

Golden Rule

Thomas Carson

Roughly, the golden rule says that we must treat others as we would be willing to have them treat us or, alternatively, that we must not treat others in ways in which we are unwilling to be treated ourselves. The golden rule entered the Judeo-Christian tradition from ancient Mesopotamia. The Mesopotamian tradition includes the following injunction: “Son, that which seems evil to thee, do not to thy companion.” Shortly before the birth of Christ, Rabbi Hillel said: “That which is hateful to you, do not do to your neighbor: that is the whole of the Torah, while the rest is commentary thereon; go and learn it.” By “neighbor,” Hillel meant all fellow human beings (Wattles 1996: 49–50). Jesus gives two statements of the golden rule in the New Testament: “And as you want that people do to you do thus to them” (Luke 6:31); “All things therefore which you will that people do to you, do thus to them for this is the law and the prophets” (Matthew 7:12). The following statement is attributed to Muhammed: “That which you want for yourself, seek for mankind.” Confucius (who predates Hillel, Christ, and Muhammed) said: “Do not impose on others what you do not desire others to impose on you.” (See Wattles 1996: 15–67 for these references.) One of the earliest statements of the golden rule is found in the Hindu Upanishads (800–600 BCE): “Let no man do to another that which would be repugnant to himself; this is the sum of righteousness; the rest is according to inclination” (Hertzler 1934: 420). Hertzler (420–1) also attributes versions of the golden rule to Zoroastrianism, the Buddha, and Laozi.

Several well-known objections to the golden rule are widely taken to be decisive. The first objection is that the golden rule prohibits many acts that are morally right. Sometimes it is right to do things to others that they do not want us to do (and, thus, sometimes it is right to do things to others that we would not want done to us were we in their position, with their wants and desires). For example, sometimes it is morally right for judges and jurors to punish murderers. However, judges and jurors who punish murderers violate the golden rule, because, if they were in the murderer’s place and had the murderer’s desires, they would not be willing to be punished. Kant regards this as a fatal objection to the golden rule: “the criminal would on this ground be able to dispute with the judges who punish him” (Kant 1993: 27; *see* KANT, IMMANUEL). Second, the golden rule permits people with unusual preferences to perform wrong acts. For example, it permits masochists to inflict pain on nonmasochists, because the masochists are willing to have others inflict pain on them.

R. M. Hare gives the most important defense of the golden rule in recent philosophy. Hare’s defense is based on his theory about the meaning of moral concepts. He claims that moral judgments are universalizable prescriptions that are overriding (Hare 1981: 55; *see* HARE, R. M.). He holds that moral judgments are prescriptions or commands.

International Encyclopedia of Ethics, Edited by Hugh LaFollette.

© 2022 John Wiley & Sons Ltd. Published 2022 by John Wiley & Sons Ltd.

DOI: 10.1002/9781444367072.wbiee188.pub2

“Lying is wrong” entails the prescription/command “Don’t lie.” To say that moral judgments are universalizable means that, if one makes a moral judgment, one is committed to making the same moral judgment about any similar case, unless there is a morally relevant difference between the cases. To say that moral judgments are overriding means that a person who makes moral judgments takes the prescriptions expressed by them to override (to be more important than) all other conflicting considerations, such as considerations of prudence, etiquette, and the law. According to Hare, statements of the form “It is morally wrong, all things considered, for you to do X, but, nevertheless, you would be justified in doing X” are self-contradictory. He also thinks that it is inconsistent to say that it is morally wrong (on balance) for you to do X but still command or advise you to do X (*see* MORAL JUDGMENT; OVERRIDINGNESS, MORAL; PRESCRIPTIVISM; UNIVERSALIZABILITY).

Hare holds that we must be able to give reasons for the moral judgments we make. If I say that an action is right or wrong, I must be able to point to some (nonmoral) feature of the action in virtue of which it is right or wrong. The answers/reasons we give to questions about why actions are right or wrong commit us to general moral principles. For example, if I say that an action is wrong because it involves breaking a promise, then I am committed to the principle “It is wrong to break promises.” Hare writes:

When we make a moral judgment about something, we make it *because* of the possession by it of certain non-moral properties . . . moral judgments about particular things are made for reasons; and the notion of a reason, as always, brings with it the notion of a rule which lays down that something is a reason for something else. (Hare 1963: 21)

We can test and criticize moral principles by imagining hypothetical cases and asking whether we are willing to endorse the judgments and prescriptions they commit us to. Few of us are willing to accept the consequences of the view that it is *always* wrong to break a promise, since it commits us to the view that it would be wrong to break a promise to meet someone for a brief coffee break in order to render life-saving first aid in an emergency. Therefore, we will try to come up with better principles, for example: “It’s wrong to break a promise, unless doing so is necessary in order to bring about very good consequences or to prevent very bad consequences.” Hare says: “What we are doing in moral reasoning is to look for moral judgments and moral principles which, when we have considered their logical consequences, and facts of the case, we can still accept” (1963: 88; *see* MORAL REASONING).

One very important test of moral principles is whether we are willing to accept their implications (and sincerely endorse the prescriptions they commit us to) for hypothetical situations in which we switch places with other people. If I judge that it is permissible for me to do something to you, I am committed to saying that it would be permissible for someone else to do the same thing to me if I were in a relevantly similar situation. According to Hare’s test, my doing X to you and saying that it is morally permissible for me to do this commits me to not objecting to you or others doing X to me if I am in a hypothetical situation relevantly similar to yours. Hare’s “switching places” test is tantamount to a version of the golden rule. Suppose that a

dishonest plumber claims that it is morally permissible (or even obligatory) for him to defraud his customers and bill them for unneeded repairs that cost them thousands of dollars. To be consistent, he must say that it would be morally permissible (obligatory) for others to defraud him and those he cares about in relevantly similar (hypothetical and/or actual) cases. Since moral judgments are prescriptive, he is committed to not objecting to others defrauding him or those he loves in relevantly similar circumstances. A normal person who is concerned for her own welfare and the welfare of those she loves cannot do this.

The idea of morally relevant differences is extremely important for the purposes of Hare's theory. Often people disagree about whether or not certain differences are morally relevant. For example, people disagree about whether there is a morally relevant difference between killing someone and failing to save someone's life, and about whether there is a morally relevant difference between abortion and infanticide. Some have concluded that the universalizability principle and Hare's version of the golden rule are useless as principles of moral reasoning, because any application of them presupposes certain views about what sorts of differences are and are not morally relevant and, in turn, such claims presuppose answers to disputed moral questions.

Hare attempts to answer this objection by showing that one can effectively use the universalizability principle in moral arguments without making any *controversial* assumptions about which kinds of differences are morally relevant and which are not. He says that the only assumption one needs to make is that one's own personal place in the universe is not morally relevant. If two cases differ only in that two individual people reverse their positions and all of the universal properties remain the same, then there is no morally relevant difference between them. Hare writes:

Our present argument has no need of a definition of universalizability in terms of relevant similarity. In this book we shall be appealing . . . only to *exact similarity*, and shall not need therefore to say, before the argument begins, what is and what is not relevantly similar . . . therefore we can use hypothetical cases exactly similar in their universal properties, and differing only in the roles played by individuals. (Hare 1981: 62–3, my emphasis; see also Hare 1963: 107)

Here is a variation of an argument that Hare gives in *Freedom and Reason* (1963). Suppose that someone says that it is permissible for him to own you as a slave. You ask if it would be permissible for you to own him as a slave, and he says “no.” You accuse him of inconsistency, and he replies that there is a morally relevant difference between the two cases; namely, that he is European and you are African. How can we show that this is not a morally relevant difference without presupposing the falsity of the man's moral judgment, in which case the consistency argument is useless for establishing the falsity of his moral judgment? If the man imagines a hypothetical case in which he is an African and you are a European, he will find that he objects to your owning him as a slave. This thought experiment about a hypothetical case reveals the inconsistency of his view.

Hare has greatly influenced recent discussions of the golden rule. Gensler and Carson defend consistency versions of the golden rule. Carson (2018: 152–3) offers

the following proof of the golden rule (GR; this proof is a modified version of arguments given earlier in Gensler 1986, 1996, 1998; also see Carson 2010: 131):

1. Consistency requires that if you think that it would be morally right for someone to do a certain act to another person, then you must grant that it would be morally right for someone to do that same act to you in relevantly similar circumstances.
2. Consistency requires that if you think that it would be morally right for someone to do a certain act to you in certain circumstances, then you must not object to him/her doing that same act to you in those circumstances. (To object to an action is to claim to be justified in feeling anger toward the agent because of her action and thinking that the agent ought to feel guilt because of it. If I object to an action that you perform, I am claiming that I am *justified in strongly disapproving* of you for doing what you did.)

Therefore,

GR. If you think that it would be morally right for someone to do a certain act to another person, then you are inconsistent if you object to someone doing the same act to you in relevantly similar circumstances.

(For an assessment of an earlier version of this argument in Carson 2010: 131, see Gensler 2013: 195–8.)

The argument is valid: the conclusion follows from the premises. Both premises are consistency requirements. Premise 1 addresses questions about the consistency of a person's different moral beliefs. Premise 2 addresses questions about whether a person's moral beliefs are consistent with her attitudes and actions.

Premise 1 follows from the universalizability principle. If it is right or wrong for someone to do something to another person, then it is right or wrong for someone to do the same thing to *me* in relevantly similar circumstances. Consistency requires me to judge acts done to others in the same way in which I judge acts done to myself. Premise 2 says that my attitudes must be consistent with my moral judgments. If I say that it is morally permissible for you to do something to me, I cannot object to you doing it. If I think that it is permissible for you to beat me at chess, then I cannot object to your beating me at chess. (I don't have to allow you to beat me at chess or to want you to beat me, but, in order to be consistent, I cannot claim to be justified in strongly disapproving of you for beating me.)

The force and power of this version of the golden rule derive from the fact that, since we *do object* to other people doing certain things to us (or to our loved ones), we cannot consistently say that it is morally permissible for anyone to do these things to others. For example, suppose that I am a salesperson for a pest control company. I try to manipulate customers into signing up to a very expensive course of treatment. I lie and tell them that they have "carpenter ants" in their homes. I also greatly exaggerate the harm that carpenter ants can cause. This costs people lots of money and exposes them to poisons in their homes. I claim that it is morally right for me to do this. But I am inconsistent, because I object to members of other professions lying to me to manipulate me whenever doing so is to their advantage. I very much object to my physician or lawyer or accountant or car mechanic doing something like this to me or to my loved ones.

The Hare–Gensler–Carson version of the golden rule (GR) can answer the two objections to the golden rule mentioned earlier. A masochist who inflicts pain on non-masochists violates GR, because he objects to this being done to him *in relevantly similar circumstances*. A relevantly similar case would be one in which he is a nonmasochist who finds having pain inflicted on him very distressing and objects to others inflicting pain on him. (The people he is causing to feel pain are nonmasochists who object to having pain inflicted on them.) A masochist who adequately imagines the position of a nonmasochist having pain inflicted on him against his will objects to having pain inflicted on him in cases in which he is in a relevantly similar position. (For more on this point, see Gensler 1996: 99; 1998: 110–12; Carson 2010: 137–8).

GR avoids Kant's objection about the moral permissibility of punishing criminals against their will. Kant's objection assumes that the golden rule commits us to the following:

[A] If I claim that someone's doing act X to person S is morally permissible, then, on pain of inconsistency, it must be the case that I would not object to someone doing X to me were I in S's position with S's desires.

Many traditional versions of the golden rule say or imply A, so Kant's objection is a fair objection to many traditional versions of the golden rule. However, GR does not imply A; it only implies the following:

[B] If I claim that someone's doing act X to person S is morally permissible, then, on pain of inconsistency, I must *now* not object to someone's doing X to me in a relevantly similar situation (among other things, I must now not object to someone's doing X to me in the hypothetical situation in which S and I switch places and I have all of S's properties).

Hare clearly accepts B and rejects A (see Hare 1963: 108). The difference between A and B depends on the difference between (a) *the preferences I would have in a hypothetical situation* in which I switch places with another person and actually *have* her preferences and aversions; and (b) *my present preferences for a hypothetical situation* in which I switch places with another person and acquire her wants and aversions. Suppose that I take a mind-altering drug that causes me to become so severely depressed that I want to kill myself. I *now* prefer that, if I were to take such a drug and become suicidal, others should forcibly intervene to prevent me from taking my own life. However, if I were determined to commit suicide, I would strongly prefer that others not try to stop me and I would strongly object to their doing so. B (but not A) is consistent with the view that it would be permissible for others to intervene to stop me from killing myself in such a case (see Carson 2010: 140; see Hare 1963: 115–17 for a discussion of the issue of punishment).

A and B also have very different implications for cases in which someone tries to act on immoral desires. Suppose that a Hutu member of the Interahamwe (the hate group that perpetrated the Rwandan Genocide in 1994) wants to murder his Tutsi neighbors. According to A, you cannot consistently hold that it would be right for

you to use force to stop him from killing his neighbors, because, if you were in his position and had the motivations of a fanatical member of the Interahamwe, then you would object to being forcibly restrained from killing. B, however, allows for the possibility that you are consistent in holding that it would be morally right for you to stop him from killing his neighbors. You might (now) be willing to be coerced should you acquire and try to act on immoral desires. B gives intuitively more plausible results than A for these cases and for a wide range of other cases.

My GR and Hare and Gensler's versions of the golden rule allow us to make consistency arguments *within morality*. They cannot justify morality or show that it is unreasonable to be amoral (see Hare 1963: 100; 1981: 184; Carson 2010: 148; for other sorts of arguments against being amoral, see Hare 1981: 190–6; Carson 2010: 148–50). To paraphrase Hare, these versions of the golden rule constrain the kinds of moral judgments we can make if we make any moral judgments at all (see Hare 1981: 6–7).

See also: HARE, R. M.; KANT, IMMANUEL; MORAL JUDGMENT; MORAL REASONING; OVERRIDINGNESS, MORAL; PRESCRIPTIVISM; UNIVERSALIZABILITY

REFERENCES

- Carson, Thomas 2010. *Lying and Deception: Theory and Practice*. Oxford: Oxford University Press.
- Carson, Thomas 2018. "The Range of Reasonable Views about the Morality of Lying," in Eliot Michelson and Andreas Stokke (eds.), *Lying Language, Knowledge, Ethics and Politics*. Oxford: Oxford University Press, pp. 145–60.
- Gensler, Harry 1986. "A Kantian Argument against Abortion," *Philosophical Studies*, vol. 49, pp. 83–98.
- Gensler, Harry 1996. *Formal Ethics*. London: Routledge.
- Gensler, Harry 1998. *Ethics*. London: Routledge.
- Gensler, Harry 2013. *Ethics and the Golden Rule*. London: Routledge.
- Hare, R. M. 1963. *Freedom and Reason*. Oxford: Oxford University Press.
- Hare, R. M. 1981. *Moral Thinking*. Oxford: Oxford University Press.
- Hertzler, J. O. 1934. "On Golden Rules," *Ethics*, vol. 44, pp. 418–36.
- Kant, Immanuel 1993. *Grounding for the Metaphysics of Morals*, trans. James Ellington, 3rd ed. Indianapolis, IN: Hackett.
- Wattles, Jeffrey 1996. *The Golden Rule*. New York: Oxford University Press.

FURTHER READINGS

- Allinson, R. E. 1992. "The Golden Rule as the Core Value in Confucianism and Christianity," *Asian Philosophy*, vol. 2, pp. 173–85.
- Cadoux, A. T. 1912. "The Implications of the Golden Rule," *Ethics*, vol. 3, pp. 272–87.
- Gewirth, Alan 1978. "The Golden Rule Rationalized," *Midwest Studies in Philosophy*, vol. 3, pp. 133–47.
- Hare, R. M. 1952. *The Language of Morals*. Oxford: Oxford University Press.

- Hoche, Hans-Ulrich 1978. "Die Goldene Regel," *Zeitschrift für Philosophische Forschung*, vol. 32, pp. 355–75.
- Locke, Don 1968. "The Trivializability of Universalizability," *Philosophical Review*, vol. 77, pp. 25–44.
- Rønnedal, Daniel 2015. "The Golden Rule and the Platinum Rule," *Journal of Value Inquiry*, vol. 49, pp. 221–36.
- Singer, Marcus 1961. *Generalization in Ethics*. New York: Knopf.
- Singer, Marcus 1963. "The Golden Rule," *Philosophy*, vol. 38, pp. 293–314.
- Weiss, Paul 1941. "The Golden Rule," *Journal of Philosophy*, vol. 3, pp. 421–30.

REVISED PROOFS