

ARTICLE

The Golden Rule: A Naturalistic Perspective

Nathan Cofnas 

Balliol College, Oxford, UK
E-mail: nathan.cofnas@balliol.ox.ac.uk

Abstract

A number of philosophers from Hobbes to Mill to Parfit have held some combination of the following views about the Golden Rule: (a) It is the cornerstone of morality across many if not all cultures. (b) It affirms the value of moral impartiality, and potentially the core idea of utilitarianism. (c) It is immune from evolutionary debunking, that is, there is no good naturalistic explanation for widespread acceptance of the Golden Rule, ergo the best explanation for its appearance in different traditions is that people have perceived the same non-natural moral truth. De Lazari-Radek and Singer employ all three of these claims in an argument meant to vindicate Sidgwick's 'principle of universal benevolence'. I argue that the Golden Rule is the cornerstone of morality only in Christianity, it does not advocate moral impartiality, and there is a naturalistic explanation for why versions of the Golden Rule appear in different traditions.

Keywords: Golden Rule; evolutionary debunking arguments; principle of universal benevolence; impartiality; utilitarianism

1. Introduction

The Golden Rule – *do unto others as you would have them do unto you* – appears in some form in all of the great religious and ethical traditions. Some Western scholars have proclaimed it to be the universal foundation of morality. Since the second century, Christian thinkers have pointed to it as the essence of the natural law (du Roy 2008; Gensler 2013: 79–80; Wattles 1996: 72) that is “written in [the] hearts” of everyone (Romans 2:15 NKJV). Hobbes (1651/1996: 14.5) describes the Golden Rule in its negative form (i.e., *do not do unto others what you do not want them to do unto you*) as “the law of all men.” Westermarck (1906: 103) avers that “St. Augustine was right in saying that ‘Do as though wouldst be done by’ is a sentence which all nations under heaven are agreed upon.” In recent times, people have compiled examples of the Golden Rule from different traditions to show that this principle is the basis of ethics across cultures (see Wattles 1996: 4).

Some philosophers see the Golden Rule as affirming the value of moral impartiality, and potentially the core idea of utilitarianism. Mill (1863/2015: 130) holds that Jesus's Golden Rule expresses “the complete spirit of the ethics of utility.” Westermarck (1906: 101–2) says that versions of the Golden Rule found across cultures all express the same “notion of disinterestedness.” De Lazari-Radek and Singer (2012; 2014) argue that the

Golden Rule is ‘similar’ to Sidgwick’s (1907) ‘principle of universal benevolence’ (PUB), which enjoins us to impartially maximize the good of all sentient beings.¹

Debunking arguments target a belief that p by showing that the causal process that gave rise to the belief does not track the truth about p (see, e.g., Kahane 2011; Nichols 2014; Sauer 2018). Debunking arguments do not establish that a targeted belief is false – only that it is *unjustified* in light of our discovery of its causal origins (Kahane 2011: 108; Sauer 2018: 29).

Suppose we construe our moral beliefs as being true in a non-natural realist sense – as being objectively correct according to some independently existing standard (Shafer-Landau 2003: 15). *Evolutionary* debunking arguments in ethics claim that our evaluative judgments are “saturated with evolutionary influence” (Street 2006: 114), that is, the influence of natural selection. Natural selection aims at fitness rather than objective moral truth. The (alleged) discovery that our moral beliefs (realistically construed) were shaped to a large extent by the morally indifferent force of natural selection renders them unjustified (see also Joyce 2006; Ruse & Wilson 1986).

Moral realists sometimes claim that belief in the Golden Rule has no good evolutionary explanation and therefore remains unscathed by evolutionary debunking arguments. Huemer (2005: 217) contrasts “moral beliefs about incest, special obligations to family, and reproductive practices,” which may have plausible evolutionary explanations, with acceptance of the Golden Rule, which does not. According to Parfit (2011: 536), the Golden Rule was “independently proclaimed and accepted in several of the world’s earliest civilizations.” He says that belief in the Golden Rule – like the belief that “everyone’s wellbeing matters equally” – is “clearly *not* the product of evolutionary forces” (ibid.: 538), so is immune from evolutionary debunking. According to de Lazari-Radek and Singer (2012; 2014), the fact that thinkers in different traditions agree that the Golden Rule is the ‘essence of morality’ has no plausible naturalistic (i.e., debunking) explanation and is evidence for the objective correctness of PUB.

De Lazari-Radek and Singer, who provide the most developed version of this argument, propose a three-step procedure for “establishing that an intuition has the highest possible degree of reliability.”² One must show that the following conditions hold:

1. careful reflection leading to a conviction of self-evidence;
2. independent agreement of other careful thinkers; and
3. the absence of a plausible explanation of the intuition as the outcome of an evolutionary or other non-truth-tracking process. (de Lazari-Radek & Singer 2012: 26; cf. de Lazari-Radek & Singer 2014: 195)

In regard to PUB, they summarize their argument as follows:

We form the intuition as a result of a process of careful reflection that leads us to take, as Sidgwick puts it, “the point of view of the universe.” This idea is not specific to any particular cultural or religious tradition. On the contrary, the leading

¹When PUB is combined with a utilitarian theory of the good (i.e., the good consists in well-being, or happiness, or something along those lines) it is equivalent to the principle of utility. Without a theory of what “the good” consists in, PUB is “empty of content” (de Lazari-Radek & Singer 2012: 27; see also Kahane 2014).

²Cf. Sidgwick’s (1907: 3.11.2) “[f]our conditions” that must be satisfied to show that a proposition has “the highest possible degree of certainty.”

thinkers of distinct traditions have independently reached a similar principle and have regarded it as the essence of morality. In addition to the well-known Jewish and Christian versions of the Golden Rule, we find similar ideas in the Confucian, Hindu, and Buddhist traditions. Admittedly, these rules do not require us to adopt universal benevolence, but they do require impartiality ... Finally, there is no plausible explanation of this principle as the direct outcome of an evolutionary process, nor is there any other obvious non-truth-tracking explanation. Like our ability to do higher mathematics, it can most plausibly be explained as the outcome of our capacity to reason. (de Lazari-Radek & Singer 2014: 193; see also de Lazari-Radek & Singer 2012: 25–26)

I argue that the appearance of the Golden Rule in different traditions does not reflect independent agreement among careful thinkers about a principle similar to PUB, moral impartiality, or the principle of utility. Furthermore, there is a plausible naturalistic explanation for why many moral teachers would independently promulgate Golden Rule-like principles. In the following section, I examine the Golden Rule as it is understood in Judaism, Christianity, Confucianism, Hinduism, and Buddhism. Versions of the Golden Rule appear in other traditions as well (Terry 2011), but I focus on the aforementioned five because they are the ones that de Lazari-Radek and Singer refer to. I argue that the Golden Rule is the ‘essence of morality’ only in Christianity. The Golden Rule as it is understood in any of these traditions is not ‘similar’ to PUB (or any related principle) in the ways required for de Lazari-Radek and Singer’s argument to work. Furthermore, thinkers who endorsed the Golden Rule did not purport to take the ‘point of view of the universe’.³ In the third section, I provide a naturalistic explanation for why versions of the Golden Rule have appeared in different traditions. In the fourth section, I offer a naturalistic explanation for why Mohists and Anglican utilitarians came to advocate impartial morality (though without appealing to the Golden Rule).

2. The Golden Rule across cultures

At first glance, rather than being something about which ‘careful thinkers’ tend to independently agree, moral principles like PUB and the principle of utility are associated with a small number of mostly anglophone philosophers in fairly recent history. But, as noted, de Lazari-Radek and Singer (2012: 25–26; 2014: 193) claim that “the leading thinkers of distinct traditions” – they mention Judaism, Christianity, Confucianism, Hinduism, and Buddhism – have taken the “point of view of the universe” and “independently reached a similar principle [in the form of the Golden Rule] and have regarded it as the essence of morality.”

Let’s begin with Christianity. Though Jesus does not comment explicitly on ‘morality’ (his concern is the religious law), the Gospels quote him stating that the Golden Rule is the essence of the religious commandments: “whatever you want men to do to you, do also to them, for this is the Law and the Prophets” (Matthew 7:12). In another passage, he agrees with a lawyer’s suggestion that one inherits eternal life by “lov[ing] the Lord your God with all your heart, with all your soul, with all your

³For a critical analysis of some other elements of de Lazari-Radek and Singer’s argument see Crisp (2012).

strength, and with all your mind’, and ‘your neighbor as yourself” (Luke 10:27). Paul expounds:

Owe no one anything except to love one another, for he who loves another has fulfilled the law. For the commandments ... are *all* summed up in this saying, namely, “You shall love your neighbor as yourself.” Love does no harm to a neighbor; therefore love *is* the fulfillment of the law. (Romans 13:8–10)

In the Sermon on the Mount, Jesus offers these instructions, which put the Golden Rule in context:

Love your enemies, do good to those who hate you, bless those who curse you, and pray for those who spitefully use you. To him who strikes you on the *one* cheek, offer the other also. And from him who takes away your cloak, do not withhold *your* tunic either. Give to everyone who asks of you. And from him who takes away your goods do not ask *them* back. And just as you want men to do to you, you also do to them likewise. (Luke 6:27–31)

There are a few things to notice about the Golden Rule as it was taught by Jesus. First, it is clearly not concerned with ‘sentient beings’ generally, but people specifically. Singer (1975/2002: 191) himself notes that “[t]he New Testament is completely lacking in any injunction against cruelty to animals, or any recommendation to consider their interests.” There is no evidence that Jesus or any leading Christian thinkers thought that people merit love in virtue of being sentient or experiencing pleasure and pain (see *ibid.*: 191–96). Second, even restricted to human beings, the Golden Rule does not recommend impartially maximizing the good. Jesus tells his followers to allow themselves to be abused and to love their abusers, whom Jesus promises to later “cast ... into the furnace of fire” (Matt. 13:42), which is described as an “everlasting punishment” (Matt. 25:46). A principle that enjoins us to disregard our own well-being vis-à-vis our abusers, and which can countenance a punishment of “everlasting fire” (Matt. 25:41) for those who “offend” (Matt. 13:41) God, seems to be profoundly *dissimilar* from PUB in important respects (again, even restricted to members of our species). Third, Jesus does not purport to take the ‘point of view of the universe’. He explicitly takes the point of view of a specific agent, namely, *God*. He interprets God’s law from the perspective of *God’s* priorities, which do not appear to be informed by PUB – and certainly not by the principle of utility. Although God may ‘love’ people, he will nevertheless “gather out of his kingdom all things that offend, and those who practice lawlessness, and will cast them into the furnace of fire” (Matt. 13:41–42). To suggest, as de Lazari-Radek and Singer do, that Jesus took the ‘point of view of the universe’ to discover a moral principle that is ‘similar’ to PUB is a big stretch.

The version of the Golden Rule in the Hebrew Bible is most naturally read as non-impartial even with respect to human beings: “Do not take revenge and do not hold a grudge against *bnei amechoh* (= sons of your people/nation) and love *reiachoh* (= your fellow) like yourself” (Leviticus 19:18). Most of the rabbinic commentators interpret *reia* (fellow) to refer to Jews, or just religious Jews. Maimonides defines *reia* in this context as a Jew who is your “brother in Torah and mitzvos [i.e., religious commandments]” (*Mishneh Torah*, Avel 14:1; in Maimonides *n.d.*). Even if *reia* is interpreted universally (as it was by some Jewish authorities), there is no indication in the Hebrew Bible that this is the ‘essence of morality’.

Proponents of the “modern claim that the central character and essential originality of Judaism lies in its universalist, humanistic ethics” (Reinhard 2005) often cite an exchange recorded in the Talmud between the sage Hillel and a potential convert. The would-be convert demanded that Hillel teach him the entire Torah while he stood on one foot, to which Hillel replied: “What you hate, do not do to others. All the rest is commentary. Go study” (*Shabbos* 31a; in Schorr 2001). At first glance this sounds similar to the Christian idea that the essence of the law is the Golden Rule. However, Hillel’s statement needs to be understood in context. In a search of the classic Rabbinic literature, Navon (2010) found 80 instances where 15 different mitzvos or states of affairs are declared to be ‘equal to all’ the Torah or the commandments. Things that are equated with all the Torah or all the commandments include the sabbath, circumcision, *tzitzis* (putting fringes on four-cornered garments), studying the Torah, peace, charity, and living in Israel. The idea that Hillel’s statement to the convert shows that the essence of Judaism or Jewish ethics is the Golden Rule is not supported when considering it in the context of the tradition.

Even if we take Hillel’s statement at face value, it would not mean that something like PUB is the essence of Jewish ethics. His injunction does not say anything about maximizing the good of all people – let alone of all sentient beings. It is a general exhortation to prosocial behavior. And Hillel does not purport to take the ‘point of view of the universe’.

De Lazari-Radek and Singer (2012: 26, n. 44; 2014: 193) quote the *Mahabharata*: “One should not behave towards others in a way which is disagreeable to oneself. This is the essence of morality. All other activities are due to selfish desire.” (The word translated as ‘morality’ is *dharma*.) But presenting this quote in isolation may paint a misleading picture of Hindu ethics. Bakker (2013: 49) observes that:

the Hindu commentaries do not pay special attention to the Golden Rule as a subject in its own right ... The Hindu authors focus on themes like *dharma*, self-control, *ahimsa* (non-violence), asceticism, the effects of certain deeds on afterlife and the importance of living in accordance with the caste hierarchy, while the Golden Rule itself is scarcely mentioned separately.

As Davis (2008: 147) says, “Indic formulations of the Golden Rule” (including the one in the *Mahabharata*) “repeatedly point toward a principle that is much more significant within classical Indic ethical discourse, namely the principle of *ahimsa*, nonharming or nonviolence” (quoted in Bakker 2013: 49). It may be that some Western commentators eager to find commonalities between Christianity and the other great religions have assigned the Golden Rule in Hinduism a meaning and significance that it does not have. A few statements in the Hindu literature are superficially similar to the Golden Rule of the Gospels, but they are rooted in fundamentally different philosophical and ethical perspectives.

De Lazari-Radek and Singer cite but do not quote passages from Confucius and the Buddha.

The *Analects* records the following exchange with Confucius:

Tzu-kung asked, “Is there a single word which can be a guide to conduct throughout one’s life?” The Master said, “It is perhaps the word ‘*shu*’. Do not impose on others what you yourself do not desire.” (*Analects* 15.24; in Confucius 1992)

How this statement should be interpreted is not obvious, but the idea that the essence of Confucian morality is a Christian-style Golden Rule is implausible on its face. It simply does not fit with any of the conspicuous features of Confucianism. Consider the following observations by Csikszentmihalyi (2008: 157):

The earliest Chinese expressions of Golden Rule-style injunctions existed somewhat uneasily within a system that otherwise emphasized acting out of a set of virtues. While post-Buddhist Confucians were better able to integrate the general principle of reflexivity into their moral system, they still had difficulty reconciling it with classical aspects of their tradition. A close examination of both early and late traditional writing on Golden Rule passages in the Confucian canon reveals that the scope of the application of the rule was often restricted, sometimes even to the point of being used as a metaphor for reflexivity in action rather than as a moral imperative.

Like Hillel, Confucius seems to be making a general exhortation to prosocial behavior, not instructing us to impartially maximize the good of all people or sentient beings, and not purporting to take the ‘point of view of the universe’. Confucians explicitly reject the idea of impartial moral concern. Mencius called Mozi a ‘beast’ for promoting an ethics of ‘inclusive care’, which Mencius interpreted (perhaps wrongly) to mean that people do not have special moral obligations to their fathers (Fraser 2016: xi). (More on Mozi in Section 4.)

De Lazari-Radek and Singer (2012: 26, n. 44; 2014: 193, n. 25) refer to the following statement by the Buddha: “What is displeasing and disagreeable to me is displeasing and disagreeable to the other too. How can I inflict upon another what is displeasing and disagreeable to me?” (*Samyutta Nikāya* 353–54; in Bhikkhu Bodhi 2000). Nothing the Buddha says here or elsewhere suggests that this principle is the essence of morality, that the good of all beings should be maximized impartially, or that he purports to take the ‘point of view of the universe’.

Again, these are the five examples de Lazari-Radek and Singer themselves cite to support their claim that ‘leading thinkers’ in different traditions took the ‘point of view of the universe’ to discern a foundational moral principle that is ‘similar’ to PUB. The analysis above seems to cast doubt on this claim. Nevertheless, the Golden Rule is indeed widespread. Terry (2011: 14) collects numerous examples of the rule (in both its positive and negative form) and concludes that it is a “self-evident, universal doctrinal [principle] of ethics found throughout the history of civilization in one form or another.” The fact that some version of the Golden Rule (however it’s interpreted) can be found in the literature of virtually every major tradition does not mean it is widely regarded as the foundational principle of morality. As argued above, it seems to have such significance only in Christianity. But why does the Golden Rule keep appearing at all? Could it be that thinkers in different traditions are grasping at the same ‘self-evident’, mind-independent moral truth? Or is there a naturalistic explanation for why thinkers have independently formulated similar-sounding moral principles?

3. The Golden Rule as a means of encouraging empathy and prosociality

When de Lazari-Radek and Singer (2012: 26) explicitly outline their procedure for defending an intuition, they say (as quoted above) that one must establish “the absence

of a plausible explanation of the intuition as the outcome of an evolutionary or other non-truth-tracking process.” Note that they refer to “evolutionary or other non-truth-tracking process[es].”⁴ However, despite asserting that belief in PUB does not have any evolutionary or “other obvious non-truth-tracking explanation,” they do not devote a single sentence to exploring any possible *nonobvious*, non-truth-tracking processes besides evolution by natural selection that could be responsible for it. Elsewhere, they present the evolutionary explanation and the realist explanation for belief in PUB as the only two options:

In the absence of an appeal to our evolved capacity to reason as the basis for our ability to grasp moral truth ... it is difficult to see what plausible evolutionary explanation there could be for the idea of equal concern for the interests of complete strangers [including all sentient beings] who do not belong to one’s own group. (ibid.: 20)

In practice, de Lazari-Radek and Singer conflate *evolutionary* debunking arguments with debunking arguments generally, which leads them to conclude that if a moral belief lacks a Street-style evolutionary explanation (Street 2006) – that is, if the underlying evaluative disposition would not have promoted fitness in the ancestral environment – then it cannot be debunked.

I suggest that there is a plausible naturalistic (albeit not directly evolutionary) explanation for why the Golden Rule arose independently in many traditions.

From a naturalistic perspective, morality serves a biological and/or social function of facilitating cooperation, at least within groups or certain segments within groups (see Boehm 2012; Harman 1977; Joyce 2006; Mackie 1977; cf. Curry et al. 2019; Sterelny & Fraser 2017). It is not surprising that moral educators in different traditions, who were concerned with the commonweal and understood basic human psychology, would hit upon some of the same rhetorical and pedagogical strategies. Some educators realized that they could foster prosocial, cooperative behavior by encouraging people to imagine themselves in others’ shoes, thereby harnessing our capacity for empathy. (Empathy, in turn, is the product of natural selection, so there is no reason to think that it is a moral-truth-tracking emotion; Cofnas 2020: 3184–85.) Nevertheless, the Golden Rule has not been regarded as a central moral principle in any of the great religions besides Christianity. (And we have seen how, even in Christianity, the Golden Rule is quite different from PUB.)

Nagel (1970) seems to consider but then reject an idea along these lines. He suggests that “we are all in some degree susceptible” to the (Golden Rule-style) moral argument, “How would you like it if someone did that to you?” (ibid.: 82). He speculates: “It could be that the thought of yourself in a position similar to that of your victim is so vivid and unpleasant that you find it distasteful to go on persecuting the wretch” (ibid.: 82). He dismisses the idea that this affective response accounts for the force of the argument with the following rhetorical questions: “what if you [do not have this] affective response? Or alternatively, why cannot such considerations motivate you to ... take a tranquilizer to quell your pity, rather than to desist from your persecutions?” (ibid.: 82–83). Even if we discount our emotions, he assumes we would still recognize the

⁴De Lazari-Radek and Singer (2014: 195) use slightly different phraseology: “the absence of a plausible explanation of the intuition as a non-truth-tracking psychological process.” But their argument is essentially the same.

validity of the Golden Rule via *reason*. But his rhetorical questions may have straightforward answers. It is *not* the case that everyone is, in all circumstances, in some degree susceptible to ‘How would you like it if someone did that to you?’ exhortations. Some people *do* lack the relevant affective response. Others *are* motivated to assuage their conscience with drink and drugs. The value of asking people to put themselves in the place of their victim is not that this always motivates people ‘in some degree’ to desist from harmful actions, but that it *sometimes* works for people who happen to possess the requisite affective and cognitive dispositions.

4. A religious route to impartial morality

Some thinkers have followed a religious line of reasoning, which led them to conclude that the interests of all human beings count equally – not from the perspective of the universe, but of *God*. These thinkers were not concerned with sentience or pleasure and pain per se, so they had no regard for animals and did not advocate PUB or the principle of utility. But they did arrive at a principle that is genuinely similar to PUB and the principle of utility in an important respect, and which, unlike the Golden Rule, advocates genuine impartiality (among human beings). However, the theologians in question were making inferences about the intentions of a divine agent described in their religious tradition. As I argue below, assuming there is a naturalistic explanation for their religious beliefs, there is a naturalistic explanation for the inferences made on the basis of those beliefs. Taking the point of view of a divine agent is fundamentally different from taking the ‘point of view of the universe’.

First, there was Mozi, who lived during the Warring States period (fourth century BCE) in China. Mozi is famous for his doctrine of ‘inclusive care’, which enjoins us to care about all the people of the world like ourselves. He based his ethics on the perspective of Heaven, which was seen as a quasi-personal god (Fraser 2016: 16). “Heaven’s conduct is broad and impartial,” he said (ibid.: 36). From its perspective, all people count equally: “Now in the world there are no great or small states – all are Heaven’s towns. Among people there are no younger or elder, noble or common – all are Heaven’s subjects” (ibid.: 36). According to Mozi, action should be guided by the following principle:

Does it benefit people? Then do it. Does it not benefit people? Then stop ... In all statements and all actions, what is beneficial to Heaven, ghosts, and the people, do it. In all statements and all actions, what is harmful to Heaven, ghosts, and the people, reject it. (ibid.: 138)

In Mozi’s philosophy, the “benefit [*li*] of all,” despite its utilitarian ring, does not refer to “the total sum of welfare in society or to individuals’ average level of welfare” (ibid.: 141). As Fraser (2016: 138) explains, “benefit” includes three kinds of goods: “material prosperity, an adequate population or family size, and sociopolitical order, including social stability and personal and national security. These three goods are the concrete criteria against which the Mohists evaluate statements, conduct, practices, and institutions.” Although his theory of the good was very different from that of Bentham, Sidgwick, or any other advocates of PUB in the West, Mozi called on us to impartially maximize the good with respect to human beings.

In the long run, Mohism “had little direct influence on the Chinese philosophical tradition” (ibid.: 19). As mentioned earlier, Mozi’s doctrine of impartiality was explicitly rejected by the Confucians who came to dominate Chinese philosophy.

Several decades before Bentham (1789) formulated the principle of utility, some Anglican theologians developed a religious version of what scholars now call utilitarianism, which was restricted to our species (or ‘rational beings’). Like Mozi, they reasoned that God has an interest in impartially maximizing the good. They combined this idea with a theory of the good that was much closer to Bentham’s, namely, the good consists in maximizing pleasure and minimizing pain in rational beings (Heydt 2014).

Berkeley (1712/1871: § 6) argued that, because God has the power to bring us everlasting bliss or damnation, “the principles of Reason” demand that we act in “conformity to His will.” In addition, we are duty bound to obey God as “the maker and preserver of all things.” Since God is infinitely good, His end must be the “good of men.”

[A]ntecedent to the end proposed by God, no distinction can be conceived between men; that end therefore itself, or general design of Providence, is not determined or limited by any respect of persons. It is not therefore the private good of this or that man, nation, or age, but the general well-being of all men, of all nations, of all ages of the world, which God designs should be procured by the concurring actions of each individual. (ibid.: § 7)

Another leading theological utilitarian, John Gay, argued similarly that ‘obligation’ is the requirement to act to promote our happiness (Gay 1731/1781: xxx–xxx). “[O]bligation is evidently founded upon the prospect of happiness, and arises from that necessary influence which any action has upon present or future happiness or misery” (ibid.: xxx). “[B]ecause God only can in all cases make a man happy or miserable ... we are *always* obliged” to conform to “the Will of God” (ibid.: xxxi).

Now it is evident from the nature of God, *viz.* his being infinitely happy in himself from all eternity, and from his goodness manifested in his works, that he could have no other design in creating mankind than *their* happiness; and therefore he wills their happiness; therefore the means of their happiness: therefore that my behaviour, as far as it may be a means of the happiness of mankind, should be such ... Thus the will of God is the immediate criterion of virtue, and the happiness of mankind the criterion of the will of God ... (ibid.: xxxi)

Gay argued that those who define virtue without reference to “the will of God” must be guilty of one of two absurdities. They “must either allow that virtue is not in all cases *obligatory* ... or they must say that the good of mankind is a sufficient obligation.” “But,” he asks rhetorically, “how can the good of mankind be any obligation to *me*, when perhaps in particular cases, such as laying down my life, or the like, it is contrary to my happiness?” (ibid.: xxxiii).

One might have thought that, for Mozi and the Anglican theologians, ‘Heaven’ and ‘God’ were somehow equivalent to ‘the universe’ – that these religious thinkers were simply using the vocabulary available to them to communicate Sidgwick’s core idea. But the quotes above suggest that they were explicitly making inferences about the *intentions* of a *particular divine agent* based on the actions ascribed to it. For them, the divine agent was not a metaphor, the literal reality of which did not matter for their conclusions. For Mozi, there is no reason to think that we would be bound by the doctrine of inclusive care if Heaven disappeared or changed its intentions. The argument of the Anglicans can be summarized as follows: (a) God, being good, must desire good for humans, (b) we should do what an (omnipotent and omniscient) God wants in

order to be rewarded and to avoid punishment, (c) from God's perspective the good of all people is equally important, (d) ergo we should act to impartially maximize the good of all people. All three premises leading to the conclusion collapse if we reject God. Although one could make a case for substituting 'the universe' for 'God' in premise (c), there's no indication that the Anglican utilitarians would have accepted that, and, without (a) and (b), the argument falls apart anyway. The Anglicans were clear about their view that "moral obligation requires God and God's sanctions" (Heydt 2014: 26). As noted, Gay (1731/1781: xxxiii) ridiculed those who believed otherwise.

If the argument above is right, Mohists and Anglicans came to advocate moral impartiality only because they held some very particular beliefs about God. In order to give a complete naturalistic explanation for their moral conclusions, we have to explain the religious beliefs that motivated them. Why did they believe in a Creator and/or Designer who wants good for human beings? Why did the Anglicans ascribe to this Being the power and motivation to reward and punish people? (The Mohists also believed in divine retribution by Heaven, ghosts, and spirits, but this principle did not play such a crucial role in their argument for moral impartiality – at least not explicitly.)

Regarding the attribution of infinite power and goodness to God, Hume (1757/2007: 6.5) offers some interesting speculations. Originally, he suggests, pagans "represent one god as the prince or supreme magistrate of the rest." To ingratiate themselves to this special deity, people compete with each other to "invent new strains of adulation," which become increasingly extravagant, "till at last they arrive at infinity itself, beyond which there is no farther progress." All other gods eventually dissolve into insignificance or even nothingness, supplanted by an all-powerful, all-knowing, and perfectly good Creator.

Norenzayan (2013) argues that belief in 'Big Gods' who enforce prosocial behavior was culturally successful because it made large-scale cooperation possible. The gods of hunter-gatherers often have strikingly humanlike limitations and parochial concerns, with little interest in morality (ibid.: 7). But when some groups developed concepts of powerful, morally concerned gods, this triggered a "runaway process of cultural evolution" (ibid.: 8). Shared belief in a powerful god who demands prosocial behavior allowed people to form large, anonymous communities that would otherwise fall apart if many would-be cheaters did not fear divine retribution. These communities, being large and extra cooperative, tended toward expansion, spreading their way of life – including their religious beliefs – across the world. It is only recently that, in a handful of developed countries, sophisticated, secular law-enforcement institutions have taken the place of God.

Hume's and Norenzayan's theories are not necessarily in conflict. It could be that belief in Big Gods initially proliferated because it facilitated cooperation among large groups. One god in the original pantheon became magnified in significance via the process described by Hume.

Once achieved, belief in an all-powerful and moral God does not *inevitably* lead to the conclusion that He/It demands moral impartiality. But from the premises that God (or Heaven) is very powerful or omnipotent, a moral paragon, and concerned with human well-being, theologians may draw the conclusion that, from the perspective of that divine agent, the well-being of all people matters equally.

5. Discussion

Some Western scholars claim that the Golden Rule is the cornerstone of morality across many if not all cultures, and that it affirms the value of moral impartiality. Some moral

realists contend that widespread acceptance of the Golden Rule has no naturalistic explanation, and is best explained as the result of people independently recognizing the same non-natural moral truth. I have argued that the Golden Rule is a central moral principle only in Christianity. It does not advocate anything like moral impartiality – or PUB or the principle of utility. And the appearance of the Golden Rule across cultures has a straightforward naturalistic explanation, namely, some moral educators independently recognized the pedagogical value of encouraging people to empathize with others in their moral community. The Mohists and the Anglican utilitarians did advocate genuinely impartial morality *among human beings*, but they did so on the basis of explicitly religious premises. Since (I have argued) it is reasonable to suppose that the Mohists’ and the Anglicans’ religious beliefs have naturalistic explanations, so do the conclusions about impartiality that were derived from them.

All that being said, I think Gensler (2013: 22) is right that the Golden Rule does capture “much of the spirit behind morality.” All moral systems depend on (more or less selective) empathy to function. By inviting us to imagine ourselves in the position of others, the Golden Rule highlights the importance of empathy and helps us cultivate this key moral emotion. Rönnedal (2015) has shown that, if we make certain assumptions that are plausible in some contexts, the Golden Rule – “If you want others to do A to you, then you ought to do A to them” – is equivalent to the Platinum Rule – “If others want you to do A to them, then you ought to do A to them.” Although the two rules often generate the same moral demands, “psychologically [they] might have different effects” (ibid.: 235). The *Platinum* Rule reminds us that not everyone shares our desires, so does not literally want us to do unto them as we would have them do unto us. The *Golden* Rule reminds us that we treat people well by acting *against* their wishes in cases where we ourselves would not want our desires to be indulged (e.g., when our inebriated friend wants to borrow our car). The Golden Rule and the Platinum Rule are not absolute moral principles, but “rules of thumb” that are “useful in moral education, at least at some stages” (ibid.: 235). However, from an *anthropological* perspective, the Golden Rule (or the Platinum Rule) does not express a substantive moral principle that has been widely held across cultures.

The Golden Rule may be a useful starting point for constructing normative ethical theories. In particular, it can serve as the basis – or as a slogan – for the sort of humanistic, impartial ethics favored by many theorists today. But to do this it will need to be enriched by a number of qualifications (Carson 2010: ch. 6; 2013; Gensler 2013; Rönnedal 2015; Wattles 1996: ch. 13) that were not part of how it was historically interpreted in most of the religious and ethical traditions in which it appeared. If I am right that belief in the Golden Rule has a ‘debunking’ explanation, that does not mean that it is without value. Nor does it mean that the Golden Rule does not represent a moral truth – if morality is *naturalistically* construed (e.g., Sterelny & Fraser 2017). It does, however, undermine the justification for our belief that the rule represents a *non-natural* moral truth.

Theoretically, non-natural moral realists only need to find a single moral belief that lacks a plausible naturalistic explanation, which would be immune from the challenge posed by debunking arguments. But this will not be as easy as pointing to a moral belief that apparently would not have been favored by natural selection. Debunking arguments do not need to be *evolutionary*. Beliefs are debunked by revealing them to be the product of any causes – not just evolutionary ones – that do not track truth in the relevant domain. If the sociological processes that produced our moral beliefs can be reduced to non-moral-truth-tracking causes, the resulting beliefs can be debunked.

Acknowledgments. Thanks to Roger Crisp, Guy Kahane, Dale Miller, Andreas Mogensen, Kim Sterelny, and two anonymous reviewers for helpful comments on earlier drafts of this article.

Competing interests. The author declares none.

References

- Bakker, Freek L.** 2013. Comparing the Golden Rule in Hindu and Christian Religious Texts. *Studies in Religion* 42(1): 38–58.
- Bentham, Jeremy.** 1789. *An Introduction to the Principles of Morals and Legislation*. London: T. Payne and Sons.
- Berkeley, George.** 1712/1871. Passive Obedience. In Alexander Campbell Fraser (ed.), *The Works of George Berkeley, D.D.*, pp. 103–39. Oxford: Clarendon Press.
- Bhikkhu Bodhi** (trans.). 2000. *The Connected Discourses of the Buddha: A Translation of the Saṃyutta Nikāya*. Boston: Wisdom Publications.
- Boehm, Christopher.** 2012. *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. New York: Basic Books.
- Carson, Thomas L.** 2010. *Lying and Deception: Theory and Practice*. Oxford: Oxford University Press.
- Carson, Thomas L.** 2013. Golden Rule. In Hugh LaFollette (ed.), *The International Encyclopedia of Ethics*, pp. 2186–92. Oxford: Wiley-Blackwell.
- Cofnas, Nathan.** 2020. A Debunking Explanation for Moral Progress. *Philosophical Studies* 177(11): 3171–91.
- Confucius.** 1992. *The Analects*. Trans. by D. C. Lau. Hong Kong: Chinese University Press.
- Crisp, Roger.** 2012. Katarzyna de Lazari-Radek and Peter Singer, ‘The Objectivity of Ethics and the Unity of Practical Reason’. <http://peasoup.us/2012/12/ethics-discussions-at-pea-soup-katarzyna-de-lazari-radek-and-peter-singer-the-objectivity-of-ethics-1/>.
- Csikszentmihalyi, Mark A.** 2008. The Golden Rule in Confucianism. In Jacob Neusner and Bruce Chilton (eds.), *The Golden Rule: The Ethics of Reciprocity in World Religions*, pp. 157–69. London: Continuum.
- Curry, Oliver Scott, Daniel Austin Mullins, and Harvey Whitehouse.** 2019. Is It Good to Cooperate? Testing the Theory of Morality-as-Cooperation in 60 Societies. *Current Anthropology* 60(1): 47–69.
- Davis, Richard H.** 2008. A Hindu Golden Rule, in Context. In Jacob Neusner and Bruce Chilton (eds.), *The Golden Rule: The Ethics of Reciprocity in World Religions*, pp. 146–56. London: Continuum.
- de Lazari-Radek, Katarzyna, and Peter Singer.** 2012. The Objectivity of Ethics and the Unity of Practical Reason. *Ethics* 123(1): 9–31.
- de Lazari-Radek, Katarzyna, and Peter Singer.** 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.
- du Roy, Olivier.** 2008. The Golden Rule as the Law of Nature, from Origen to Martin Luther. In Jacob Neusner and Bruce Chilton (eds.), *The Golden Rule: The Ethics of Reciprocity in World Religions*, pp. 88–98. New York: Continuum.
- Fraser, Chris.** 2016. *The Philosophy of the Mōzi: The First Consequentialists*. New York: Columbia University Press.
- Gay, John.** 1731/1781. Preliminary Dissertation Concerning the Fundamental Principle of Virtue or Morality. In Edmund Law (ed.), *An Essay on the Origin of Evil* (5th ed.), pp. xxiii–xlv. London: R. Faulder.
- Gensler, Harry J.** 2013. *Ethics and the Golden Rule*. New York: Routledge.
- Harman, Gilbert.** 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.
- Heydt, Colin.** 2014. Utilitarianism before Bentham. In Ben Eggleston and Dale E. Miller (eds.), *The Cambridge Companion to Utilitarianism*, pp. 16–37. Cambridge, UK: Cambridge University Press.
- Hobbes, Thomas.** 1651/1996. *Leviathan, or the Matter, Forme, & Power of a Common-Wealth Ecclesiasticall and Civill*. Ed. by J. C. A. Gaskin. Oxford: Oxford University Press.
- Huemer, Michael.** 2005. *Ethical Intuitionism*. New York: Palgrave Macmillan.
- Hume, David.** 1757/2007. The Natural History of Religion. In Tom L. Beauchamp (ed.), *A Dissertation on the Passions; The Natural History of Religion: A Critical Edition*, pp. 33–87. Oxford: Clarendon Press.
- Joyce, Richard.** 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Kahane, Guy.** 2011. Evolutionary Debunking Arguments. *Noûs* 45(1): 103–25.

- Kahane, Guy.** 2014. Evolution and Impartiality. *Ethics* 124(2): 327–41.
- Mackie, J. L.** 1977. *Ethics: Inventing Right and Wrong*. New York: Penguin Books.
- Maimonides, Moses.** n.d. Mishneh Torah: Avel - Chapter 14. https://www.chabad.org/library/article_cdo/aid/1181895/jewish/Avel-Chapter-14.htm#t=he.
- Mill, John Stuart.** 1863/2015. Utilitarianism. In Mark Philp and Frederick Rosen (eds.), *On Liberty, Utilitarianism and Other Essays*, pp. 115–55. Oxford: Oxford University Press.
- Nagel, Thomas.** 1970. *The Possibility of Altruism*. Princeton, NJ: Princeton University Press.
- Navon, Mois.** 2010. Equal to All the Mitzvot in the Torah. <http://www.divreinavon.com/pdf/EqualToAll.pdf>.
- Nichols, Shaun.** 2014. Process Debunking and Ethics. *Ethics* 124(4): 727–49.
- Norenzayan, Ara.** 2013. *Big Gods: How Religion Transformed Cooperation and Conflict*. Princeton, NJ: Princeton University Press.
- Parfit, Derek.** 2011. *On What Matters*, vol. 2. Oxford: Oxford University Press.
- Reinhard, Kenneth.** 2005. The Ethics of the Neighbor: Universalism, Particularism, Exceptionalism. *Journal of Textual Reasoning* 4(1).
- Rönnedal, Daniel.** 2015. The Golden Rule and the Platinum Rule. *The Journal of Value Inquiry* 49(1–2): 221–36.
- Ruse, Michael, and Edward O. Wilson.** 1986. Moral Philosophy as Applied Science. *Philosophy* 61(236): 173–92.
- Sauer, Hanno.** 2018. *Debunking Arguments in Ethics*. Cambridge, UK: Cambridge University Press.
- Schorr, Yisroel Simcha** (ed.). 2001. *Talmud Bavli - Shabbos*. New York: Mesorah.
- Shafer-Landau, Russ.** 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Sidgwick, Henry.** 1907. *The Methods of Ethics*. London: Macmillan.
- Singer, Peter.** 1975/2002. *Animal Liberation*. New York: HarperCollins.
- Sterelny, Kim, and Ben Fraser.** 2017. Evolution and Moral Realism. *The British Journal for the Philosophy of Science* 68(4): 981–1006.
- Street, Sharon.** 2006. A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127(1): 109–66.
- Terry, Q. C.** 2011. *Golden Rules and Silver Rules of Humanity: Universal Wisdom of Civilization*. West Conshohocken, PA: Infinity.
- Wattles, Jeffrey.** 1996. *The Golden Rule*. Oxford: Oxford University Press.
- Westermarck, Edward.** 1906. *The Origin and Development of the Moral Ideas*. New York: Macmillan.