

Moral Reasoning. Moral Motivation and the Rational Foundation of Morals.*

Luz Marina Barreto

Universidad Central de Venezuela

Abstract

In the following paper I will examine the possibility of a rational foundation of morals, rational in the sense that to ground a moral statement on reason amounts to being able to convince an unmotivated agent to conform to a moral rule - that is to say, to “rationally motivate” him (as Habermas would have said) to act in ways for which he or she had no previous reason to act. We will scrutinize the “internalist’s” objection (in Williams’ definition) to such a claim and confront it with a, however, obvious fact: our need to convince or, at least, to influence agents to act according to moral judgements that entail strong validity claims. Therefore we need to find a better ground for our moral norms than that provided by a narrow concept of practical reason, which ultimately offers the basis for the internalist’s intuitions.

Classical philosophers believed that to have a “right” sense meant also to *think* accordingly, that there was no intelligence without a social or moral side and that a person who lacked moral sense was not, at the same time, very intelligent. This contradicts some deep Humean intuitions, and in any case requires a more complex concept of practical reason than the one we normally encounter in internalist moral philosophers. In the following pages, I would like to reflect on this kind of social reason, or intelligence, or moral rationality, in an attempt to couple, in a coherent vision of moral sense within the framework of practical reason, intuitions that appear in Kant and Adam Smith with empirical research about what the lack of moral sense could mean.

* For helpful comments and suggestions on an earlier draft, I am most grateful to Monica Betzler, Thomas Wallgren, Rebecca Beke and Mark Gregson.

Introduction

Moral judgments seem to imply a strong truth or validity claim. When we say "an action is morally wrong", the predicate 'wrong' refers here to something about which we would not want to make any concessions. In ordinary language, such judgements are normally understood as strict demands. They are not like the suggestions of rational agents of the type "I would not cut carrots with this knife if I were you", nor are they simple expressions of intense personal feelings or dislikes.

When we say that "to torture people or sentient beings for the pleasure of seeing them suffer is morally wrong", the point is not that we do not *like* it or that we may try to understand those who do such things. Rather, we feel that no-one is allowed to torture other beings, and that everybody should be acquainted with the feelings of repulsion that these actions arouse in those who endorse the norm sincerely. It would be odd if someone asked for explanations when we forbade them to molest children or abuse or humiliate people. This strong normativity is the distinct mark of moral judgments, precisely what distinguishes them from mere conventions and other imperatives. It obliges in a way that bears no resemblance to other kinds of practical recommendations, particularly in modern societies.

This powerful normativity suggests that there are good arguments for the strong feelings of obligation associated with a moral norm. However, is it possible to find such convincing grounds in modern societies? This proves difficult within a realistic or objectivistic framework, since a rational agent may very well derive satisfaction from despicable attitudes. It is an undisputed fact that people can be, and throughout the ages have been, delighted by things that appear to us almost unspeakable, such as the kind of pleasure experienced by Roman citizens in their public games. On the other hand, if we shift our perspective and try to look for the sources of this strong normativity inside our subjective nature, we run the risk of diluting it into mere psychology: the moral attitude would merely be the remnants of an earlier socialization process that took place within a moral

community. In consequence, no strong validity claim could be implied here and no rational agreement among equals could ever be justified.

In the following pages I will examine one of the attempts that has been made to find a solid basis for our moral imperatives - the rational foundation (the others are based on religion or tradition). It is important to bear in mind the Habermasian definition of rational foundation, in the sense of his concept of communicative action, in order to fully understand the scope of this attempt. According to this definition, to ground a moral statement on reason amounts to being able to convince an unmotivated agent to conform to a moral rule - that is to say, to rationally motivate him to act in ways in which he had no reason to act previously. However, we will see that, according to the internalist, this type of rational motivation is, by definition, impossible within the conceptual framework of a standard theory of rational choice, for to do something in this sense means simply doing what we want or prefer. This explains why the attempts to ground the strong validity claim of moral judgments on the standard or narrow concept of rational choice have not yielded many encouraging results.

No doubt we would all be happy with this situation if it were not for the powerful validity claim that accompanies moral judgments. We need to convince unmotivated agents because we do not want to be their victims and we do not want our loved ones to be their victims. A world where we cannot show the truth of our strong moral convictions, such as not torturing people or animals, does not seem to be right. This is why the internalist contention of moral motivation is so unsatisfactory in the end. We need to find a better ground for our moral norms than that provided by a narrow concept of practical reason. Furthermore, we would like to be sure that we are not deceiving ourselves when we state that moral norms are followed, not because we are afraid of feelings of inner sanction, nor because we are simply unable to abuse or despise sentient beings, but because it is correct or right to abide by the moral imperatives in the way we do.

It seems, then, that we need to reflect seriously on the possibility of grounding morals on reason. Classical philosophers believed that to “feel” well,

i.e. to have a “correct” sensibility, meant also to think well, that there was no intelligence without a social or moral side and that a person who lacks moral sense cannot be, at the same time, very intelligent. This contradicts some deep Humean intuitions, and in any case requires a more complex concept of practical reason than the one we normally encounter in internalist moral philosophers. Therefore I would like to reflect on this kind of moral reason, or intelligence, in order to find a link between intuitions concerning the nature of practical reason with empirical research about what the lack of moral sense could mean.

For to simply postulate a wider concept of practical reason or rational deliberation on practical judgments would not be enough. To say that the moral stance supposes not only the right affective attitude but also that this attitude might result from, or might complement, reflective or rational powers of the sort that are not normally associated with a healthy moral character is an empirical hypothesis. I will also, therefore, present some empirical evidence concerning the lack of moral sense, and I will use this to support the need to construct a better concept of practical reason. Thus the philosophical analysis of the concept of a moral rationality aims fundamentally at suggesting to the moral psychologist and to the neuroscientist a possible direction for empirical research.¹

I.

It has been said that modern moral discourse is mainly characterised by the aspiration to find a rational foundation for morals. It explores the possibility of social norms and received traditional obligations being firmly grounded in rational argument. Its aim has been to replace old traditional views with arguments based on universally accepted traits about human nature and the good of people, which for the universalist should be evident to everyone. However, for those of us who are continually horrified by the amount of abuse and crime imposed on people by some who do not seem to share with us a common notion

of human decency, defining the meaning and perspectives of a rational foundation for morals has become increasingly challenging and problematic.

Unavoidable ambiguities in the concepts of practical reason, motivation, validity or truth, even in the meaning of moral obligation and constraint, have diluted the once noble aspiration to create a space for the transparent discussion of practical norms and to reach a final agreement based on arguments accepted by everyone. Instead, two different ways of eluding the whole task and abandoning its aims have emerged. The first consists of divorcing reason from motivation and makes the latter dependent on affective features that are supposed to be unrelated to the reflective and rational powers of the human mind. The second attacks the notion of reason as such, engulfing it in motivation and thereby surrendering all attempts to reach a rational agreement on practical issues. This allows the agent to abandon himself to irrationality, to mere force or to changing moods that are perceived as not needing justification.

It is also difficult to figure out what it means to produce a rational foundation for morals. Those who believe in the “rational motivation”² of an interlocutor think it is possible to find rational criteria for the assessment of the truth-value of a practical judgement, as well as criteria for making sense of someone else’s actions. In contrast, from the point of view of the relativist, it is not only impossible to share rational criteria for the assessment of action judgements, but to be rational has an entirely different meaning: it merely conveys the private interests or preferences of an agent and thus contains nothing that could be shared on an inter-subjective ground. In Hume’s terms³ rational choice does not refer to the ends of an action, but only to the means we

¹ As analysis of basic concepts, the philosophical task is precisely to provide the theorist with possible means for a critical and constructive exploration of an object domain. See, for instance, P.F. Strawson, 1992, *Analysis and Metaphysics*, Oxford University Press.

² For instance, Jürgen Habermas. The fact that, for Habermas, the foundation of moral norms links normativity (*Normativität*) and rationality through the notion of communicative understanding processes (*rational motivierende Verständigungsprozesse*) that may rationally motivate a previously unmotivated interlocutor is crucial to his concept of practical reason. See, for instance, his *Theorie des kommunikativen Handelns*, Introduction, Suhrkamp Verlag, Frankfurt am Main, 1981, his *Erläuterung zur Diskursethik*, Suhrkamp, Frankfurt am Main, 1991, or his *Faktizität und Geltung*, Suhrkamp, Frankfurt am Main, 1998. The German phrase I quote here appears in this last book, p.20. I have extensively analysed this aspect of Habermas's thought in my *El lenguaje de la modernidad*, Monte Ávila Latinoamericana, Caracas, 1994.

decide to employ in order to achieve what we want. Therefore, if someone wants to incur some form of self-sacrifice in order to follow moral duty, it would not make much sense to call their action “irrational” as long as they do what they consider they want to do, after sufficient personal deliberation.

According to a narrow, standard theory of rational choice, a rational agent follows personal preferences that need not be universally accepted. These preferences or ends are supposed to be stable and, since they are assessed by economic theory, there is no common measure to compare the diverse utilities deriving from their fulfilment. This is bad news for the universalist because it makes a rational agreement on motives or preferences - in the case of moral actions - impossible. Although such actions seem valid in an inter-subjective sense for the agent who endorses them sincerely, a narrow concept of practical reason makes the universality of moral respect towards others implausible.

Let us examine why. We are indebted to Bernard Williams and his conception of moral internalism for a short and precise description of the problem at hand. Because he identifies motives for actions with reasons, as must be the case within the standard or narrow theory of rational choice, he denies the possibility of persuading or of rationally motivating an interlocutor to act according to a moral norm in the absence of a pre-existing “internal” motivation to do so. For if I do not have a motive to enrol in the army, nor will I have a *reason* to do so, and therefore no amount of “rational discussion” will persuade me to do something for which I have no reason. In contrast, the opposite conception, which Williams calls “externalism”, stresses that it is possible to convince an unmotivated interlocutor to follow a moral prescription if one can show that it is in the interlocutor’s best interest to conform to the rule, even if he is not previously motivated to do so. In this case, a moral norm will be true or valid (so states the philosophical tradition) if it is “rational” in this “external” sense. The recommendation would be a reason not in the internal, private way,

³ See David Hume, *Treatise on Human Nature*, Oxford University Press, 1990, p. 416.

but in a way that is susceptible of inter-subjective agreement.⁴ In this sense, reasons are not motives, but good reasons for every possible agent.

According to the rational theorist or, as Williams calls him, the externalist theorist, a practical prescription (or a moral norm) is true if and only if it is always in your *best interest* to comply with it, even if you do not *want* to do it - that is, even if you do not find now, in your particular system of motives, a reason to comply with it. We clearly have here two interpretations of what it means to be “rational”. On the one hand, we have an objective or externalist point of view, and on the other, a narrow conception of reason, where the rationality of an action depends on the requirements it serves. These are agent-centred: the value of a certain action rests on what the agent considers valuable or good for him.

This idea of reason is very persuasive. The notion of rationality normally refers only to what satisfies an agent’s preferences. Therefore, the value of an action is relative to the agent’s set of interests, preferences or motives. The problem is that such a conception of reason is at odds with that special feature of moral action we pointed out before: its apparent universalist character. As a consequence, the internalist theorist is forced to justify the validity claim of an obligation without resorting to reason, understood as the capacity to offer universalist criteria leading to inter-subjective agreement about the good of an action. This being the case, the foundation of a moral attitude is not sought in the rational arena but somewhere else, most probably in the psychology of the moral agent, since we need to explain empirically why a particular person does not seem to share with other moral agents what we consider relevant motivations. This could explain why the internalist theorist regards moral motivation as something affective and discards cognitive aspects that could be the object of rational agreement within an argumentative discussion aiming at some sort of theoretical truth. And this could also explain why we tend to separate the idea of a “genuine” moral motivation from a “rational”, self-

⁴ See Bernard Williams, “Internal and External Reasons” in *Moral Luck*, Cambridge University Press, 1980.

centred motivation. It is our narrow conception of practical reason that justifies the emphasis we put on being in possession of a moral sense as a condition for a rational agreement on moral norms.

Exiled from the philosophical discussion about the nature of a foundation for morals, the theory of rational choice is abandoned by moral philosophy in favour of a reflection centred on the nature of moral duty and, ultimately, on moral personality, where to be a moral person would merely refer to the capacity to experience moral feelings. Like Hume, the internalist also thinks that a moral person is not more reflective than others or more open to rational deliberation, but has a better or more “temperate” character.⁵

However, in the last few decades, several lines of argument have been explored in order to expand the standard theory of rational choice to a less “irrational” - that is, closer to a universalist idea of the sources constituting the validity of a moral norm - notion of moral motivation. Or, to put it another way, in order to reconcile these opposing views concerning the normative content of a moral rule, so that we could have a weaker notion of the rationality of an action, expressing motives and ends which an interpreter can reconstruct, and a stronger one, where we *demand* from the agent certain motives and ends towards other living beings. The normative content we expect from a moral action is therefore much stronger than the content we expect from a morally neutral action, such as which flavour of ice cream I will ask for next, for nobody would mind if I were to choose vanilla instead of chocolate but they would find it strange if I asked for a steak-flavoured ice cream.⁶

⁵ Hume, op. cit., p. 417 and 418. Also Bernard Williams's conclusions in *Ethics and the Limits of Philosophy*, Fontana Press, 1985 and Simon Blackburn's remarks in his *Ruling Passions*, Clarendon Press, Oxford, 1998: “The differences between the wanton and a member of the party of mankind are distinctions within the brood of dispositions, and have nothing to do with the authoritative exercise of the moral law, nor with the true operation of autonomy” (p. 251).

⁶ The idea that there is an implicit normative assumption when we assess actions has also received special attention in Simon Blackburn, 1998. But his notion refers to a weaker sense of what it means to conform to a normative structure. In Blackburn's sense, in order to be intelligible an action must reflect at least some of the wide range of practical norms. However, the moral demand is not an expectation of intelligibility in this sense, but a strong demand to conform to what we expect from a moral person. We do not only want to understand a moral agent, but demand from him certain attitudes, and therefore some strong values.

The standard theory of rational choice, which is a narrow version of rationality since it does not prejudge the universal validity (or truth) of the ends and preferences at issue, works well under conditions of certainty - that is, when the agent knows what payments or utilities to expect from specific actions; or within a parametric framework, when the agent does not have to face other people's decisions or does not have to lessen the impact of other actions on his eventual benefits. Things become complicated when the payments or utilities that you expect are or can be affected by other agents' actions. There are several models of strategic interaction that specify which are the most rational strategies for someone who has to interact with other "players". The results can be assessed either from the point of view of the strategies themselves or from the point of view of their expected utilities. From the point of view of the strategies, a result is a *Nash equilibrium* if the strategy maximises its value, that is, if it is the best strategy, given the strategies of other players. From the point of view of the payments or utilities, a strategy yields optimal results (or is a *Pareto's optimum*) if it maximises the utility of the agent given the expected utilities of other players. The ideal situation is when the strategies in equilibrium produce optimal results. However, there is a game where this coincidence does not take place: the Prisoner's Dilemma. Here we have only two strategies: either to defeat the opposite party or to cooperate with them. A Nash equilibrium is possible when all parties decide to defeat the counterpart. However, the optimal payment results not from defection or fraud but from cooperation. Hence the dilemma: we have here a conflict between the best strategy and the best result. Or, in simpler words, the best rational strategy yields the worst results for the rational agent, whereas the best one is the most "irrational".⁷

There is another way to show the same point and this is the one that inspires Rawls and other contractualist social theorists. The main idea is that the rational agent wants the best results he can attain given the outcomes of others. But, as happens in Prisoner's Dilemma, the best strategy seems to be to exploit

⁷ For extended discussion see Gutiérrez, Gilberto, *Ética y decisión racional*, Editorial Síntesis, Madrid, 2000, p.131 ff.

the cooperative dispositions of others. The problem is that if everybody took a free-rider attitude, exploiting others for their own benefit, this would be worse for everybody. In the social contract, for instance, if everybody behaved as in the state of nature, they would all lose the better outcomes that result from cooperation, as well as the possibility of fulfilling their own conception of the good without being interfered with by others. Here we have a coordination problem that could only be solved if those implicated in the social contract gave up some of their freedom in order to better fulfil their ends, providing that other agents did the same. In Rawls's model, this strategy gives rise to two principles of justice that allow all agents to reasonably pursue their particular rational plans of life without being obstructed by those of other people. Thus the same idea applies to the Prisoner's Dilemma as to the social contract: if someone has to interact several times with the same individuals, it is better to avoid the unbridled pursuit of their own satisfaction in order to enhance the chances for mutually beneficial cooperation.

Even under this interpretation, however, the contractualist model turns out to be unsatisfactory as a complete theory of moral motivation. For, even if you know that you have to interact several times with your partners within a contract, there is no straightforward argument that favours the abandoning of the relative advantages you may have as a rational agent. These relative advantages do make a difference regarding your position within a society ruled by a social contract. Of course, you may say that a society that applies social rules differentially is less efficient. Or "unfair", as they may hinder you from following your personal conception of the good. Given the uncertainty of human affairs, by which you may or may not enjoy your personal advantages over a long period of time, it seems advisable to prefer principles of distributive justice that ignore your personal assets or have been founded "under a veil of ignorance". But if you happen to know which are your personal advantages and have a reasonable expectation about how long you will probably enjoy them, it does not seem rational to avoid using them when faced with a good opportunity.

In fact, the role that this abandoning of advantages plays in a supposedly rational theory of the moral attitude is still unclear. This is why Rawls's critics claim that you would only agree to ignore your personal assets in favour of a general rule of distributive justice if you already had some kind of pre-existing moral motivation to do so. Perhaps, if you did not know what your relative advantages were, you would prefer to abide by a general or fair rule. But since we normally do know, it is impossible to show why someone who does not already have an internal motivation would prefer to favour a general conception of the good rather than their own.

An interesting instance of a position that reconciles both options, an internalist with a rational foundation, is David Gauthier's theory of a rational moral.⁸ Gauthier's initial idea is that, as we have already pointed out, Rawls's foundation of principles of justice cannot legitimately be called "rational", for it has been reached only after an agent has ignored his personal assets or relative advantages regarding other agents. In Gauthier's view, your acquiescence to the conditions of the contract - and your acceptance of the outcomes of the argumentation process which has taken place under conditions analogous to the original position - has to be the result of a genuinely rational strategy. Furthermore, the principles of justice, as well as moral norms, need to be considered in the light of a rational strategy which should be accepted, as Rawls suggested, as part of a theory of rational choice. None of these conditions apply if, on the one hand, you are already prepared to consider - from the beginning, and as a potential member of a situation similar to the original position - your personal interest within the framework of a general rule; and if, on the other hand, you are also prepared to abide by the conditions produced by the contract even when - once the veil of ignorance is removed - you are endowed with personal advantages.

The intuition here is that no rational person will ever be willing to consider the personal interest of other agents as important as her own interest.

As we have already pointed out, rationality means to regard value as agent-centred. To suppose that a rational agent might be willing to comply with rules that apply to all rational agents in the same measure is another way of saying that there are some “objective” values, in contrast to value created by a personal and subjective set of preferences and motivations. Therefore, Gauthier’s challenge is to deepen the rational character of the contract by removing all conditions that would suggest other motives for complying with the rules produced by it, different from the rational pursuit of each person’s own interest. This is why Gauthier regards a moral system as a rational strategy.

There are, in principle, two ways of deriving cooperation from a theory of rational decision. The first depends on the well-known argument concerning interactions that extend through time: given more than one possible interaction, a rational agent will try to build up trust among the parties so he can guarantee their future cooperation and, therefore, the fulfilment of his long-term personal interest.

The second - Gauthier’s argument - is more subtle. According to Gauthier, even if you have to interact only once with a group of potential co-operators, you will lose the opportunity of reaching an optimum personal outcome if you do not give up your selfish, exploitative stance, in favour of a cooperative one.⁹ Thus this is not simply a question of the reputation you may attain if you do not defeat others, but of realising that if the others are prepared to cooperate - that is, if you expect others to cooperate with you and if you know they are capable of identifying your willingness to cooperate - then you will achieve a better outcome from this conditional cooperation than if you abstain from cooperating by taking a selfish stance, for you would be expelled from the contract and therefore you would lose the potential gains that can only be obtained among a community of co-operators. The next step is to show that cooperation can be the

⁸ See Gauthier, David, 1998: *Egoísmo, moralidad y sociedad liberal*, Paidós. See also his *Morals by Agreement*, Oxford University Press, 1986.

⁹ Gauthier, D., “El egoísta incompleto” (“The Incomplete Egoist”), in Gauthier, 1998. Originally published in the *Tanner Lectures on Human Values*, Vol. 5, University of Utah Press and Cambridge University Press, 1984.

best rational strategy for agents preoccupied only with their own interest. Or, as Gauthier has stated, that rational selfishness, confronted with the chances of gain offered by cooperation and mutual trust, is inconsistent and self-defeating.

But the willingness to cooperate with and trust others requires you to have a personal stance that lasts over the years. In other words, it requires you to have some particular psychological traits. At this point you may distinguish moral motivation from rational motivation by insisting that being a person with a moral character (that is, being a person with genuine moral motives in your subjective motivational set) is different from merely acting on the mandate of a reasonable recommendation. For if you were this kind of person, you would not be tempted to break the agreements reached with others (and, furthermore, you would also want to reach an agreement and cooperate with others, thus avoiding the perils for your own welfare brought about by rational self-interest). Or you may, as Gauthier does, try to couple the convenience of being a moral person with a rational prescription by suggesting that reason commits you to developing a long-term moral attitude. Gauthier's solution to the dilemma posed to reciprocal cooperation by a rational attitude (where rational means only to maximise your own conception of the good - that is to say, to maximise the benefits resulting from the satisfaction of your personal values and preferences) entails showing that a rational person may well try to avoid the self-defeating aspects of a selfish frame of mind by acknowledging that she would better serve her personal interest through cooperation with others.

However, Gauthier's argument has the following flaw: he points out that if a rational agent knew that others would cooperate with him and if they knew that he was a potential co-operator, then he would prefer to cooperate in order to increase his benefits. But this argument cannot be derived from situations resembling the Prisoner's Dilemma: in fact, the rational thing to do for someone who is playing games of this sort, or faces similar situations in real life, is to exploit the other person if he knows with certainty that the other will cooperate.

Of course, you may say that someone who has to interact several times with another person needs to build up trust, but again, this kind of temporal argument was the one Gauthier was precisely trying to avoid, since he wanted to show that it is a better strategy to cooperate with potential co-operators even if you have to do it just once.

There is only one situation, if we really follow the lesson resulting from playing Prisoner's Dilemma, in which you would cooperate and in which this would be a rational action: this is when you do not know what the other person has chosen or will choose, because once you know what she has chosen - that is to say, once you know that the other has either confessed or not confessed, or cooperated or not cooperated - your rational action in either case must be (if you have to interact once and only once with this person) not to cooperate.

In consequence, Gauthier's argument must be: cooperate when you do not know what the others have done. If you, by chance, land in a collective situation where you find yourself having to interact with others who you have not seen before and will never see again, cooperate: you may well find yourself in the kind of interaction where, if you refuse to cooperate with others, you will miss the optimum outcome (as in Prisoner's Dilemma).

But having to cooperate in uncertain situations - that is, when you do not know how others' decisions will affect your benefits - implies precisely that you already have a subjective moral motivation. What rationality, then, is advising you to do, since having a selfish attitude can be self-defeating, is to BE a moral person. Reason seems here to recommend having a sort of cooperative disposition, a genuine moral motivation in what Williams called one's subjective motivational set. This is how Gauthier ended up endorsing the internalist point of view, when he was trying to argue in favour of the opposite, externalist position.

So, if you find yourself in the sort of situation where you do not know what others have done or will do, it could be psychologically impossible for you to pretend to be a trustworthy co-operator if you are not already this kind of

person. And if you are not, you won't be able to make sincere promises to other people so that they can really trust you in the future, in which case you will reduce your chances of promoting your personal well-being. But, on the other hand, if you are able to promise to cooperate sincerely, then you are the kind of moral person who did not need a rational argument to begin with.

II.

In the rest of this paper, I will explore two aspects of Kant's concept of practical reason that might contribute to a better understanding of what it means to be a truly rational person. In doing so, however, I will leave aside the most metaphysical features of Kant's notion of autonomy. My aim is rather to highlight the capacity for self-reflection and the sense of being restricted by moral law that seem to be implied by the notion of a rational agent - if we understand him as an empirical and not noumenal self - and to make the resulting image of a rational agent plausible through empirical data concerning the lack of moral sense.

Kant's notion of practical reason has frequently been used to support the case for a rational foundation for moral imperatives. Kant believed that, as a practical rule, to be grounded on reason meant to pass a universality criterion, in very much the same way that a statement referring to an objective state of affairs needs to awaken some kind of general agreement to be considered true. But for Kant the universality of a scientific statement did not float in the air but rather rested on the unity of understanding and perception - that is, on common features belonging to the human cognitive framework. It is this unity of the transcendental subject that supports the consent around a given assertion about a state of affairs. This is why the model of a moral norm is the categorical imperative. However, while it is clear that, for a given objective statement, pure reason offers the necessary cognitive framework that sustains the truth claim of synthetic judgements, where is the truth claim to be grounded in the case of practical rules? Is there a common framework, shared by all "rational beings", as Kant often said, which would guarantee the unity of criteria needed by a

practical rule in order to be considered universal, or which could ground a presumed universality claim concerning practical rules?

If a practical rule is to command the same kind of unanimous consent that characterises statements resulting from the formal or natural sciences, it has, above all, to override my personal tastes or dislikes with its persuasive force. Rational autonomy means, for Kant, that the human will knows itself to be restricted by a sense of duty. It is only through the internalised moral rule that a rational agent realises his condition as a genuinely free agent - that is to say, genuinely free to restrict himself, to abstain, from doing what he wants when he knows what he must do. It is only through the feeling of being utterly restricted by duty that the rational agent discovers the freedom of his will as an absolute framework within which the validity of a norm expresses itself. This kind of freedom is for the philosopher a reverse image, an unending possibility for the agent who wants to assume it entirely. As such, the autonomy that the moral rule uncovers does not show itself completely; it appears before him not as a blank page, but like a territory one needs to explore. It reveals all its potential only as the agent finds himself to be limited and restricted by the moral rule. Then he discovers that he is free to follow or not what is demanded of him by his sense of duty.¹⁰

This is the reason why a moral norm, which has the characteristic of revealing my freedom not just to do what I want, cannot be considered a means to a personal end. And, therefore, the usual contractualist argument only finds a secondary place within the Kantian system. Kant was keenly aware of the objections that would have to be faced by a foundation for morals based on an agent-centred notion of value. In contrast, what he had in mind was a practical rule whose universality was capable of overriding my personal wants or preferences, making me respect others' feelings and preferences even if they collided with mine, and which would express the feeling of being strongly restricted by moral law.

¹⁰ See *Kritik der praktischen Vernunft*, A 53, p. 139.

The Kantian notion of freedom of will, which begins with the awareness of the restrictions imposed on us by moral law, is very interesting, for it depicts a persuasive image of rational freedom as such - that is to say, detached from its roots in morality. According to economic thought, for instance, to be a rational agent means always to incur opportunity costs: each course of action has a price insofar as it deflects scarce resources (in particular, valuable time) that could have been invested in alternative courses of action. This is the so-called Law of Diminishing Marginal Utility, a basic assumption about human attribution of value: the more one has of a good, the less an economic agent will value additional units of it, for with each new unit the opportunity costs increase.¹¹ C.K. Chesterton, in some memorable pages, sheds light on the same aspect of the freedom of will - that which is discovered and experienced in the face of our limitations, moral or mundane. Every intentional action implies a sacrifice willingly made. Moreover, for Chesterton, as well as for the economists, every act of evaluation implies a sense of the limited means we have to enjoy life. A rational agent who thought he could have it all, with no restriction, would not simply be “selfish”, he would be wrong from an ontological point of view. This sort of agent does not exist. As Chesterton convincingly showed, the notion of a rational agent, as an individual dominated by the will for power, is a metaphysical invention.¹²

A moral norm, according to Kant, has a universal validity claim that should be grounded on reason, like an assertion emerging from natural sciences. Kant's notion of practical reason offered us the possibility of finding not only a rule that meets the criteria of universality required by a moral norm (the categorical imperative), but also the common ground whose features, shared by all rational beings or, as we would say now, by humankind, would guarantee the agreement that supports the practical rule. Modern contractualists, as we have already seen, have also tried to find a rational foundation for morals stemming from an

¹¹ See, for instance, Lionel Robbins, “The Nature and Significance of Economic Science”, in Daniel Hausman (Ed.), 1994, *The Philosophy of economics*, Cambridge University Press.

¹² Cfr. C.K. Chesterton, 2000, *Ortodoxia*, Editorial Alta Fulla, Barcelona.

agent-centred notion of value. We have also already seen the limits of these attempts. However, Kant's idea, which I think is worth considering, is that the concept of an agent-centred notion of value implies *analytically* some features of rational agency that would support the case for a moral consideration of rational beings, and would not imply, as Kant intended, a straightforward reference to moral feelings or personal inclination unrelated to the regular, rational and self-centred pursuit of happiness.

This idea states that rational nature is characterised by the fact that it has to choose. At the same time, the objective limitations that force us to choose offer the framework for all rational attribution of value, despite the fact that there are many different kinds of consideration determining the criteria for the process of choosing (in Kant's case the restrictions are basically moral - an internalised moral law). But the basic fact is this: we have to choose amidst restrictions imposed sometimes by objective circumstances, sometimes by a moral law, and at other times simply by other choosers. To be a rational agent means, ideally, that we decide for ourselves our ends and choose what we consider most valuable among several alternative courses of action. To be rational, then, points to the fairly obvious fact that we usually decide what we would like to do and how we are going to do it. Of course, not everybody enjoys this sort of autonomy - we will come back to this in a minute - but, in general, to be rational is to value something that we, as rational beings, consider valuable. Hence the problem of those who despair of trying to ground morals on reason, since different rational agents value things differently depending on their personal preferences.

Kant was aware that from the mere fact of our condition of rational agents with personal ends it does not follow that we should respect everybody's ends to the same degree. But Kant knew that once you accepted as satisfactory the agent-centred definition of rationality, you also had to acknowledge that the crucial aspect of this notion is that a rational agent has chosen or may choose his ends with autonomy and that this autonomy implies a sense of limitation. To take

this step you only need to define rationality as having chosen certain ends, for whatever reason. This is why there is no obligation to respect everybody's ends, because what makes them valuable is only that a rational agent values them. To put it simply, no end, just because it is someone's end, is valuable in itself, and Kant insists on that in several passages of his *Groundwork*.¹³

In consequence, if there is a ground for respecting someone's ends it is not because of the ends as such, but because they are a rational agent's ends, the ones he has decided he wants and values. Therefore we do not value every end but rather an end in itself, which is the ground of every end, that which makes an end something valued. This is the core of the third formulation of the categorical imperative. But, we may still ask, why should we value above all the person who makes attributions of value? I think that what Kant had in mind was a twofold idea of the meaning of rational agency. Firstly, that each rational agent, and therefore *particularly me* as a rational agent, cherishes and knows what is to have the freedom to make attributions of value. Secondly, that I will value this capacity in me and in other rational beings if I am aware of the sense of limitation and restriction on other beings that it entails. To be a truly free agent inherently involves the gaze of the other upon me. This is why, for Kant, moral respect begins with respect towards ourselves, an aspect of moral sense that, by the way, has been neglected by modern interpreters of Kant's practical theory.¹⁴ Kant's idea of practical reason offers, therefore, not merely the *form* a rule has to exhibit to be considered valid for an agent (the categorical imperative), but also a concrete idea of what it means to be a rational agent, an idea that is implied analytically in our condition of agents, even if we might sometimes make wrong choices.

Now let us explore the meaning of this for the understanding and constitution of a moral motivation. Each of us knows what is to value something and what is to have access to the means that enable us to reach the goals we have striven for. This is what Rawls called in his *Theory of Justice* the

¹³ See, for instance, *Grundlegung zur Metaphysik der Sitten* BA 68, p. 62, footnote.

“Aristotelian motivational principle”, or the view that is contained in Ronald Dworkin’s “challenge model”, in which having the opportunity to lead the life that one values, regardless of the impact an action would add to one’s own welfare, is what matters. In the challenge model what counts is performance - that is to say, to pursue an end for the end itself and not because it would have an impact on anyone’s life.¹⁵

Dworkin’s distinction between a model of impact and a model of challenge, whereby the first stresses the potential objective value of a rational agent’s set of ends and the second a view of human life as a challenge for the rational person who wants to do what he thinks is worth doing, casts light on two essential, albeit sometimes confusing, aspects of rationality. Let us recall Williams’s argument against the rational foundation for morals: that there is no reason to follow a prescription emanating from someone else’s motivational set. But it is one thing not to be able to recognise as valid another person’s ends and another, very different thing, not to be able to identify oneself with the enjoyment - that is to say, with the sense of pursuing something that one considers important - that comes from simply being a rational agent.

Now, if we see Kant’s rational foundation of morals in the light of something analogous to Dworkin’s model of challenge or Rawls’s Aristotelian motivational principle, we can understand why Kant insists on the third formulation of the categorical imperative - that you have to treat human beings as ends in themselves and not merely as means - in order to find a universal support for moral recommendations. What ought to be valued, in Kant’s view, is the intrinsic value of rational activity as put forward by rational agents and not the relative values of single ends. And you will want to do this not simply because, as a self-interested agent, you want to be given free rein to pursue what you prefer (for this would take us back to the contractualist argument of the sort advanced by Gauthier), but because you know that being a rational

¹⁴ See *Grundlegung*, BA 82, p. 71.

¹⁵ See Dworkin, Ronald, 1988, *Foundations of Liberal Equality*, The Tanner Lecture on Human Values, Stanford University Press, University of Utah Press, p. 57.

agent actually means that what you essentially value is not merely your single, contingent ends but the freedom, the autonomy, you enjoy as such.

Kant suggests that this conclusion follows from the notion of rationality itself. I think we could agree with him in that to be rational also means to be reasonable -that is, to be able to reflect about the things we prefer or value and, above all, about our condition as agents who make attributions of value. Therefore, if we think, as Williams does, that to be rational does not necessarily entail respecting other people's prescriptions, it nevertheless implies the capacity to turn the attention towards our own condition as agents who want the freedom to make attributions of value. But now: can this capacity support a form of moral respect towards others? Should I respect others just because I am the ground of the things I value?

The internalist view, and its expressivist interpretation, rejects rational foundations for morals on the grounds that our usual notion of rationality supposes an exploitative attitude in relation to other agents that cannot be reconciled with the selfless stance that characterises morality. But in Kant's view, to be rational does not simply entail the willingness to use other things and other persons as a means, but rather a capacity to make attributions of value within an environment characterised by inner and outer restrictions. Within this capacity, we are the highest end, precisely the kind of end that gives its own ends their sense of importance, while at the same time conscious that we are framed by objective circumstances and other choosers. Now, Kant believed that when we see ourselves in this light, we are able to see others in the same way. Of course, he could not prove it, as I cannot prove it now, because this is an empirical assumption: what it states is that to be rational means to be aware of one's own capacity to make attributions of value, and this kind of awareness lets us imagine how important it is also for others to have the freedom to choose their own ends. With this interpretation, I think we can understand why Kant thought that a foundation for morals follows from the concept of rationality. Moreover, a subtler exploration of the meaning of being a rational agent might

produce a better understanding of the origins of the feelings of mutual sympathy that constitute an important aspect of moral motivation.

It is, however, Adam Smith who postulates imagination as the link that leads one subjectivity to recognise itself in another. For Adam Smith, imagination and sympathy are two distinct steps in the constitution of a moral sense and the former, as in Kant, is an attribution of the rational agency. It emerges initially from the gaze we cast on how we would feel under circumstances similar to those in which another rational agent might find himself. When we see another's grief with sympathy, states Smith, it is not that we abandon or forget ourselves, but rather that we return to ourselves to see how we would feel if we were them. And the criteria we use to assess our feelings derive from our condition of agents that make attributions of value in order to fulfil our conception of the good life, for we ask ourselves whether such unfortunate circumstances would also make us unhappy or not and to what extent.¹⁶

For the majority of internalist moral philosophers, having been socialised within a moral community is a precondition for the sort of reflectivity that directs on oneself the attention we might receive from another - the impartial spectator. But, even if we may not be so sure about what its origins were, we would be well advised to avoid reducing it exclusively to the moral stance. It seems to me that, as Smith suggested, our capacity to see ourselves as agents that value our happiness is an aspect of rationality itself. It is closely associated with the theoretical meaning of truth: that which transcends our mere subjectivities. Proper rationality demands the disposition to see ourselves as limited, along with other rational beings.

We have here, then, several properties of rational deliberation that are not only inherent to the moral stance, but also constitutive of the kind of cognitive independence that makes the assessment of truth-value possible. These are: reflectivity, imagination, and the capacity for identification and recognition

¹⁶ See Smith, Adam: *The Theory of Moral Sentiments* in Schneider, Herbert W. (Ed.): *Adam Smith's Moral and Political Philosophy*, Hafner Publishing Company, New York, 1948, p. 81.

in another's situation. Normally, when we reflect on moral feelings and moral socialisation, we tend to suppose that they force us to subordinate our attention and preoccupation about our own goals and interests to those of another sentient being. We think moral feelings promote the oblivion of self. This is the reason why Hume's knave or the free-rider is normally seen as someone that, while lacking moral sense, is however very capable of promoting his or her own interest, for we tend to suppose that when someone lacks moral sense, they are nevertheless able to take care of themselves as rational agents.¹⁷

A better picture would be to conceive rational deliberation in practical issues as a capacity that affects not only my relationship with other agents, but also my relationship to my interests, my goals and myself. According to this picture, the incapacity to reflect on our condition as rational agents supposes also the incapacity to identify ourselves with another agent and imagine ourselves in his situation: our condition of rational agents collapses when we are unable to consider other people's interests. Even if this sort of reflectivity is conditioned by a previous moral socialisation (so that someone who has not been socialised within a moral community would not be capable of rational deliberation in this broad sense), the lack of it also affects the capacity of an agent to effectively promote his interests and goals and to relate to them in a free or autonomous way. In this respect, it would be true, after all, that an amoral person is also irrational.

But now, you may well say this is all very counter-intuitive or even too good to be true. You can be a very bad but a very successful person at the same time, it might be objected. However, can we really describe these people as *rational*? Is the individual who lacks moral sense really trying rationally to promote his own interests and goals? According to the picture I have been trying to construct, someone lacking moral sense would not only have difficulties in gaining reflexive access to their own system of preferences and motives, which is

¹⁷ "But he is not - writes Blackburn about Hume's knave- on the face of it irrational: indeed, to manage his knavery effectively he must be intelligent as well as daring. In spite of Kant's dream, it is better then, to rest

a condition for rational deliberation, but also in having an accurate image of themselves as rational agents.

Since the conceptual reflection on the nature of practical deliberation has led us to an empirical thesis, we will now make it plausible by using what has been gained conceptually to illuminate empirical research on this matter. We may try to understand, for instance, how a person lacking moral sense really ought to be assessed. Are there descriptions of this sort of person that match our conception of rational agency and moral motivation?

Yes, there are. Now let me draw some conclusions for our conception of moral motivation from this notion of reflective rationality. The first one is that a person lacking moral sense may not really be rational either, as some internalists have suggested. Someone who uses others only as a means could, according to this semi-Kantian view of rationality, be seeing themselves just as a means for ends they cannot really deliberate over. And someone who does not have the freedom to reflect about their own ends is not being rational in a proper sense: they are not making any real choice but compulsively pursuing the things they prefer. This is compatible with some known facts about what we could call "amoral" personalities. In my opinion, these kinds of people tend to be either sociopaths or, more interestingly, suffer from narcissistic personality disorder.¹⁸ While the former have been seen as the clearer instance of people lacking moral sense, the latter dwell unnoticed among us in higher numbers: these are the kinds of people with a good sense of reality that nevertheless invariably exploit others and make them subservient to their own agenda. The narcissist characteristically exemplifies the kind of person who lacks the features we normally ascribe to a person with a healthy moral sense. These are the capacity to empathise with others, to feel guilt or remorse if another is being hurt by

content with Smith and Hume. The knave is vicious and odious. We have already the words to express our contempt: it does not add anything except rhetoric to call him also irrational". Blackburn, 1998, p.223.

¹⁸ My description of the narcissistic personality disorder is based on the landmark work of the Chilean American psychiatrist Otto Kernberg. See, in particular, *Desórdenes fronterizos y narcisismo patológico*, Paidós, Barcelona, 1993 and *Relaciones amorosas: normalidad y patología*, Paidós, Barcelona, 1995.

one's own actions and to feel ashamed when excluded from the moral community. Narcissists, for reasons I cannot explore in depth now, feel themselves so threatened by others, and so overcome with aggressive impulses towards people with whom they intend to have intimate relationships, that they are not able to integrate feelings of love with their aggressive tendencies into a unified image of a person. This trait forces them to end close relationships abruptly, since they would be unable to tolerate the impulses of their aggressive nature.

The etiology of narcissism shows not only an impaired process of socialisation that makes the narcissist incapable of the moral feelings normally associated with having a moral motivation. It is also believed that the strong impulses of aggressive nature are directed primarily towards oneself and this would explain the need to project them. Therefore, my argument is that the impaired capacity to feel empathy towards others supposes also a lack of freedom regarding one's own wants and preferences - that is to say, it supposes that the narcissist, trapped in a world dominated by ill will, has no other choice but to act as if he had to defend himself against a world that conspires to destroy him. What I am trying to say is that this characterises not only a personality that is not moral, but also very probably (but again, this is speculation, for it is an empirical assumption that must be duly tested by the moral psychologist) a flaw of a cognitive nature: the inability to relate to oneself as a free rational agent. This pleads in favour of a more differentiated view of the role rationality plays in the constitution of moral motivation. But, as I have been insisting, we need a concept of rationality that includes a better concept of what it means to be a rational agent.

Adam Smith had an intuition that supports what I am trying to say. It is not that a person who lacks moral sense loves himself or herself *more* than any other. It is that they love themselves as *badly* as they love others. So it seems to me that he believed that real love for oneself needs a rational spectator, someone who teaches us to love ourselves as others could love us: "As to love our

neighbour as we love ourselves is the great law of Christianity - he wrote - so it is the great precept of nature to love ourselves only as we love our neighbour, or, what comes to the same thing, as our neighbour is capable of loving us."¹⁹

My suggestion would be that a person who can relate to herself as a free rational agent - that is to say, who is free to deliberate about her attributions of value - could very well empathise with the necessity of freedom that others have, in order to also make attributions of value. Although not sufficient, this is a necessary condition for moral sense, for it is the first step that makes moral respect towards others possible. Perhaps Kant thought it was sufficient, since he made it the substance of one of the formulations of the categorical imperative. We cannot be as sure as he was about our capacity to treat all people at all times as ends in themselves, but I think we can consider that a conception in which to be rational also implies to be aware of the importance of being free to choose - which by definition has to be universal, as Kant foresaw - is a pretty accurate vision of reason.

References

Barreto, Luz Marina, *El lenguaje de la modernidad*, Monte Ávila Latinoamericana, Caracas, 1994.

Blackburn, Simon, *Ruling Passions*, Clarendon Press, Oxford, 1998.

Chesterton, C.K., *Ortodoxia*, Editorial Alta Fulla, Barcelona, 2000.

Dworkin, Ronald, *Foundations of Liberal Equality*, The Tanner Lecture on Human Values, Stanford University Press, University of Utah Press, 1988.

¹⁹ Smith, Adam, 1948, p. 88.

Gauthier, David, *Egoísmo, moralidad y sociedad liberal*, (translated and introduced by Pedro Francés Gómez), Editorial Paidós, 1998.

Gauthier, David, *Morals by Agreement*, Oxford University Press, 1986.

Gutiérrez, Gilberto, *Ética y decisión racional*, Editorial Síntesis, Madrid, 2000.

Habermas, Jurgen, *Theorie des kommunikativen Handelns*, Suhrkamp Verlag, Frankfurt am Main, 1981.

Habermas, J., *Erläuterung zur Diskursethik*, Suhrkamp, Frankfurt am Main, 1991.

Habermas, *Faktizität und Geltung*, Suhrkamp, Frankfurt am Main, 1998.

Hausman, D. (Ed.), *The Philosophy of economics*, Cambridge University Press, 1994.

Hume, David, *Treatise on Human Nature*, Oxford University Press, 1990.

Kant, Immanuel, *Schriften zur Ethik und Religionsphilosophie*, Wissenschaftliche Buchgesellschaft, Darmstad, 1983.

Kernberg, Otto, *Desórdenes fronterizos y narcisismo patológico*, Paidós, Barcelona, 1993.

Kernberg, O., *Relaciones amorosas: normalidad y patología*, Paidós, Barcelona, 1995.

Rawls, John, *The Theory of Justice*, Oxford University Press, 1971.

Smith, Adam: *The Theory of Moral Sentiments* in Schneider, Herbert W. (Ed.): *Adam Smith's Moral and Political Philosophy*, Hafner Publishing Company, New York, 1948.

Williams, Bernard, "Internal and External Reasons" in *Moral Luck*, Cambridge University Press, 1980.

Williams, Bernard, *Ethics and the Limits of Philosophy*, Fontana Press, 1985.