The Golden Rule as It Ought to Be

Abstract

The Golden Rule, most commonly expressed in the form "do to others what you would have them do to you", has attracted criticism for failing to provide practical guidance in case of moral disagreement and for being susceptible to irrational outcomes. I argue that the alleged limitations are not a defect but just what makes the Golden Rule an effective tool of socio-ontological transformation towards *ideal* agency.

Keywords

The Golden Rule, Metaethics, Social Ontology, Metanormative Realism, Religious Ethics.

The Golden Rule and the Categorical Imperative

The Golden Rule was historically formulated in a number of ways: "In everything, do to others what you would have them do to you..." Matthew 7:12; "Treat others as you treat yourself..." Mahābhārata Shānti-Parva 167:9; "Love your neighbour as yourself..." Leviticus 19:18. Despite being widely regarded as self-evident, the rule has attracted scholarly criticism for failing to provide practical guidance in case of moral disagreement and for being allegedly open to irrational outcomes (Wattles 1996, 6-8). Immanuel Kant, the most prominent critic of the Golden Rule, has proposed an alternative principle, known as the Categorical Imperative, aiming to overcome its perceived deficiencies: "Act only according to that maxim whereby you can, at the same time, will that it should become a universal law." (Kant 1998, 4:421) The underlying idea that a universal law can be individually determined seems radically at odds with the Biblical context of the Golden Rule, which prescribes that individuals do not have moral authority over others (Matthew 7:1, Luke 6:37, James 4:11). S. B. Thomas (1970, 199) nevertheless argues that "the Categorical Imperative and the Golden Rule are two sides of the same coin, so to speak, the former providing a clarification of the rational scope of the latter, and the latter providing the spiritual basis for the correct application of the former." I will defend this hypothesis, albeit in support of a different interpretation of the Golden Rule. I argue that the alleged metaethical limitations of the rule are not defects but make it uniquely effective as an immediately intelligible transformative tool geared to socio-ontological grounding: it expresses social-reflexivity as a constitutive feature of rational agency.

The Categorical Imperative was an attempt not so much to re-interpret the Golden Rule but to transition from this circumstantial view of morality, conditional on personal inclinations, feelings or self-interest, to Metaphysics of Morality grounded in "unconditional good" (Kant 4:399-401, 4:414). Kant reasoned that the Golden Rule fails in its alleged purpose because a consistent moral formula cannot be subject to contingent personal inclinations but, rather, our inclinations must be subordinated to universal moral norms. Vis-a-vis the Golden Rule, the principle of universality implies that consistent moral evaluation cannot be just about my interest vs. the interests of people I happen to be directly interacting with but about the interests of all people simultaneously. This is a crucial consideration since no action happens in social vacuum and what we do (or fail to do) with respect to one person may have profound secondary consequences for others (Reinikainen 2005, 160). How should I relate to a unique state of another if their constitution is different from mine? I cannot imagine what it would be like to be someone else insofar as we are constituted differently but only insofar as we are the same, perhaps only insofar as we are conscious, rational agents in different circumstances. I may still be unable to relate to the relevant circumstances because I do not experience them in the same way as someone who is constitutively different. What does the Golden Rule or the Categorical Imperative amount to in the face of such asymmetries?

Kant attempted to solve the problem of asymmetry in the operation of the universal law by subordinating the moral criteria of action to what all rational agents have in common: "Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end." (Kant, 4:429) The central idea behind this move was to ground

morality in what all rational agents value about themselves: the uniquely human capacity to bestow *worth* according to our will.[1] Humanity for Kant (4:389) means just conscious, rational agency: his moral formula therefore implies that we are rationally committed not to infringe on the rational agency of others at least insofar as they themselves do not infringe on the rational agency of anyone else.[2]

Notwithstanding the fact that all humans share the property of being conscious, rational agents, we may still be in radical disagreement about beliefs and values, and this kind of difference is arguably the biggest obstacle to consistent application of any moral or metaethical principle, including the Categorical Imperative. Given that individuals X and Y are alike insofar as they are both conscious, rational agents who value their own agency, how should X apply the Categorical Imperative with respect to Y (and vice versa) if Y's idea of what counts for an infringement of rational agency is contrary to X's idea of the same? In such cases the contrary values are not included in the domain of common properties that the Categorical Imperative could attach to, it is therefore imperative to consider how different values could affect what X and Y have fundamentally in common. Another way, I can rationally treat others 'as myself' or with the same consideration as I have for myself only insofar as we are the same in the most relevant respect, because it would be nonsensical for different things to be considered the same in any respect that they are in fact different. This is a question of social significance, for example, in the debate about moral permissibility of abortion: are zygotes, embryos and foetuses relevantly the same as newborns, infants, toddlers, teenagers and adult humans? That which is different can be made common only in terms of properties that are held in common, so if our conflicting values motivate actions that affect those values which we hold in common then it is only in terms of those common values that actions can be consistently judged. But even perfect agreement about values will not necessarily lead to consistent application of the Categorical Imperative.

There is a quasi-paradoxical possibility of multiple agents being ostensibly committed to the Categorical Imperative and to common values while being implicitly in violation of the Categorical Imperative: I call it the case of 'virtuous tyrants'. If X and Y both agree on universalising the virtue of domination as the ultimate expression of rational agency, so that they respect one another's efforts to enslave or kill the other, then the rule is violated not by the absence of commonality or reciprocity of values but by the common inconsistency between explicit and implicit value-commitments. Specifically, the professed common commitment to the 'virtue' of domination contradicts the common commitment to value one's own agential capacities implied by the shared preference for acting in a particular way. Another way, if X values Y's aim of enslaving or killing X, then X is committed against its value-commitment to X's own agential capacities. The Categorical Imperative thus stands or falls not just on the basis of universality and reciprocity, but also on the basis of logical consistency of our explicit and implicit value-commitments; reciprocal inconsistency of values precludes consistent reciprocity of value. According to Kant, "Some actions are so constituted that their maxim cannot even be *thought* without contradiction as a universal law of nature, far less could one *will* that it *should* become such." (Kant 4:424) The Kantian formula, despite sophisticated *a priori* grounding, presents us with perhaps an insurmountable practical challenge, the answer to which is not contained within the formula and must be consistently worked out by every agent in his or her own capacity, under changing circumstances, rationally, while putting aside any personal biases.

Whereas the Kantian formula is too intellectually demanding to consistently negotiate in practice, the Golden Rule seems to fall short of the Categorical Imperative with respect to grounding: a mere assertion that something like the Golden Rule is universally imperative does not of itself constitute an authoritative argument in favour of the rule. It may be further objected that a rule which is not grounded in objective facts or an *a priori* argument is not a rule at all but an empty assertion. Without such grounding, interpretations of the Golden Rule may vary widely and even reach contradictory outcomes. Historical variants of the rule do not explicitly identify who counts as 'your neighbour' or as 'others', potentially allowing for unlimited exclusions that are often found within the same textual sources (Leviticus 20:9-13; Deuteronomy 13:6-11). The rule is also deemed compatible with (or even reducible to) 'an eye for an eye'

[1] (Markovits 2014, 103); "The value of humanity itself is implicit in every human choice." (Korsgaard 1996, 3.4.8)

[2] "Kantian internalism's central claim is that we behave irrationally when we fail to recognize others like us as our equals, in the sense that their goals and needs matter as much, objectively, as ours do. The exercise of our procedural rationality involves us in the task of examining our own ends in a manner that does not dismiss those of others." (Markovits 2014, 196)

(Leviticus 24:19-20). A further theoretic step is therefore required to show that the alleged rule has objective normative force, that is, for a particular range of conditions there must be consistent, universally applicable reasons that count in favour of adherence to the rule that outweigh any reasons to reject it.

Social Reflexivity as the Ontological Ground of Agency

Modern philosophers have ascribed many allegedly 'implicit' features to the Golden Rule which may be untrue. Specifically, the Golden Rule says nothing about circumstances or differences between persons, nothing about sympathy, universality, equality, the greater good, or the interests of others. On a literal reading, the rule seems to refer only to actions that would be desirable irrespective of situation, asking us to commit to absolute values rather than to case by case situational evaluation. For example, if I do not want to be killed then I ought not to kill, ever, under any circumstances. Alternatively, if I want my life preserved by others then I ought to preserve the lives of others. This is consistent with the religious idea of Sin as something absolute, independent of context: "Thou shalt not kill", period. The rule does not tell what specific actions are wrong or what we ought to desire, but the moral content could emerge heuristically, as a cumulative effect of our subjective wants and actions being confronted with the subjective wants and actions of others under various circumstances. The quality of *what we want* could thus progressively shape *what we ought to want*: a shared moral intuition geared to self-interest among beings of the same ontological kind. Without presupposing the ontological symmetry of person-to-person relations the notion of morality would be nonsensical. The Golden Rule therefore expresses the relational grounding of moral content, but I will attempt to show that the relevant grounding is more fundamental.

According to S. B. Thomas, focussing on the Golden Rule in the context of Christian theology, "the awareness of himself as a Man-Type, rather than a man-particular, must have no doubt dominated Jesus' view of himself, and it may be pointed to as precisely that aspect of his self-awareness which makes him an instantiation of the Kantian imperative." (1970, 195-6) A perfectly rational application of the Categorical Imperative is, on this view, just how an ideal agent would apply the Golden Rule: as the ontological essence of Man-kind. "That which is prohibited by the Categorical Imperative, at the level of rational morality, is prohibited by Jesus' very being, at the existential or religious level. Anyone who partakes of his being, by finding in his projected in-dwelling in Him the sense of his own identity as Universal Man, will share this existentialized Categorical consciousness with Him also." (Thomas 1970, 196) Kant approximates this mode of universality only in the final formulation of the Imperative, declaring Humanity (understood as conscious rational agency) as the value-commitment underpinning all action.

The proposed socio-ontological interpretation of the Golden Rule can be formally substantiated. Thomas Nagel argued that "to recognize others fully as persons requires a conception of oneself as identical with a particular, impersonally specifiable inhabitant of the world, among others of a similar nature." (Nagel 1970, 100) This works also in reverse: for an organism to have "conscious experience at all means, basically, that there is something it is like to be that organism." (Nagel 1974, 436) Crucially, the core phenomenological question of 'what it is like to be me' exemplifies a fundamental property of self-consciousness and cannot be meaningfully answered just in terms of 'me', as 'I *am* me' or 'I am *like* me', without falling prey to circular reasoning or triviality: "[Self-identity] is certainly a relation formally or logically speaking, but it also holds trivially, it's trivially true of everything…" (Strawson 2013) Like a finger that cannot point at itself, human agency is not ontologically self-sufficient but is constituted in terms of identity relations with other beings of the same kind. It follows that I can be *myself* only indirectly, socially, by consistently identifying *with* what I identify others *as*, insofar as others identify reciprocally.[3]

In light of the above insight we can recast Nagel's earlier formula as an explicitly ontological principle: to exist as a person requires a conception of oneself as identical with a particular, impersonally specifiable inhabitant of the world, among others of the same nature. Nagel has thus implicitly demonstrated that reciprocal recognition is an ontological foundation of conscious identity and therefore of agency; an idea which may go back to the ancient (albeit uncertain) etymological roots of the word *ánthrōpos* ('man'): one

[3] Cf. "The individual self will only emerge through the course of social externalization, and can only be stabilized within the network of undamaged relations of mutual recognition." (Habermas 2003, 34)

who is alike, the likeness of man. This socio-ontological thesis is rigorously developed in [Redacted] https://doi.org/10.1007/s11406-019-00149-6. In summary, personhood is not ontologically self-sufficient; nothing can maintain a meaningful identity just in term of itself. To identify as *x* therefore presupposes another, non-identical individual *y* having the same relation to *x* as *x* has to *y*. This is possible only if *x* and *y* are identical in the restricted sense of belonging to the same kind. More formally, *x* exists as a subject of **f**-kind only if *x* is an individual that relates to itself by relating to a different individual *y* that relates to itself by relating to *x*, in terms of properties (**f**) common to them both.

> "The degree of existence of the subject $\mathbf{I}_{\mathbf{f}}\{x\}$ is maximal iff *x* is an individual that relates to itself by relating to all other individuals *n* that relate to themselves by relating to *x*, in terms of properties (**f**) common to them all." (Ibid.)

Whereas some aspects of personal identity are contingent, others are intrinsic to rational agency. If my actions or attitude would negate any intrinsic property or the moral status of any other conscious agent who is therefore (intrinsically) an agent like me, I would be undermining my own ontological integrity and, implicitly, my conscious existence and moral status as an agent. Another way, whatever intrinsic property I value about myself I am rationally committed to value about other beings of the same ontological kind, or I stand to negate it about myself.

If this is correct then the Golden Rule is not just a cultural or religious artefact but an expression of a universal law: social reflexivity grounds individual consciousness and is therefore objectively binding, insofar as we value our own existence as conscious rational agents. Moreover, if the degree of existence as a conscious rational agent is conditional on the consistency of reflexive relations with others, then the rule is also a means of transformation towards ideal agency (approximating the likeness of God)[4], or, in the negative, a pathway of devolution from Man-kind towards unconscious animal existence. This socio-ontological condition of personal consciousness is plausibly crystallised in the first Biblical formulation of the Rule, "Love your neighbour as yourself" (Leviticus 19:18), and fully universalised over conscious agents in Jesus' pronouncement "Love your enemies" (Matthew 5:44; Luke 6:27).

From Metanormative to Practical Considerations

Some philosophers have pointed out that "To accept the Golden Rule ... is not to adopt a specific moral code but simply to adopt the moral point of view..." (Blackstone 1965, 174) I suggest that even this distinction fails to capture the essence of the Golden Rule while already claiming too much. It is not evident that the Golden Rule entails a moral point of view; I have shown above that it may be a metanormative principle grounded in ontology, a practical expression of the foundation upon which we realise our agency and self-interest. It does not of itself resolve any specific moral dilemmas but posits a metanormative framework of action based on the premise that preservation of our interests depends in a fundamental way on preservation of the interests for others. Another way, it does not prescribe how to be rational about ethics in particular circumstances but only radically limits the capacity to be irrational about our intentions, which may in turn guide social evolution towards ideal agency.

Wattles (1996, 166) writes that "The Golden Rule cannot be the supreme principle of morality in the sense of functioning as the sole normative axiom in a deductive system of ethics, because it cannot operate in a value vacuum." I agree insofar as that the Golden Rules is not a moral principle of morality but, I contend, it does not entail a value vacuum. Irrespective of our subjective, agential preferences the rule entails a practical commitment to social reflexivity and, indirectly, to human agency. The value of human agency is implicitly affirmed in every action; it is the source of every contingent value commitment. This universal source of value is conditional on social reflexivity, which is therefore the objective normative standard for all agents. In addition to the socio-ontological considerations already discussed, the basic formulation of the rule also expresses a profound psychological insight; it asks us to consider before we act how we want to be treated by others. The moment of reflexive contemplation that the rule is asking us to commit to has the capacity to mitigate impulsive, emotionally-driven actions, but it is also transformative; a bonding experience able to progressively harmonise social relations. "The

[4] This transcendental feature is explicitly formulated in 2 Corinthians (3:18): "And we all, with unveiled face, beholding the glory of the Lord, are being transformed into the same image from one degree of glory to another."

practice of the golden rule may be advocated precisely as a heuristic device to help the agent to become aware of the kinship of humankind." (Wattles 1996, 180) By reflecting on the Golden Rule we become aware how our actions and attitudes can lead to responses that generate value. Based on this information we may adjust our standard of behaviour in order to reliably satisfy our intrinsic value-commitments by respecting the same value-commitments in others. In this way moral intuition is socially evolved. I suspect that this was the main reason for the Golden Rule to find expression in most religions and cultures (Wattles 1996, 4). It was arguably never intended as an infallible moral formula - "a criterion of the moral rightness of interpersonal actions" (Gewirth 1978, 133) - but as a transformative tool for rationally imperfect beings to reconcile the common need for cooperation with their competing subjective claims on limited resources.

There are some important current criticisms of the Golden Rule. "The rule, it has been charged, cultivates blindness to the otherness of the other, since it assumes a basic commonality between agent and recipient. Some challenge the notion of a common humanity, citing the pervasive influence of differences such as gender, race, and class, and the uniqueness of individual personality." (Wattles 1996, 174) I retort that without the presumption of a common humanity (as conscious rational agency) we would not be of the human-kind, we would not exist as conscious rational agents and there would be no possibility of mutual understanding. Our ontological commonality underpins the capacity for individual uniqueness, which is therefore subordinate to this higher normative principle. In essence, our likeness-to-kind does not negate individual uniqueness but, rather, makes it meaningful.

When our personal identity is constituted in terms of likeness-to-kind we are committed to value certain relations more than others. "Every beast consorts with its own kind, and shall not man with his fellow? Like to like is nature's rule, and for man like to like is still the best partnership." (Sirach 13:19-20) Our relations with family and friends are naturally more deeply integrated, more reflexive than our relations with strangers on account of a more intimate phenomenological affinity, a more perfect likeness, and this may explain different degrees of ethical commitments, but we are nonetheless ontologically and therefore normatively bound in a similar way to all conscious rational agents of the human kind, which is *our* kind. The Golden Rule is transformative; its diligent application teaches us how to integrate our consciousness. Conversely, to violate the rule is to fracture the socio-ontological ground of our personhood and thus metaphysically dis-integrate, become less real to oneself.

In conclusion, many philosophers attempted to improve the Golden Rule by effectively replacing it with the form 'do onto others what is objectively right to do, and be perfectly rational about it' (Gewirth 1978, 141). For the rule to do what some philosophers presumed it ought to do we would have to be sufficiently informed about social ontology, normativity and perfectly rational about our intentions... but we are not: "We do not walk around with a set of definite desires about how we want to be treated in various types of situations." (Wattles 1996, 166) Moreover, any attempt to detach the rule from irrational intentions would negate its positive psychological effect precisely where rational normative evaluation has already failed. The rule dogmatically invites us to apply a fundamental metanormative principle while simultaneously creating a space for reflexive kinship which in turn mitigates our impulsive, emotionally-driven responses. Unlike the Categorical Imperative which rationalises the socio-ontological grounding of conscious agency, the rule works directly on the socio-ontological level, beneath rationality, and can therefore be transformative irrespective of our vague or inchoate desires. The ostensibly deficient Golden Rule may work in practice whereas analytically demanding moral formulas work only in theory.

References:

Blackstone, W. T. "The Golden Rule: A Defense." *Southern Journal of Philosophy*, 1965: 172-177.

Gewirth, Alan. "The Golden Rule Rationalized." *Midwest Studies in Philosophy*, 1978: 133-147.

Habermas, Jürgen. *The Future of Human Nature.* Cambridge: Polity Press, 2003.

Kant, Immanuel. *Groundwork of the Metaphysics of Morals.* Cambridge: Cambridge University Press, 1998.

Korsgaard, Christine M. *The Sources of Normativity.* Cambridge: Cambridge University Press, 1996.

[Reference Redacted for Anonymous Review] https://doi.org/10.1007/s11406-019-00149-6.

Markovits, Julia. *Moral Reason.* Oxford: Oxford University Press, 2014.

Nagel, Thomas. *The Possibility of Altruism.* Oxford: Clarendon Press, 1970.

Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review*, 1974: 435-450.

Reinikainen, J. "The Golden Rule and the Requirement of Universalizability." *Journal of Value Inquiry*, 2005: 155-168.

Strawson, Galen. "'Self-intimation'." *Phenomenology and the Cognitive Sciences*, 2013: 1-31.

Thomas, S. B. "Jesus and Kant." *Mind*, 1970: 188-199.

Wattles, Jeffrey. *The Golden Rule.* Oxford: Oxford University Press, 1996.