal that presents itself as the ultimate criterion of value. Further, such hypostatization of the ideal, *free* me of responsibility for such acts, as they are not only in accordance with the ideal, but issue from it – the location of the production (creation) of *reason* for this act is not in me, but in the determination of the content of ideal that makes it ideal. Freeing of responsibility is also *relieving* of responsibility; by removing potential guilt, I also lose the potential for reward, and ultimately might not be able to call such an act *my own*, rather than mere mechanical following of a quasi-universal rule. Moral integrity is neither a reflection nor an expression of sovereignty of personhood of a reasoning being (a spiritual, intelligible being such as the individual); rather, it is a consequence deduced from the solidity of the *sincerity* and resoluteness of conscience based in consistent (obedient) following a heteronymous ideal to which everything is sacrificed, and which is not only valuable, but *infinitely* valuable, certainly *more valuable* than we ourselves are, and more valuable than anything else that could be desired and then wanted (for which decision to act or not act upon could make one responsible). That the ideal is unconditionally more valuable than the human as an individual results in the suspension of the absolute and irreducible value of the individual: universal distribution of the comparative “more valuable” introduces the category of *price*, ultimate comparability with other values and reduction to other values. This is what Kant refers to as “market price” [Kant, 1996, 91–93].

For its part, the “market” too can be diversely determined. The association that imposes itself with regard to this kind of price in the moral sense is the threat of hell and bribery of heaven in the Christian worldview. The “market” is our whole world in that case, like the valley of tears, while we, including our lives and personhoods, have a price; we are appropriately paid by a quantity of what is desired, or undesirable. It is no coincidence that in the canonic articulation of religious morals, the Golden Rule holds a special place, nor that it can be found in a more or less explicit form in the canonical texts of all religions⁹. After all, like other idealist value systems, the Golden Rule is quite handy for religions, since the content of permissible and desirable wishes is already strictly defined, irrespective of the degree of ‘sincerity’ or fanaticism, in accepting these ideas. Indeed, the greater the fanaticism, the easier and more straightforward the application of the Golden Rule. This only makes sense, given that the defined content is ever more clearly determined. It allows for the removal of all independent sources of wishes, while the content of ‘all’ wishes becomes determined and fixed. Yet it also has catastrophic consequences on moral integrity, since it is not the subject who “has wishes,” but rather the wishes (which is to say, their subject, the ideal, the heteronymous source of value) that, as it were, “have the subject.”

To recapitulate the problem of the Golden Rule, we can finally put in direct relation to the categorical imperative. What makes the relation of the Golden Rule and categorical imperative significant is that both principles lay claim to grounding morality, and both seemingly in universality. The possibility of grounding, basing, and establishing, morality is itself very important. Both principles claim to be one

⁹ For more on how the Golden Rule is articulated in most world religions, see: [Rost, 1986].

such grounding of morality, which does not depend on specific and transient moral understandings. Even the Bible¹⁰, as we have seen, considers the Golden Rule a general moral principle and test of acceptance of specific moral rules.

Nevertheless, what presents the perhaps greatest value of formalism in the categorical imperative is that it allows us to precisely establish that the categorical imperative *is not* a variation of the Golden Rule. The emphasis in the Golden Rule is on relationality, and not pure universalism, as is the case in the categorical imperative. The second, more important distinction, is that the determination of the practical principle as general instruction for acting is formal in the categorical imperative, while for the Golden Rule it is a matter of content. The Golden Rule, in its negative formulation, the one closer to the categorical imperative, states that one should not do unto others what one does not wish done unto oneself. But what is wished is always some content, and unless there is an additional limit, then in principle, the Golden Rule allows any action. And although this is an often-cited criticism of the categorical imperative also, it is, due to the imperative's formal determination, without grounds.

This can be clearly illustrated via a brief analysis of the phrase 'morally right' and the application of this analysis to the Golden Rule. It shows that these two procedures can *only* be differentiated in that one is right, or good, while the other not; in all other respects, they are completely equal. However, given that value statements in their *meaning* contain some *commitment* and that this is even the crucial part of that meaning, even the demand of consistency cannot be satisfied if two acts, otherwise overlapping in every relevant description, are valued differently (if one is lauded, while the other condemned). This allows persons in *the same* situation to act *differently*, yet still thinking that they are acting morally right. There is no act for which some version of the Golden Rule cannot be found to make that act morally correct and acceptable. An extreme example can be found without even sifting through philosophical literature: *Hitler* (cf.: [Hare, 1972]), for example, thought that it was entirely right to kill Jews, and he did so. To act differently would be against his *conscience* and would be a kind of betrayal of his deepest beliefs and ideals. Yet he can in all seriousness call upon the Golden Rule, on the condition that he *sincerely* believes in these ideals, meaning that, should he by some turn of events (say, by tracing his family lineage) find out that he was Jewish, that he also be liquidated. All it takes for one to refer to the Golden Rule is to call upon a maxim (any maxim!) expressed in a formulation that encompasses the rule. The only condition is that one *sincerely* accepts the maxim. It is *sincerity* that constitutes a given *conscience*, which one can refer to as sufficient basis of justification for one's actions. Any maxim can be a moral criterion on condition it represents *valid*, factually valid, personally and socially, position of valuation, as long as it is independent of the kind of valuation itself.

All these difficulties are overcome if morality is grounded *formally*. In Kant's ethics, the categorical imperative is a unique criterion for moral evaluation of human action, without extinguishing the possibility of autonomy and ascribability of freedom and responsibility, while moral values, such as justice or honesty, having the potential to achieve universality and objectivity preserving impartiality as

¹⁰ In particular if Lk. 6:32 is taken as the elaboration and grounding for Lk. 6:31, which states "And as ye would that men should do to you, do ye also to them likewise".

equality of application of the moral criterion *in different situations*. Whereas in the necessarily relativist interpretation of the Golden Rule we faced the situation of acting differently in similar situations (in the morally relevant sense) while believing to be acting morally right, in the case of the categorical imperative, we have a substantially altered standpoint: all situations are seemingly initially different, while their moral similarity is determined in the same act as their moral relevance – in the application of the categorical imperative. What is *measured* is infinitely different, but what is *used to measure* is always the same, in that it is one, but one only in that it is a measure. The results of the application of this measure to different situations will thus be different, which allows for what is ‘measured’ to be always and infinitely different, while the ‘measure’ in general relate to the whole world without reducing its differences to the content of a single dimension (or limited number thereof), nor to a description of any content – leaving the world, thus, open and (in general) infinite. Whatever the measured actually is, whatever content is measured (evaluated) morally, it is always measured according to only one and the same *condition*: that it is the subject of a real desire, that it is set as the goal of action, and that a real aim for its achievement exists. (In which case responsibility follows directly and evidently, as opposed to impossibility of any plausible responsibility for wishes which emerge of their own accord, without question or description, and which exist entirely independently from us.)

But this condition does not at all limit or determine the content of the desire, it merely connects with the representation of the object in the maxim of desire, relating it with what Kant calls *the will* (from which, through the freedom as a mental faculty, issue the maxims of action, the decision to act and the action itself). The very absence of such a limit is the expression of free will: no specific goal follows from this condition, nor is any possible goal therefore excluded. Further, no mutual relation is created between different goals due to this condition: only limits of given reality in the world in which those goals are set can make them mutually compatible or incompatible. There is no given ultimate goal. All goals all equally subject to evaluation exclusively according to whether they are subjects of some, if even only potential, desire, and not the degree of their achievement or their content. Their ultimate meaning is in their actualization, but the principle of this meaning logically precedes the actualization. As Kant put it, the metaphysics of morals cannot be grounded in anthropology but is rather applied to it.

Золотое правило и категорический императив

Йован Бабич – Профессор этики. Белградский университет. 11000, Џика Ljubina 18-20, Belgrad, Serbia.

ORCID: 0000-0002-7591-4902

e-mail: jovanpbabic@gmail.com

Золотое правило – «поступай с другими так, как ты хотел бы, чтобы они поступали с тобой», или в его негативной формулировке «не делай другим того, чего не хотел бы, чтобы делали тебе», является одной из самых древних формулировок моральных критериев. В той или иной форме мы встречаем его в канонических текстах практически всех религий, претендующих на какую-либо универсальность. На первый взгляд,

категорический императив утверждает и требует того же, что и Золотое правило. В связи с этим возникает вопрос: является ли категорический императив разновидностью Золотого правила или, наоборот, Золотое правило – это другой способ выражения того требования, что мораль должна быть одной и той же во всем мире? Другими словами, можно ли свести эти два «правила» друг к другу? Если это действительно так, то это был бы исключительный пример корреляции между практикой и теорией, тщательно разработанный в моральной философии Канта. Увы, если аргументация, приводимая далее в этом тексте, верна, она покажет, что категорический императив не является разновидностью Золотого правила, равно как и Золотое правило не является популярной формой требования, выдвигаемого категорическим императивом. Более того, будет показано, что Золотое правило вообще не представляет собой деонтологическую позицию и что оно не способно защитить от произвола, не способно быть критерием особого рода оценки, которую мы называем моральной, или выражать кардинальное требование универсальной объективности и беспристрастности. В отличие от Золотого правила, категорический императив, благодаря его ценному свойству, которое Кант обозначил как «формализм», как раз представляет такую позицию.

Ключевые слова: Золотое правило, позитивная и негативная формулировки ЗП, категорический императив, Иммануил Кант, формализм категорического императива, универсализация, беспристрастность

Литература / References

The Bible: Genesis; Gospel of Matthew; Gospel of Luke; First Epistle to Corinthians; Second Epistle to Corinthians.

Bruelisauer, B. "Die goldene Regel. Analyse einer dem Kategorischen Imperativ verwandten Grundnorm", *Kant-Studien*, 1980, Vol. 71, No. 3, S. 325–345.

Brandt, R.B. *Ethical Theory*. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1959.

Gewirth, A. *Reason and Morality*. Chicago: The University of Chicago Press, 1978.

Hare, R.M. *Freedom and Reason*. Oxford: Oxford UP, 1963.

Hare, R.M. "Peace", *Applications of Moral Philosophy*. London: Macmillan, 1972, pp. 48–66.

Hartmann, N. *Ethik*. Berlin & Leipzig: Walter de Gruyter & Co., 1935.

Kant, I. "Groundwork of the Metaphysics of Moral", in I. Kant, *Practical Philosophy*, ed. and trans. by M. Gregor. Cambridge: Cambridge UP, 1996, pp. 37–108.

Kant, I. "The Metaphysics of Morals", in I. Kant, *Practical Philosophy*, ed. and trans. by M. Gregor. Cambridge: Cambridge UP, 1996, pp. 353–604.

Rost, H.T.D. *The Golden Rule*. Oxford: George Ronald, 1986.

Sidgwick, H. *The Methods of Ethics*. Macmillan and Co., London, 1893.

Singer, M.G. *Generalization in Ethics*. London: Eyre and Spottswode, 1963.

Smart, R.N., "Negative Utilitarianism", *Mind*, 1958, No. 67, pp. 542–543.

Tarassenko-Struc, A. "Interpersonal Invisibility and the Recognition of Other Persons", *Explorations in Ethics*, ed. by D. Kaspar. London: Palgrave Macmillan, 2020, pp. 219–242.