The Gige Ring and the New York Hot Dog: A Kantian Constructivist Analysis of Moral Motivation

Jacopo Morelli[1]

1 University of Padua

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

In this article I will try to reconstruct a brief history of the philosophical debate on the myth of Gige's ring, showing the different positions and answers to it, arguing that the Kantian approach may offer the best reply to Glaucon's challenge, rejecting the underlying account that we only have reason to do what promotes our happiness. I will then present the proposal of Kantian constructivism, explaining why the concept of the constitutive principle of reason makes it possible to provide an answer to Glaucon without having to fall into forms of metaphysical foundationalism. I will expand this theme through the analysis of a simple everyday case (a walk in the streets of New York) in order to provide an analysis of the relationship between moral motivation and normativity for Kantian constructivists.

Jacopo Morelli

University of Modena-Reggio

INDEX

Introduction

The ring of Gyges is a mythical magical artifact mentioned by Glaucon in Plato's book "Republic", to argue against justice in itself. Since the ring makes its bearer invisible, it is easy for him to behave by ignoring the consequences of his actions, simply trying to maximize his interests. Glaucon argues that in these circumstances every agent with the ring should act

for his own benefit, and that it would be foolish not to do so. To assert this implies embracing a concept of ethics based solely on prudential reasons. But moral reasons are not prudential reasons, because even assuming that the pastor with Gyges' ring can always get away with no consequences, his actions would still remain morally wrong. This is an intuition[1] that we all seem to have, even a fictionalist like Joyce, although he adds two specifications: 1) that the feeling of injustice surrounding the action of the peasant is due to evolutionary reasons and 2) that this feeling, despite being found in most agents, it is not rationally justifiable albeit pragmatically useful (Joyce, 2016). What I will argue in this article is instead that it is possible to rationally justify action, even in a hypothetical case like that of the ring of Gyges, that is to say, that it is possible to explain why the pastor's work is not only morally condemnable but rationally wrong.

Firstly, I will try to reconstruct a brief history of the philosophical debate on the myth of Gige's ring, showing the different positions and answers to it, arguing that the Kantian approach may offer the best reply to Glaucon's challenge, rejecting the underlying account that we only have reason to do what promotes our happiness. I will then present the proposal of Kantian constructivism, explaining why the concept of the constitutive principle of reason makes it possible to provide an answer to Glaucon without having to fall into forms of metaphysical foundationalism. I will expand this theme through the analysis of a simple everyday case (a walk in the streets of New York) in order to provide an analysis of the relationship between moral motivation and normativity for Kantian constructivists. Since the ontological conception of normativity as 'external' seems to be at odds with the internal authority and the motivational power of practical judgments, it is difficult to understand why an ontology of previous and independent normative facts must generate reasons that are authoritative and effective from the practical point of view for those who are about to take action. Precisely because of this difficulty, Kantian constructivism retrieves the interpretation of autonomy as a form of self-regulation: we are subject only to those laws on which we legislated, that is, those laws that were built, constructed, by ourselves. Only by independently reflecting on why a given principle is right (therefore rational) can the subject really be motivated to act according to it[2].

1. Ring of Gyges: the myth and some possible solutions

In Book II of Plato's Republic, Socrates' main interlocutor, Glaucon, raises a vivid and powerful challenge to justice and morality in general. Glaucon challenges Socrates' claim that justice is desirable for its own sake and for the sake of its consequences. He argues instead that most people consider justice burdensome and onerous but practice it for the good consequences that are a result of having a reputation for justice. He presents the case of Gyges, a Lydian shepherd who possesses a ring that made its wearer invisible. Thanks to the ring, the shepherd established himself and his descendants on the throne of Lydia. The importance of this story is pointed out to us by Glaucon, in expounding the thesis that injustice is superior to justice, which he wants Socrates to refute. "Now if there were two such rings, and the just man would put one on, and the unjust man the other, no one, as it would seem, would be so adamant a to stick by justice and bring himself to keep away from what belongs to others and not lay hold of it, although he had license to take what he wanted from the market without fear, and to go into houses and have intercourse with whomever he wanted, and to slay or release from bonds whomever wanted, and to do other things as an equal to a god among humans. And in so doing, one would act no differently from the other, but both would go the same way?" (Republic, 359c-360b)

Glaucon's contention is that Gyges does not have reason to be just in this circumstance, since being just will not promote his happiness: in this type of circumstance would be right, for every agent, to pursue injustice, "for someone who did not want to do injustice, given this sort of opportunity, and who did not touch other people's property, would be thought most wretched and most foolish by everyone aware of the situation" (Republic 360d). The ring puts its wearer outside the circumstances of justice, showing that justice is not something that has value per se, but only a social imposition on nature. Glaucon portrays the natural man as pre-social and pre-moral, and, from this standpoint, morality seems to be a "necessary evil": the idea that morality reflects only human weakness and the inability to take net advantage of others.

For the most part, philosophers have responded to Glaucon's challenge by rejecting one or more of the key claims[3] underlying his account: (1) the contractarian or conventionalist theory of justice; (2) the claim that we are essentially acquisitive and competitive, and that human happiness crucially involves the possession of material goods and power; (3) the claim that we only have reason to do what promotes our happiness. Plato attempted to answer Glaucon's challenge by simultaneously rejecting (1), the contractarian account of justice, and (2), the account of human happiness, and several philosophers have since followed this approach, expanding the analysis and showing the various contradictions inherent in Glauco's assumptions. For many, if morality is represented in this Sophist way, as a conventional constraint on the natural mode of human behavior rather than as integral to human good, then any genuine adherence to morality is undermined. But as MacIntyre (1967) noted, the Sophist conception of the natural man is not so purified from civilization and "nomos" as the Sophists taught, but merely based on the previous system of values. "What he [the natural man as understood by the sophists] express is not nature, but the social attitudes of the Homeric hero" (pg.17), with power and pleasure being their exclusive interest.

However, these lines of thought often limit themselves to the simple*pars destruens*, criticizing the foundation and reduction of morality to the mere sphere of the social contract, but failing to provide equally solid foundations for explaining why the shepherd's action with the ring is wrong. Often these proposals follow Plato in a foundationalist, realist approach to ethics, which identify and locate the good in a certain area of the ontological realm (tending to be metaphysical) and then base moral practice on a process of recognizing these entities. The problem with realist conceptions of morality and justice, in addition to general foundationalist difficulties, is that the normativity they enact is disengaged from the autonomous deliberative capacity of rational agents[4].

Other philosophers maintained a broadly conventional or contractarian account of justice, but challenged (2), the account of human nature and human happiness. Glaucon claims that we are by nature acquisitive and competitive and thus take material goods and power to be the main components of happiness. But surely this is a limited and one-dimensional account of human nature. We are also social creatures: we require companionship and cooperation and cannot flourish in solitude. Thus, if we understand the goods that are scarce in the circumstances of justice to include those that we acquire through the companionship and cooperation of others, then even equipped with Gyges' ring we would still need justice. In his Morals by Agreement (1986) David Gauthier argues that Glaucon is mistaken in representing "justice as a necessary evil because he supposes that it affords us only a second-best means to our ends." He argues that the cooperation made possible by justice is not a second-best way of realizing ends that might sometimes be achieved in other ways, for "in

cooperating we make the most effective use of our powers to attain ends that would otherwise lie beyond our individual capacities. And we find value in ways that no solitary being could experience." Glaucon is wrong, then, to think that Gyges would have no need for justice, since humans "are neither all-powerful nor all-sufficient" (Gauthier 1986: 345; see also Gauthier 1990).

The problem with this type of approach, however, is the absence of a justification as to why always be just and, more specifically, it may not show that we have reason to be just to everyone. Glaucon may concede that people are not affectively and socially self-sufficient and so need to be just to some people in order to be happy, without conceding that we need to be just to everyone, particularly those with whom we do not care to have cooperative relationships. David Gauthier's idea that rational agents, aiming to maximize their expected utility, would in suitable conditions choose to become "constrained maximizers, does not provide solutions to not-suitable situations such as the one in which the Lydian pastors find himself. In Gauthier's view there are no requirements to constrain our maximizing activity, neither to enable us to cooperate nor to show respect for each other, unless it is rational to constrain our maximizing activity. Therefore, Gauthier argues that it is rational for Glaucon to constrain his maximizing activity, while failing to provide him with genuine self-motivating and truly normative reasons, since for Gauthier too the reasons for action are solely prudential[5]. The crucial point is that people who are harmed by invisibility are wronged, they are victims of injustice, not only because of the deliberation of the subject (who is only interested in its maximization), but especially because of the factual conditions of injustice that any impartial subject can ascertain, and it is from these conditions of injustice that the deliberation starts.

Finally, many philosophers respond to Glaucon's challenge by rejecting (3), the view that we only have reason to do what promotes our happiness. The core idea is that we have reason to do what is just or right, regardless of the impact on our happiness, indeed, regardless of the impact on any of our ends, whether those ends are narrowly self-interested or altruistic. Among the various philosophers[6] that subscribe to this thesis, there are the Kantian constructivists, according to whom all rational beings owe just treatment to other rational beings, which are considered as ends in themselves, and have reason to act accordingly. To better understand these questions about practical rationality we need to comprehend the central difference between the two main types of reasons that we have already mentioned several times during this discussion: motivational and normative reasons. By the term "reasons" we seem to use a single concept that is used to answer different questions: the question of whether there is a reason for someone to do something (normative) and the question of what someone's reason for acting is (motivating)[7]. Normative reasons were conceived of as facts, and so were regarded as mind-independent: the facts are what they are independently of whether anyone knows them or thinks about them. By contrast, motivating reasons were conceived of as mental states of agents and, as such, as entities that depend on someone's thinking or believing certain things (Audi 2001 and Mele 2003 are representative examples—but see also Mele 2013).

In recent years, however, this assumption has been challenged, giving rise to several disputes about the ontology of reasons[8]. Whichever source we seek to trace in normative reasons (whether rationality, desires, or values[9]), the practical function they must serve is still to motivate action as well as justify it. Therefore, any account of normative reasons must offer a plausible explanation of the relationship between the normativity of reasons and the capacity that reasons have to

motivate agents to act. It is necessary to provide a coherent account that can explain how thinking that there is a reason for me to do something can motivate me to act, and to act *for* that reason. Desire-based accounts of reasons might seem to have an advantage in this respect, compared to constructivists: if normative reasons depend on my antecedent motivations, desires, and plans, then it is plausible that I am motivated to do what I believe will help satisfy or grow those motivations. But desire-based accounts seem to fail to account for another central feature concerning normative reasons, namely the normativity that influences agents independently of their motivations[10]. Why should we act rightly if we do not seem motivated to do so? This continues to be at the heart of Glaucon's challenge.

The solution of the Kantian constructivists lies in bringing normativity back inside the agent itself, rather than outside. Kant himself was committed to the “constitutivist view” that the source of the categorical force of moral obligations lies in the constitutive features of rational agency (Bagnoli 2011). The normativity of reasons in this conception is provided not by desires or conceptions of the good, but by the rationality and deliberation of autonomous agents. The distinctiveness of this conception lies in the idea that reason must be examined by reason itself in order for its verdicts to be justified, bringing normativity back into the agents, but still making it public. Reasoning is public not in the sense that it is a form of social negotiation, but in the Kantian sense that it has a reflexive and dialogical structure, and, for this reason, capable of producing reasons that are addressed to other rational beings: it is precisely reasons that are among the most valuable tools for building cooperative practical interactions.

Why should we then think that we always have a reason for being just, even when justice has no relation to our ends? Because we are rational agents and can only judge it unjust and wrong if someone takes advantage of others by using a power such as Gige's ring. It is inherent in our nature as rational agents to detect this injustice; if we nevertheless decide to ignore it for our own personal gain, because in the condition we are in it benefits us to ignore what seems rational to us, we can do so, because we are free agents, but we would remain conscious that we are acting irrationally, namely, in the sense we have specified so far, unjustifiably (if another agent had discovered us we would have no normative reason to give him to justify our action). Even in a hypothetical scenario such as Gige's ring, in which I do not have to provide any justification to other agents, my nature as an autonomous rational agent remains active, continuously and passively, and therefore I know that I am acting immorally because I critically self-judge my actions and cannot inter-subjectively justify them. But why should knowing this motivate me to act morally, if I believe that I will not be found out and therefore do not have to give any reason (that I could not give) for my actions? And what specifically is the deliberative process by which I arrive at reasons that I could provide to publicly justify my actions? In the second chapter, we will answer these questions, starting with the second one first and then, by analyzing a concrete case (that of the New York Hot Dogs) seeking the way we judge and discern normative reasons, to understand how these reasons actually succeed in motivating our actions.

2. The New York Hot Dog: can moral motivation be rationally justified?

You're in New York City, walking down 5th Avenue; you've just picked up a hot dog along the way and find a shriveled child at the curbside holding out his hand and pointing with his other hand, without speaking, to the hot dog. You can judge from his appearance that he is literally starving. What is the right choice to make? The answer seems obvious and

uncontroversial: give food to the child so he can feed himself. But we know that with philosophers it is often not so easy. Williams (1993), for example, argued that no one has reasons for doing something unless he himself is motivated to do it. An agent has a reason for doing F only if it is possible to identify a "sound deliberative route" that allows us to trace the reason in question back to the subjective motivational complex. Someone who deliberates past the hungry child might say that he was simply hungry, and this was the reason that led him to ignore any other reason. The deliberative path, which led him to the practical reason for ignoring the child would be, for Williams, valid, even though selfish.

Williams believes that it is utterly foolish to accuse someone who does not recognize certain reasons of being irrational: the claim to identify definitive practical reasons that apply indifferently to all is, in his view, impossible. He who should respond to the appeal of certain reasons may be accused of being, from time to time, cruel or selfish or thoughtless, but not of acting irrationally; Williams does not deny that those who are indifferent to reasons can be blamed, which is probably what everyone would do in the case taken as an example, but the point he wants to emphasize is that these criticisms cannot be made by asserting that there is a valid rational procedure irrespective of the agents' subjective motivational complex, let alone by saying that the reasons in question were true prior to their being recognized as such by the agents themselves[11]. Deliberating rightly, for Williams, is thus a process that takes place from the actual subjective motivational complex of the agents, depends entirely on it, and cannot be regulated from criteria outside the set itself that would guarantee its correctness.

But if reasoning is a public[12] human faculty, it is possible to sanction its correctness, that is, to check whether the criteria for good reasoning are met, just as it is possible to check whether or not someone is literate by checking whether or not they are correctly applying various grammatical rules[13]. The intuition of constructivism, and that of constitutivism in general, is, as mentioned, that we have reasons common to all, which are constitutive precisely of the fact that we are rational beings, and to act against them is to act irrationally, that is, to act against what practical reason itself suggests. If we base ethical acting on practical reason, we can then check whether an agent is acting morally simply by checking whether he or she is complying with the constitutive reasons. What we want to show with the example of the starving child is precisely that the one who has no reason to act in feeding him is committing an error that prescinds from his motivation, not because there is a specific constitutive reason for this individual case, but because by deliberating and weighing the various reasons correctly we can see which are the constitutive ones and which are contingent, clarifying what needs to be done.

Let us then try to make the arguments transpire and clarify the situation. First of all, we can start by checking the strength of the argument adduced by Williams' egoist, his “being hungry” as a reason for eating the hot dog. Isn't judging one's momentary hunger more important than the life of someone something that can be judged lacking in the most basic reasons? Even if feeding yourself is a basic reason, the *conatus sese conservandi*[14], there are times when it should not be indulged if there are other reasons to act otherwise. If you are bleeding out, and you are supposed to heal your wound, but you start eating because you are also hungry, are you irrational? Here even Wiliams would answer yes, precisely because, between the two actions, both afferent to the principle of self-preservation, there is one that is a priority: before I feed myself, I must consider it more important to avoid bleeding out (otherwise I will act irrational).

For Kantian constructivism, however, it is equally possible that the reason to feed oneself[15] can be overshadowed by other kinds of reasons (not just prudential ones), on pain of the same irrationality[16]. In this case, the failure to recognize that we have a reason for feeding that child that is more important than our reason for being hungry, because the two conditions of hunger are simply different and with opposite consequences (dramatic for the child, indifferent for the adult), involves a practical irrationality, an error of deliberation due to the inability to correctly apply the constitutive principles of reason; the principle in question here might simply be the second formulation of the Kantian categorical imperative, namely, "act in such a way as to treat humanity both in your own person and in the person of every other person always as an end and never simply as a means." Respecting others as end in themselves is something we understand as right simply by reflecting on it, although it is sometimes possible to ignore the evidence. This is due to the fact that we, as human agents are autonomous and therefore fallible.

Indeed, anyone who addresses the topic of moral motivation knows that he or she must address and try to explain the issue of akrasia[17], that is, weakness of will, the disposition to act contrary to one's considered judgment about what is best to do. This is a lengthy topic that deserves a longer space of discussion than I can devote to it here. I will only highlight two important points salient to the present discussion. First of all, it is an empirical fact that akrasia is more unnatural than natural, that is, basically people are more motivated to do what they judge to be right than not to do it. Indeed, it is precisely this motivation that often prompts us to do things that we would not want to do just because we feel they are right (one of the most everyday examples in this regard may be recycling). The second point is to try precisely to provide an answer to this unnaturalness of akrasia. Kantian constructivists have provided different answers on this issue, some of them very distant from each other. Compare, for example, the opposing solutions of Korsgaard and O'Neill, who insisted, respectively, on the theme of unity and fragility of the agent.

Korsgaard bases his conception of practical reasoning on the idea that the reflective agent makes a reasoned choice on the basis of a criterion that identifies him with a certain practical identity. It is therefore not necessary, in this Korsgaardian perspective, to externally justify motivation, because it is already internally presupposed in the concept of practical identity (or personal identity if we use Bernard Williams' vocabulary). Acting on the basis of what our identity requires already provides the motivational drive necessary to justify action: we are interested and involved because what is at stake is our life, and what we consider significant and important in it, namely our attachments and projects.

O'Neill would share with Korsgaard the claim that agents cannot fail to act, and that whenever they engage in practical activity and engage in reasoning, they are also committed to the conditions of acting and reasoning, but she locates these conditions in the plurality, connectedness, and finiteness of human beings, in the plural and collective dimension of acting. This abstract approach of constructivism departs, according to O'Neill, from Kantian ethics, which is designed directly for imperfect non-angelical rational agents; moreover, if human beings were not finite and interdependent on one another, there would be no need to bind them through the institution of moral obligation. For O'Neill, the standpoint of practical reason is external to the individual perspective and can be vindicated precisely because it does not rely on the determinations and attachments that come from a sense of belonging to a particular identity or community. In this respect, according to O'Neill, there is no guarantee that reason and motivation proceed together. We can find, by reasoning,

justified solutions for everyone, but these are not always matched by adequate motivations to act (see O'Neill 2000a, p. 27).

So, while Korsgaard prefers looking from within the deliberative process, from the point of view of the agent who personally decides what to do (similarly to Williams) O'Neill suggests that in order to avoid arbitrariness and inaccessibility practical reasoning must be vindicated from the outside, from the plural point of view of all who participate (even hypothetically[18]), but in doing so she seems to renounce the possibility of rational justification for moral motivation. In the proposal I am outlining, my intention is to recover the positive aspects of both of these constructivist paradigms, uniting and overcoming the concepts of unity and fragility of the agent, in that of agent integrity. In my view, the concept of integrity retains both the internal motivational character (regarding precisely the psychological health of the agent) and the public character, in a form of intersubjective transcendentalism. Our integrity as agents is given not only by the way we respect and act in accordance with the narrative unity we have about ourselves, but also and indeed above all through the intersubjective relationship we have with others: coherence with respect to our identity as agent finds its full realization precisely because I act with others in mind, reflecting on whether the way I am acting can be justified to an external agent.

Conclusions

What I wanted to do in this article, using the myth of Gige's ring as a starting point of discussion, was provide original answers to two of the oldest questions about morality, which I re-illustrate here again: “How do we determine what is right to do?” And “why should knowing what is right to do motivate me to do it? To the first question we have replied “by weighing the various reasons given in support of or against that course of action”. The procedure of rational deliberation (checking the consistency of reasoning that leads to any argument and verifying whether we are violating reasons constitutive of our reasoning) is precisely to vindicate and discern good and bad reasons. The self-authenticating nature of reason as a self-legislative non-arbitrary activity is the true source of normative authority in practical agency, and since there are constitutive principles of reason itself, we can recognize whether or not an agent is acting in accordance with practical reason. Gyges is capable, as a rational agent, of understanding why it is wrong what he is doing, and in the event that he is not, it would simply mean that he has made some error in the deliberation process in applying the constitutive reasons, or he has blundered in a more general and fundamental form of reasoning, in failing to recognize certain constitutive reasons.

Failure to convince a possible possessor of Gige's ring from avoiding behaving unjustly does not invalidate any moral framework that is not an error or a relativist theory. Joyce, while approaching the issue from his finctionalist perspective, is also aware of this when he admits that "from the fact that we could not reasonably exhort them to adopt the fictive stance toward morality it does not follow that we must calmly accept any mischief or violence they perpetrate (Joyce, 2001, p.222)". We cannot provide a principle to the possessor of the ring that will surely persuade him to use it without harming others, however, after clarifying the reasons, he will necessarily be aware of the "wrongness" of behaving in such a way. He is already endowed with a constitutive sense of justice, therefore what we can do is clarify why he has this feeling that he is behaving inappropriately by exploiting invisibility for his own gains. Clarifying means in this case both making

manifest how we arrive at such a feeling and showing why it is well-founded to have it and to act in accordance with it. If at the end of this the possessor of the ring still wants to act for his own personal gain, he may do so, but he will know that he is acting contrary not only to his natural sense of justice but also against his own reason.

As Gauthier writes: “Glaucon thinks of the just man merely as someone who recognizes the need to accept certain constraints, but whose emotions are in no way engaged by them. The just man thus lacks any sense of justice as capacity to be moved by considerations of justice as such It is then not surprising that, given the ring of Gyges, he would behave as did the Lydian shepherd. But this shows only that he is not truly the just man. Properly understood, the just man is the person who, recognizing a certain course of action to be just, finds her feelings engaged by that recognition and so finds herself moved to adhere to that course of action because of its justice. To be sure, she may be moved in other ways as well, but for the just person, this motivation is essential” (Gauthier, p.328)[19]. Moral motivation is fundamental to acting morally but acting morally is not fundamental to action. The second question is, therefore, why act morally? Why Gyge should act just? The answer we have provided is: because he has reasons (normative, not prudential) for doing so, on pain of the integrity of his very identity as an agent, which inherently means both individually (unity and psychological health) and socially (intersubjective relations with others). So, if we adopt a Kantian conception of public reasoning (not of simple social negotiation, but, inherently, both reflexive and dialogic) Glauco's argument that the Lydian pastor is perfectly rational, since he acts following the prudential reason of pursuing his own gain, is false, for the simple fact that this is not a good reason to justify actions to others (and to ourselves). And in the context described by Plato, in which it would come naturally to us to question whether what we are doing is right, we will be forced to switch from a prudential reason to a public reason, to justify this demand. And I believe it is only by recovering such a framework of public rationality that we can truly give importance to reasons as essential elements for cooperative action.

Footnotes

[1] There is a general agreement that ethical theory bears precisely the burden of proof of explaining the objective-seeming features of our moral experience.

[2] This deliberation, this selection of what is right and what is not, is not merely arbitrary, since, as rational agents, we must submit to the intrinsic rules of practical reasoning itself. This does not mean to reduce the autonomy of reason to a deterministic application of rules already provided, but to circumscribe its area of legitimacy.

[3] This reconstruction by key phrases was taken from Morris and Singpurwalla (2013).

[4] It should be made explicit that, of course, not all moral realism positions are Platonist, quite the contrary. However, it is perhaps useful to briefly address the comparison that occurred between Scanlon (1998) and Korsgaard (1996,2021) to show how constructivism differs even from those very moderate forms of moral and epistemic realism. According to realism (about reasons), substantive reasons are independent of the requirements of rationality, and work as premises in moral reasoning (Scanlon 1998, pp. 25-32). On the constructivist view, instead, rational requirements are constitutive, that

is, they do not work as premises but as organizing norms that constitute the substantive reasons (Bagnoli 2013, 2021). For Scanlon, reasons are “relational” (2014, p. 120) and they are practically significant insofar as they bring us into an agreement with others (Scanlon 2014, p. 2). Scanlon’s emphasis on relationality is supposed to avoid the caricatural charge that moral realism presupposes the existence of normative properties “out there” and that we encounter reasons “wafting by” (Mackie 1977; Korsgaard 1996a, p. 44; Scanlon 2014, p. 120). Korsgaard is unconvinced by this response and her objection against moral realism or realism about reasons is that it takes the relations to reason to be epistemic (Korsgaard 2021, p. 175). Thus, we should expect that the purported advantage of constructivism concerns a different rendering of relational normativity, which is practical rather than epistemic. To explicate this contrast, Korsgaard presses the view that the practical standpoint is to be understood as the first-person deliberative standpoint. This is an important piece in the entire argumentation because it allows Korsgaard to establish that the normativity of morality, unlike other forms of normativity, should be expressive of individual autonomy. To function as proper reasons for action, considerations must be such that they represent the stance of agency. The question of normativity arises for and matters to agents who have a distinctive kind of subjectivity, that is, capable of self-consciousness. Thus, the source of normativity is the capacity for evaluating: “whether you have reason at all depends not on what is out there in the world, but what sort of subjectivity you have. Reasons exist in the first instance in the deliberative perspective itself. But the kind of subjectivity that is needed is not the capacity to form an objective conception of reason: it is rather the kind that Kant associated with autonomy, the capacity to make a law for yourself. The essential element of a reason is its normativity, and its normativity for you rests in the fact that you legislate acting on it as a law”, (Korsgaard 2021, p. 178; cf. 1996a, §3.4).

[5] Constructivists and Gauthier have more in common than differences. Both share the idea that moral and normative facts are mind-dependent and that they depend specifically on the "deliberative" intentional states that people would have under certain hypothetical circumstances. By "deliberative" intentional state, we can refer to an intentional state such as a wanting, choosing, or evaluating that has a direction of adaptation similar to desire and is normally subject to deliberative control (regarding what would be rational to have under certain circumstances). However, the disagreement is precisely about the criteria of this deliberative control: while for Gauthier they boil down to a cost-benefit calculation, relying on an economic and individualist conception of rationality, Kantian-minded constructivists insist more on the public nature of deliberation and rationality, emphasizing its dialogical character.

[6] To cite just a few others, from very different background: Hegel, Nietzsche and Parfit. Specifically Parfit (2011), in his critique of utilitarianism, tried to show that the view that we only have reason to do what promotes our happiness is based on a perspectival deception, that is, a deception where the propositions 'maximizing happiness is what has to be done' or 'good coincides with maximizing happiness' are thought to be informative propositions, but in fact they are not: they are as informative as propositions like 'being round is the same as being blue.' What the utilitarian does is simply to say that 'maximize' and 'ought to do' coincide, without trying to bridge the manifest gap between the two facts.

[7] Alvarez (2016) provide an example in which is completely sound to ask whether there is a reason for the government of a country to tax sugary drinks (normative) while at the same time asking for the government’s reason for actually taxing the drinks (motivating). The same reason may answer both questions: the reason that favours taxing the drink may be that the tax will help reduce child obesity; and that may also be the government’s reason for taxing the drinks. In that case, the government is motivated to tax drinks by a normative reason, a reason that may justify its doing so. But we don’t always

act for the reasons that favour our actions. For instance, the government may tax sugary drinks because (or in part because) some of its members own shares in a company that sells low-sugar drinks. In that case, the reason for which the government decides to tax sugary drinks is not, or not solely, the reason that favours its doing so. The distinction between normative and motivating reasons, therefore, enables us to separate the question what reasons motivate agents to act (a psychological question) and the question whether those are good reasons: reasons that favour and justify their acting thus (a normative question). My suggestion is that we should use the term reason only for normative reason, while simply using the term motive for the motivating reason: therefore, there can be different motives why the government tax sugary drinks, but there will be only some reasons to do that, namely the normative reasons, reasons that will be justifiable to citizenry.

[8] Constructivists hold precisely that normative reasons are mind-dependent, that is, they do not exist prior to our deliberation, and that, at the same time, they are objective and non-arbitrary, since they are based on principles that are constitutive of reason itself. By constitutive principles of reason they mean those principles that are ineliminable from reasoning itself, since they constitute it and are its foundation. Constructivism, in this sense, is delineated as a conception of practical reason that is based on the agent's reflexive skills, in which the moral principles identified refer back to the reason for which that action was chosen and undertaken. The 'not being able to avoid' reference to the constitutive principles of reason itself expresses a practical, rather than logical and ontological, necessity with which the agent must necessarily come to terms when asking "what should I do?", putting his or her integrity as a responsible and reflexive (Korsgaard 2019, Bagnoli 2019) if fallible (O'Neill 1989) agent at stake. We will delve into this issue in details in the next page.

[9] While Kant and Kantian constructivists such as Korsgaard, O'Neil, and Bagnoli (but also other contemporary non-Kantian thinkers such as Smith, Scanlon, and Millgram) ground the normativity of reasons on the concept of rationality, there are approaches that ground the normativity of reasons on the concept of good or value (from Aristotle to modern thinkers such as Anscombe, Raz, and Dancy) and approaches that ground the normativity of reasons on the concept of desire or motivation (from Hume to modern thinkers such as Williams, Goldman, and Harris). Some authors (such as Quinn 1993) reject this tripartition, arguing that all meta-ethical positions on normativity are nonetheless based on a concept of rationality, distinguishing precisely between subjectivists, who hold moral principles to be rational albeit non-cognitive (and thus e.g. preferences and desires), and objectivists, who hold moral principles to be rational and endowed with cognitive content and thus capable of genuine truth (based on objective, foundational elements of the world, such as values and the good). Quinn who bases normativity on an Aristotelian concept of good views rationality precisely as the excellence of the use of the faculty of reason, the “good” condition of rational practical thought. “The pursuit [of an action] is rationalized not by the attitude, but by the apparent value that attaches to its object” (Quinn, 1993, p.247)

[10] Dancy (2000, p.106) provide a possible response to the normative constrain by stating that acting for a good reason may simply require your motivating reason to be a mental state whose content is a good reason. So, you act for a good reason if your motivating reason for, say, taking your umbrella is your believing that it is raining, which is a mental state whose content “it is raining” is a good reason to take your umbrella. Dancy’s proposal removes the psychologism apparatus of the motivating reasons since, according to him, if psychologism is right, we can never act for a good reason.

[11] For Williams, one cannot argue with those who reject the moral point of view because they feel no commitment to

others. The starting point of moral experience is sympathetic involvement with the fate of others, and this can only be produced by education if it can awaken appropriate emotions (not reasoning). One can argue against immoralism only with an interlocutor who has manifested involvement with something. It is precisely in this sense that I try to distance myself from Williams' approach, since I believe it is possible to argue in favor of the 'moral point of view,' because ethics need not be based necessarily and solely on emotions; reason-based ethical systems are possible, in the sense that even those who feel no commitment to others can be persuaded toward it by reasoning correctly. Kantian constructivism is one such ethical system.

[12] The term public underline the contrast with Williams' conception of rationality, which is almost solipsistic, disconnected from verification mechanisms external to the subject itself. The reasons end up being merely motivating reasons or at best explanatory reasons. But justifying the action is very different from simply explaining it. If I explain the reasons why I walked away in front of the starving child, I am simply showing what was the reasoning I followed, regardless of whether it was correct or not. That is why we have to talk about justificatory (or normative) reasons and not explanatory reasons. The latter are reasons to which we appeal to try to explain actions and attitudes; the former are reasons to which we appeal to try to justify them.

[13] In theoretical reasoning this is evident, but there are good reasons to argue that this is also applicable to practical reasoning. Millgram for example points out that: " One important aspect of theoretical reasoning (that is, reasoning directed toward belief) is resolving contradictions in one’s system of beliefs, and freedom from contradiction is an important contributor to theoretical coherence. How we revise our beliefs in the face of contradiction is not well understood, but we can expect that practical rationality will also require freedom from practical contradiction, that freedom from practical contradiction will be an important contributor to practical coherence, and that practical reasoning will be directed at, among other things, resolving practical contradictions” (Millgram, 2001, p.15).

[14] I used the term basic and not constitutive voluntarily. While the principle of self-preservation is a common feature of all cultures (as shown by various anthropologists, psychologists and biologists since the time of Freud and Darwin), this does not mean that it is a constitutive principle of practical reason. It may simply be an abstraction of a survival instinct common to most animals and not a formalization obtained from the faculty of reason peculiar only to human beings.

[15] Which, again, in this case will have no major repercussions on the passerby, who is simply going home where he can have more to eat. However, it could be reiterated that my wanting to eat the hot dog is not due to the simple function of feeding myself, but precisely because of the pleasure given by the taste, the pleasurable sensation of eating. What follows in the discussion applies perfectly to this answer as well: the reason for my temporary pleasure is not sufficient to balance the condition and fate that lies ahead for the child

[16] Or perhaps another kind of rationality. If the space separating constructivists and Williams remains wide, it is precisely because there is no agreement on what should be meant by rationality, in its practical side. Can it be limited to the application of Williams' principle, or must it also concern the choice of ends? This is not merely a semantic choice, but something that affects us in our everyday actions. See Richardson (2012).

[17] The Greek word ‘akrasia’ is usually said to translate literally as ‘lack of self-control’, but it has come to be used as a general term for the phenomenon known as weakness of will, although some philosophers try to distinguish between the two (Stroud and Svirsky 2021). It is also interesting to note, since the pretest of this article is a Platonic myth, that

Socrates (as portrayed by Plato in the Protagoras) argues that it is impossible for an agent to choose a course of action which one knows full well to be less good than some alternative known to be available. Anyone who chooses to do something which is in fact worse than something they know they could have done instead, must, according to Socrates, have wrongly judged the relative values of the actions.

[18] O'Neill is, however, very critical of those abstract settings of constructivism that, in her view, depart from true Kantian ethics, which is designed directly for imperfect non-angelical rational agents; after all, if human beings were not limited and interdependent on one another, there would be no need to bind them through the institution of moral obligation. O'Neill's construction is not, therefore, based on a hypothetical consensus among idealized agents as in the various contractualist theories, including Rawls' original position: the question is not how we would act if we were perfect, but what we can accept as a reason for action, given that we are imperfect and limited agents. Certainly, however, it seems possible to abstract from the concrete context in order to deliberate, and it is not always necessary to have the direct, physical presence of all the agents concerned with the problem being posed (think of the various everyday moments when we deliberate precisely in favor of an outside agent, for example, hoping to do him or her a favor: we must necessarily try to idealize what other agents would do in that context). The original proposal I am trying to outline seeks precisely to combine Korsgaard's transcendental advantages with a conception of public reason that is, as noted by O'Neill, closer to the original heart of Kantian philosophy.

[19] With respect to this specific passage, the only point I would emphasize more than Gauthier's concerns the rarity of the just man: in my view it is more anomalous to find people lacking any sense of justice (and thus completely blind and unable to grasp the very constitutive principles of his nature as a rational being) rather than normality. However, if one looks in general at the whole of Gauthier's proposal provided in “Morals by agreement” there are many more points of distance. In general, regarding different forms of cooperation and especially recognition of the other agent, see Bagnoli (2007) for a Kantian constructivist approach.

References

- Alvarez, M., 2016 *Reasons for Action: Justification, Motivation, Explanation*, in The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.).
- Bagnoli, C., 2002, *Moral Constructivism: A Phenomenological Argument*, in Topoi, 21 (1).
- ——, 2007, *Respect and Membership in the Moral Community, in*Ethical Theory and Moral Practice, 10 (2).
- ——, 2011 a, *Constructivism in metaethics*, in Stanford Encyclopedia of Philosophy, Winter Edition.
- ——, 2011, *Emotions and the Categorical Authority of Moral Reasons*, in C. Bagnoli ed., *Morality and the Emotions*, Oxford University Press, 2011, pp. 62-81.
- ——, 2013, *Constructivism about Practical Knowledge*, in *Constructivism in Ethics*, Bagnoli ed., Cambridge University Press, 2013.
- ——, 2014, *Desideri e necessità: sull’incompletezza della ragione pratica*, in La mente morale: persone, ragioni, virtù

(Galetti ed.), Edizioni di storia e letteratura, Roma.

- ——, 2021, *The Practical Significance of the Categorical Imperative*, in Oxford
- Studies in Normative Ethics, 11, Oxford Scholarship Online, Oxford.
- Blackburn, S. (1998), *Ruling Passions*, Oxford University Press, Oxford
- Dancy, Jonathan, 2000, *Practical Reality*, Clarendon Press, Oxford.
- Gauthier, 1986, *Morals by Agreement*, Clarendon Press, Oxford.
- ——, 1990, *Moral Dealing*, Cornell University Press, Ithaca.
- Greene, J.D., 2014, *The Cognitive Neuroscience of Moral Judgment and Decision-Making*, in The Cognitive Neurosciences V (ed. M.S. Gazzaniga), MIT Press, Cambridge MA.
- Herman, Barbara (2008a) Moral Literacy, Harvard University Press
- Herman, Barbara (2008b) Training to Autonomy: Kant and the question of moral education, in Herman 2008a, pp. 130-154
- Joyce, R., 2001, *The Myth of Morality*, Cambridge University Press, Cambridge.
- —— 2006, *The Evolution of Morality*, MIT Press, Cambridge MA.
- Korsgaard, Christine M., 1996, *The source of normativity*
- ——, 2021, *Valuing Our Humanity*, In Dean and Sensen 2021, 171- 191.
- Kitcher, P., 1985, Vaulting Ambition, MIT Press, Cambridge MA.
- MacIntyre, A., 1967, *A short history of ethics*, Routledge, Croydon.
- Mackie, J.L., 1977, *Ethics: Inventing Right and Wrong*, Penguin, London.
- Millgram, E., 2001, *Varieties of practical reasoning*, The MIT Press Cambridge, Massachusetts.
- ——, 2005, *Ethics done right*, Cambridge University Press, Cambridge.
- Morris, C.W. and Singpurwalla, R. (2013). *Ring of Gyges,* In International Encyclopedia of Ethics, H. Lafollette (Ed.)
- O'Neill O., 1975, *Acting on principle*, Columbia University Press, New York.
- ——, 1989, *Constructions of reasons*, Cambridge University Press, Cambridge.
- Quinn, W., 2012, *Morality and Action*, Cambridge University Press, Cambridge
- Richardson, H.S., 2012, *Practical Reasoning about Final Ends*, Cambridge University Press, Cambridge.
- ——, 2018, *Articulating the moral community: Toward a constructive ethical pragmatism*, New York: Oxford University Press.
- Scanlon, T. M., 1998, *What We Owe To Each Other*, Belknap Press, Cambridge.
- Stroud, S. and Svirsky L., 2021, *Weakness of Will*, in *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), Edward N. Zalta (ed.).
- Williams, B., 1973, *A Critique of Utilitarianism*, in *Utilitarianism: For and Against*, Cambridge University Press, Cambridge.
- ——, 1979, *Internal and External Reasons*, In Ross Harrison (ed.), *Rational action*. Cambridge University Press. pp. 101-113.