Taking Moral Intuitions Seriously, but not Dogmatically

Miriam Ronzoni (University of Frankfurt) &
Laura Valentini (The Queen's College, Oxford)

1. Introduction

Moral intuitions are virtually omnipresent in normative reasoning. Specific moral statements such as 'It is wrong to torture babies for fun' or 'One ought not to punish the innocent' strike most of us as intuitively true, and therefore as *sine qua non* of any plausible moral outlook. But what is their epistemic status?

One possibility is to think of such immediate convictions as representations of an external and objective moral reality that we 'grasp' through some special cognitive faculty: Intuition.[1] If this is how we understand intuitions, then we have good reasons to take them very seriously, and let them carry a great deal of weight in our moral arguments. The obvious difficulty with this view is that, so far, no one has been able to prove that such moral reality exists, or that we possess the necessary cognitive faculty to reach it. In fact, since people's intuitions often diverge, we seem to have good grounds for being somewhat sceptical about the existence of such a reality. Consider, for instance, the philosophically familiar scenario in which a trolley is about to kill five people unless it is diverted on a different track, in which case it would only kill one.[2] When asked whether one ought or ought not to divert the trolley, people often disagree: Some are convinced that turning the trolley is morally prohibited, others that it is required.

In light of these difficulties, we might conclude that, if intuitions do not reflect objective moral facts, then they cannot but be subjective convictions with little, if any, moral authority. If this were the case, our appeal to them in normative theorizing would be entirely unwarranted. How can a valid moral theory rest, in the best-case scenario, on our states of mind or, in the worst case, on our cultural prejudices? The only way out of this impasse seems to be a form of moral theorizing which avoids reliance on moral intuitions altogether, grounding normative principles only on allegedly safe 'rational' foundations, prescinding from our different and conflicting convictions about specific cases. Unfortunately, as we shall argue, this is not an option, either: a self-declared intuition-free moral theory either relies on dubious foundational metaphysical claims or is in fact a theory with intuitions in disguise.

It seems, therefore, that we can put faith neither in grasping moral reality through Intuition nor in any attempt to reject intuitive grips altogether and offer a systematic account of morality based on 'purely rational' considerations. We cannot rely on intuitions, but we cannot do without them, either. In light of this, is there still room for a sound moral methodology? Can philosophy say 'what we ought to do' with any degree of authority? In this paper we bring good news to moral philosophers, and answer this question in the positive.

Given the impossibility of doing without intuitions altogether, we argue that a sound moral methodology may still legitimately make a controlled use of them. Our 'recipe' for such a methodology contains two main ingredients: (1) constructivist inter-subjective justifiability and (2) reflective equilibrium. First, we argue that, since we cannot take the authority of our intuitions *as unproblematically given* – because we do not know whether they reflect an independent moral reality – authoritative normative

[1] Throughout, by intuitions (small initial) we indicate the particular representations delivered by the putative epistemic faculty of Intuition (capital initial).

[2] For a classic statement of the 'trolley problem' see Philippa Foot, 'The Problem of Abortion and the Doctrine of the Double Effect' in her *Virtues and Vices: And Other Essays in Moral Philosophy* (Oxford: Basil Blackwell, 1978).

principles must be *constructed* through deliberative procedures aiming at inter-subjective justifiability.[3] On our view, then, a normative claim is rational (hence authoritative) only if it is inter-subjectively justifiable. Second, we argue that the principles delivered by constructive procedures based on inter-subjective justifiability alone would be too abstract to be genuinely action-guiding. In order to make these principles operationalizable, we supplement them with a holistic moral methodology: Rawlsian reflective equilibrium. Crucially, within this methodology, intuitions are allowed to play a role, but only a *controlled* one: they are balanced against each other, and none of them is given absolute moral authority. The only authority they are provisionally granted, rests on their status as putative instances of sedimented inter-subjective justification. We conclude that proceeding by way of reflective equilibrium – thus giving intuitions a substantive, but controlled, justificatory role – allows us to develop genuinely action-guiding principles whilst approximating the ideal of inter-subjective justifiability and staying as clear as possible of dubious metaphysical assumptions.

Our argument is structured as follows. In section I, we briefly describe the motivations behind the view we shall call moral intuitionism and outline two serious problems with it. In section II, we consider utilitarian approaches to morality understood as a possible escape route from the problems faced by intuitionism, and argue that utilitarianism can only deliver an action-guiding moral theory by either (1) relying on contentious metaphysical claims about human nature, or (2) admitting intuitions through the back door. In section III, we discuss a second escape route from the difficulties encountered by intuitionism: constructivist approaches to justification. Proponents of constructivism hold that moral principles are valid so long as they are proven to be inter-subjectively justifiable through a hypothetical consent test. We argue that, contrary to what some constructivists think, hypothetical consent can deliver substantive (as opposed to merely formal) normative principles only if it involves an engagement with moral intuitions. A viable intuition-free constructivism is simply not an option. Hence, we conclude our *pars destruens* by admitting that, while intuitions, taken by themselves, are not a firm foundation for a sound moral theory, their complete elimination from moral theorizing is simply impossible, as the analyses of both utilitarianism and (intuition-free) constructivism show. Finally, in section IV, we consider a third, in our view superior, alternative to utilitarianism and would-be intuitionless constructivism: a constructivist approach that relies on the holistic method of reflective equilibrium. We argue that reflective equilibrium can both overcome the problems affecting intuitionism without appealing to dubious metaphysical claims and reflect the uncertain epistemic status of intuitions as potential reflections of an independent moral reality.

2. Moral Intuitionism and Its Difficulties

There seems to be no real disagreement about the claim that torturing babies for fun is wrong. And, the moral intuitionist claims, there is more to this consensus than it first appears. For not only do we agree that torturing babies for fun is wrong; we also subscribe to this claim in an unmediated way, without needing any further reflection. We think that torturing babies for fun is wrong, and we think that this belief needs no further justification; we regard it as a self-evident, obvious, and thereby *intuitive*, truth. Is it not ridiculous, if not offending, to claim that torturing babies for fun is wrong

[3] For a more detailed account of our views on this issue, see Miriam Ronzoni and Laura Valentini, 'On the Meta-ethical Status of Constructivism: Reflections on G.A. Cohen's "Facts and Principles"', *Politics, Philosophy & Economics* 7 (2008): 403-422; Miriam Ronzoni, 'Constructivism and Practical Reason: On Intersubjectivity, Abstraction, and Judgment', *The Journal of Moral Philosophy*, 7 (2010): 74-104; Laura Valentini, 'On the Nature and Possibility of Political Normativity' (manuscript).

because, on closer scrutiny, it might give pleasure to some but does not maximize overall utility? Or because abstract agents in the original position would adopt a principle that prohibits this practice? Isn't it simply *self-evidently* wrong? In fact, many would find it wrong to take this claim to be anything other than self-evident.

This simple test – with this or other widely shared moral judgments – is often used by intuitionists to make a case for their approach to moral justification. Ultimately, they claim, moral theories must rely on self-evident moral intuitions. Albeit appealing, this method has two well-known and important limits.

Firstly, intuitionism notoriously fails to provide a systematic account and ordering of moral principles. As Rawls construes it,[4] intuitionism is characterized by two related beliefs: (1) that there exist irreducible first moral principles, whose truth is grasped by intuition, and which 'may conflict to give contrary directives in particular types of cases,'[5] without this being a challenge to their truth or soundness; and (2) that there are no criteria, methods, or priority rules, to weigh these principles against one another in case of conflict – we simply have to resort to our considered judgments about what strikes us as right in these cases. Hence, Intuition plays a double role, in being the faculty through which we access first principles as well as the one through which we balance them. Not only do we only grasp normative principles by Intuition; the balancing itself among these potentially conflicting principles can only be performed through intuitive means. There is, in other words, 'no expressible ethical conception'[6] underlying this balancing; no structured theory accounting for the way we use our considered judgments and how.

Since, in Rawls's view, there is nothing intrinsically irrational about this way of proceeding,[7] the only way to dispute it may well be to offer an alternative account, capable of delivering at least equally plausible prescriptions while giving us a story about how principles are ordered and/or balanced against each other. This is the task that Rawls sets himself.[8] This assessment, however, may be more charitable than it needs be. Rawls argues that all we can do to dispute intuitionism is to come up with a better theory. This seems to be the appropriate attitude towards a theory that, all things considered, has no major independent methodological shortcomings. This is not, however, the case for intuitionism, which brings us to the second point.

Intuitions have an unclear ontological and epistemological status. Are they a direct channel towards a mind-independent moral reality? But how can we prove that, and let alone show that such a moral reality exists in the first place? We simply lack an epistemological account telling us what moral intuitions are and how exactly sound moral intuitions 'come' to us. The most obvious way out of such an impasse is to declare sound moral intuitions as being self-evident. Without further argument, however, it is by no means clear that intuitions which strike us as self-evident are not in fact shaped by the institutions and social practices we live in, and potentially mirror their biases and taboos, rather than evident truths.[9]

Many critics of intuitionism – especially from the utilitarian camp – have contended that moral intuitions are nothing more than states of mind, which may well be shaped by our psychological dispositions and our historical prejudices rather than by a reliable cognitive faculty.[10] Indeed, even in cases of wide consensus, utilitarians think that we should be cautious with immediate moral judgments, for they might replicate cultural

[4] John Rawls, *A Theory of Justice* (Oxford: Oxford University Press, 1999), pp. 30-36.
[5] *Ibid.*, p. 30.
[6] *Ibid.*, p. 34.
[7] *Ibid.*, pp. 34-35.
[8] *Ibid.*, p. 35.
[9] J.L. Mackie, *Ethics: Inventing Rights and Wrong* (London: Penguin, 1977).
[10] Richard Brandt, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979); Selim Berker, 'The Normative Insignificance of Neuroscience', *Philosophy & Public Affairs*, 37 (2009): 293-329.

prejudices which ultimately hinder the pursuit of utility, rather than reliable sedimented knowledge on how to pursue it.[11]

Moreover, moral intuitionism is strongest when it comes to showing that, in the face of some 'obviously' wrong practices, we do not seem to need, or indeed even to want, an articulate justification for their wrongness. However, it seems plausible to think it is precisely when we move beyond these practices that moral philosophy starts. It is when we do not share the same fundamental intuitions, or the same view about how two or more widely held intuitions should be balanced against each other, that we need *philosophical argument* in order to support one moral view rather than another one. From the perspective of intuitionism, it is by no means clear what we should do when our moral judgments start to diverge – other than testing and trying to dissolve or reduce the disagreement through other intuitions and examples.

Disagreement about moral intuitions is extremely problematic if we lack an account of when our moral judgments are to be trusted and why, let alone whether there is a moral reality out there that our intuitions grasp. Hence, if consensus may track shared prejudices rather than self-evidence, if intuitionism lacks satisfactory resources to deal with disagreement about intuitions themselves, and if no account for the nature and source of moral intuitions is provided, there may well be more 'irrationality' to intuitionism than Rawls charges it of. The only way would be to provide a full-blown epistemic story as to why certain individual moral judgments – whether self-evident or not – are right. And that is, as the history of moral philosophy shows, an extremely hard, if not chimerical, enterprise. Furthermore, even if successful, one such theory would not be grounded on the *intuitive* power of the relevant true judgment, but would have to provide an independent (possibly a metaphysical) justification for them.

Intuitionism, therefore, seems to have problems that urgently require attempts to device alternative strategies to moral theorizing. The two following sections briefly assess two such attempts, utilitarianism and constructivism.

3. Utilitarianism

Utilitarianism is understood as the moral theory according to which there exists only one moral principle giving guidance to the whole of our conduct: that of maximizing (*overall*) utility. In Rawls's useful reconstruction, utilitarianism is committed to the two following claims about the nature of the good and the right respectively:[12]

1) There is one rational (possibly naturalistic) good, whose goodness is fully independent from moral considerations of rightness, namely *utility*;
2) That the right is to be defined in terms that are fully derivative from the good: right is all that which maximizes (overall) utility.

For utilitarians, the weighing of intuitions and moral considerations is not a problem: the only guideline for conduct is the maximization of utility, and no moral conflicts can therefore arise. There might be problems in weighing utility, but those are problems of a different kind, raising empirical, rather than moral, puzzles. Moreover, as we have anticipated in the previous section, utilitarians strongly object to those philosophers who ground moral theorizing on intuitions. Hare goes as far as to claim that relying on moral intuitions is the common practice of 'lower', and pre-philosophical, types of moral thinking.[13] For utilitarians, immediate allegiance to some moral prohibitions derives from accumulated and sedimented human wisdom about how roughly to

[11] See, for instance, R. M. Hare, 'Abortion and the Golden Rule', *Philosophy and Public Affairs'*, 3 (1975): 201-222.
[12] See Rawls, *A Theory of Justice*, pp. 21-24.
[13] R. M. Hare, *Moral Thinking* (Oxford: Clarendon Press, 1981), pp. 46-47.

promote utility;[14] some practices have consistently proven to succeed or fail and have become, as it were, 'no-brainers'. More important still, not even this 'shortcut' explanation of the role of intuitions in moral reasoning provides us with firm reasons to rely on them; precisely because the source and nature of intuitions is uncertain, some of our strongest moral intuitions may well be accumulated knowledge about what maximizes or hinders utility, but some others might reflect nothing more than the prejudices and moral taboos of our culture, institutional environment, and social practices. For example, those who think that it would be wrong to 'turn the trolley' to save the five and kill the one, the utilitarian might complain, are simply misled by the irrational prejudice that actions are more morally salient than omissions.

The utilitarian explanation of how intuitions might come to have such an immediate moral force is certainly clever; however, it is not clear why structuring our moral thinking around the maximization of utility would be a sounder and more reliable alternative. Utilitarians typically argue that their theory provides a superior account of moral justification in that it is grounded on purely *'rational'* considerations, fundamentally traceable to the idea that we ought to maximize the good.[15] Crudely put, the reasoning behind this claim goes as follows:

P1: 'The good (i.e., utility) ought to be pursued'
P2: 'More utility is better than less'
P3: 'Nobody's utility counts more than anybody else's' (morality is impartial)[16]
C: 'We should maximize overall utility'

Since it is rational to maximize overall utility, all moral judgments which stand in the way should be dismissed as prejudices deriving from dominant cultural practices and religious doctrines. The defence of utilitarianism as a rational moral theory had some considerable force in its early days, when its agenda and its historical role was mainly to dispose of prevailing puritan and illiberal moral codes.[17] Utilitarianism was a component of the larger enlightenment agenda. Outside such historically contingent circumstances, however, utilitarianism faces serious difficulties in presenting itself as a purely rational account of morality. Indeed, the defence of its central claim that aggregate utility ought to be maximized is itself dependent either on intuitions or on unwarranted metaphysical assumptions.[18]

The claim that it is rational to maximize aggregate utility can be interpreted in two ways. On the one hand, we can interpret utilitarianism as having a *subjective account of the good (i.e., utility)*: utility, happiness, or pleasure, are the goods that ought to be maximized because they are what people ultimately always desire – to the extent that they desire other things, they desire them instrumentally, in order to achieve utility, happiness, or pleasure.[19] This line of thought incurs several well-known problems. First, and most obviously, the empirical claim that utility, happiness, or pleasure are the only things humans actually desire is highly disputable. Second, even if the claim were empirically

[14] *Ibid.*; See also Norman Daniels, *Justice and Justification: Reflective Equilibrium in Theory and Practice* (New York: Cambridge University Press, 1996), p. 4.
[15] Hare, *Moral Thinking*; Richard Brandt, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979).
[16] See for instance Hare, *Moral Thinking*, p. 44.
[17] John Stuart Mill, *On Liberty* (New Haven and London: Yale University Press, 2003).
[18] On the metaphysically problematic status of the utilitarian account of the good, and on the role played by moral agnosticism in defining the relationship between the right and the good in moral theories, see also Miriam Ronzoni, 'Teleology, Deontology, and the Priority of The Good', *Ethical Theory and Moral Practice*, forthcoming.
[19] This is the line of argument famously adopted by John Stuart Mill in his *Utilitarianism* (Chicago: University of Chicago Press, 1906).

true, it is not obvious what would make a moral theory based on its maximization in any sense more reliable, or 'rationally safe', than one based on our strongest moral intuitions – our desires can be shaped (and distorted) by culture, institutions, and practices just as much as moral intuitions. Finally, if utility is to be pursued because it is what ultimately everybody *wants*, it is not clear why it is *rational* to maximize *overall* utility, rather than equally distribute it across different agents. If utility is the good to be promoted purely on the ground that *each* of us seems to value it more than anything else, why should we bring about its aggregate, impersonal, maximization, even if this might mean that many should suffer for the grater cause of overall utility?

Utilitarians usually answer this charge by pointing out that such implausible scenarios where utility should be distributed in a disproportionate manner in order to be maximized, never occur empirically: what tends to maximize utility is a much more egalitarian distribution, for reasons due to human nature, peace, and stability. As it is widely known, deontologists are unimpressed by such response, and usually counter-reply that it does not give us a principled reason to object to some absolute evils such as slavery of torture, and is therefore at odds with our strongest moral intuitions to an excessive extent. Of course, this can strike us as being hardly anything more than a clash between two incompatible approaches: why should utilitarians worry about the potentially (if hardly possible) counterintuitive prescriptions of their theory, if their theoretical approach starts precisely from the premise that intuitions are not to be trusted?

At closer inspection, however, the issue is more complex. First of all, as we have just noted, overall utility maximization does not have a rationale if utilitarianism is based on a subjective account of the good. Secondly, and most important, the utilitarian common reply that slavery, torture, and any other disproportionate distribution of utility will never come about seems itself to be grounded on the *moral intuition* that there would be something wrong if they would: otherwise, why should we worry about such counterintuitive implications in the first place? The very fact that many utilitarians think that the slavery objection deserves a response seems an indirect acknowledgment of the force of moral intuitions.[20]

The only way out for utilitarianism seems to endorse an *objective* account of the good. This would allow them *both* to eschew indirect reliance on intuitions *and* to justify the maximization of overall utility rather than some form of distribution. The costs of taking such a route would be, however, prohibitively high. Such a route would involve relying on a full blown ontological account of what is good from an objective perspective. This was the problematic way out for moral intuitionism in order to avoid the uncertain status of intuitions, and seems the only strategy that utilitarians could adopt in order to avoid the equally uncertain implications of relying on our desires. The objective good can be construed in a variety of ways: as a metaphysical or naturalistic account of human nature or sentient beings, as what is objectively in nature's or life's interest, and so forth. The obvious problem with any of these approaches is that they are no less epistemically dubious than relying on intuitions. Talk of human essence *per se* is hard to justify, and all the more so when it involves such counter-intuitive, anti-individualistic claims as those endorsed by utilitarians; similar considerations would apply to naturalistic accounts of the good. An objective account of the good can only be grounded in shaky metaphysical foundations, as shaky as those intuitions are allegedly supposed to track.

[20] It should be briefly noted that utilitarians could revert the argument and claim that we do have strong intuitions against slavery and torture because we have come to know that they do not promote the maximization of overall utility. This, however, still leaves open the issue of why we should promote overall utility if we have a subjective account of it. An objective account, as we shall explain in the following paragraph, would incur further significant difficulties.

To conclude, then, utilitarianism has a clear advantage over intuitionism, and one already pointed out by Rawls: it delivers a master-principle that is to govern the whole of human conduct, rather than a set of individual moral judgments and potentially conflicting intuitions. But it cannot escape the fundamental methodological problem of justifying the account of the good it relies upon, be it subjective or objective.

4. Constructivism

Neither intuitionism nor utilitarianism seem to be able to provide a reliable foundation for moral principles, nor to find the right place for intuitions in moral thinking. Intuitionism gives them unwarranted prominence, whereas utilitarianism seeks to avoid them throughout, albeit trying to accommodate its prescriptions with our intuitive judgments, and thereby running into a deep inconsistency. For both approaches, the only way out seems to ground their method in a full-blown metaphysical account of morality, thus running into even larger problems. Can constructivism, the third moral method we shall consider in this paper, succeed where both intuitionism and utilitarianism fail? We shall argue that it can if 1) it lowers the ambition of what a normative approach can aspire to justify, and 2) it allows a limited role for intuitions in its process of justification, thus offering an intermediate view between purely intuitive and purely rational foundations of moral principles. The present section articulates the main features of the constructivist project and highlights the potential problems it might itself incur if it either tries to provide a purely rational foundation for ethics or excessively relies on its intuitive appeal. The next, concluding section sketches an alternative constructivist approach capable of avoiding these problems.

Constructivism, in a nutshell, characterizes a series of theories which seek to provide a rational justification for morality, but in a way that significantly differs from utilitarianism. Constructivist approaches, broadly construed, rely on an *intersubjective* account of rationality. From a constructivist point of view, the right normative principles to adopt and act upon are those which can be the object of an intersubjective consensus under appropriate circumstances.

It is necessary to specify briefly what grounds the reliance of constructivist theories on the ideal of intersubjective justification or agreement. As we have argued elsewhere,[21] we hold that the best justification for the constructivist emphasis on intersubjectivity is an *agnostic* stand towards ontological meta-ethical puzzles, such as the existence of an objective mind-independent morality and/or of an independent order of values, and the capacity of moral intuitions to grasp such an objective moral reality. As the discussion above should have illustrated, intuitions can neither be unproblematically relied upon as being our way of accessing independent moral truths, nor be dismissed as crystallized prejudices on the basis that an independent moral realm clearly does not exist.

If such an issue cannot be settled, however, we cannot simply shrug our shoulders, forget about it, and carry on our lives pursuing our *interests*, since conforming our actions to *morality* is not possible. This option is open to nihilists, who positively hold that morality is a chimera, but not to agnostics. If we are agnostics, we must acknowledge that, whenever our actions have a significant impact on others and other rational agents demand justification, replying 'there is no morality, hence I am simply doing as I please' is just as unwarranted as answering 'This is the objectively right thing to do. I just know it, even if I cannot rationally demonstrate it'. This is why agnostics take seriously the idea that reason is 'public', i.e., inter-subjective. For something to be rationally vindicated, it has to be *inter-subjectively justifiable*.

[21] For a more detailed account of our views on this issue, see, Ronzoni and Valentini, 'On the Meta-ethical Status of Constructivism'; Ronzoni, 'Constructivism and Practical Reason', and Valentini, 'On the Nature and Possibility of Political Normativity'.

This very principle lies behind the conclusion that appeals to intuitions as well as to foundational metaphysical claims cannot ground a sound moral theory. Any foundational claim or moral intuition is bound to be reasonably objected to by *some* rational agents. For instance, those who object to diverting the trolley on the one to save the five cannot be accused of being irrational by those who have the opposite intuition. The same applies to the utilitarian claim that the objective good to pursue is the maximization of overall utility. We seem to have no way of showing that certain intuitions, or claims about the objective good, command the assent of all rational agents. And in the absence of an independent and objectively binding source of authority,[22] this is a good enough reason to reject such theories. If there is no such pre-established source of authority (be it God-given or otherwise), the only normative requirements which can claim authority are those which fulfil public standards of justification, and neither intuitions nor unwarranted metaphysical assumptions (unless universally shared) can therefore qualify.

If intuitions or metaphysical claims are dubious because they lack the required inter-subjective justifiability, then we are well-advised to design a moral theory which explicitly aims to meet this requirement. This is what the constructivist enterprise is all about. It starts from the fact that we are a plurality of rational agents who have to share the same scarce resources, and whose attempts to *discover* rational (i.e., authoritative) foundations for any plausible distributive criterion have so far failed. If independent authoritative criteria cannot (at least for the time being) be discovered, then, constructivists argue, they ought to be *constructed* aiming directly at intersubjective justifiability.

But what counts as intersubjective justification? The constructivist literature offers two broad answers to this question. First, some constructivists hold that what we should seek is *real* agreement, under conditions that are variously specified according to the theory in question.[23] Such theories are, by definition, proceduralist and therefore significantly open ended as to their substantive outcome: right are the principles that real agents under the appropriate deliberative circumstances rely on. Proceduralist theories are often criticized for being incapable of providing a sufficiently concrete and realistic account of how the appropriate deliberative circumstances could be reproduced in real life and ground a reliable setting for the choice of principles. We share part of this concern; moreover, since proceduralist approaches do not have the ambition of justifying substantive normative principles, they fall outside the scope of this paper to a large extent.

Second, other constructivist theories hold a hypothetical account of intersubjective justification: right are those principles which ideally situated agents would choose in an ideally constructed deliberative procedure. Rawls's original position is the most prominent version of such an approach. Such theories do have the ambition of delivering specific, substantive principles, and maintain that, in order to achieve such a goal, one needs to reflect on what agents would agree on if they were freed of their biases, undue constraints, and differential bargaining positions – in a nutshell, if they were placed in a position that is conducive to a genuine process of intersubjective justification. Can constructivism, so conceived, provide a sound justification for normative principles?

Much relies on how the agents and the hypothetical situation are construed. It seems that, in order to provide a justification of moral principles that is genuinely

[22] Like God's law, which would plausibly give us a sufficient reason to disregard the fact that some rational agents disagree with its content.

[23] David Gauthier, *Morals by Agreement* (Oxford: Oxford University Press, 1986); Jürgen Habermas *The Inclusion of the Other* (Cambridge, MA: MIT Press, 2000); Rainer Forst, *The Right to Justification*, (New York: Columbia University Press, 2010).

intersubjective, the criteria for appropriateness should be sufficiently formal and thin – in other words, no substantive or moralized criterion should be let in through the back door. Kant's universalizability formula of the categorical imperative is one such attempt: normative principles are justifiable if we can wish for them to become laws of nature without incurring contradiction in the will. Harming others, undermining their agencies, lying, and stealing are examples of forbidden conduct, because such forms of behaviour would fail the test of universalization.

Onora O'Neill's constructivist outlook is an attempt to reformulate the Kantian model in the contemporary (mainly Rawlsian) language of an ideal and hypothetical deliberative procedure. O'Neill claims that the principles to be adopted are those which would be chosen by a plurality of agents characterized by imperfect and indeterminate rationality and independence. In such a procedure, O'Neill argues, the only principles that could be chosen are those that we can imagine to be simultaneously acted upon by all. The outcome of one such procedure would be the rejection of principles of deception, coercion and severe injury for 'nobody can [for instance] coherently think of a principle of injuring severely as adoptable by all (in any domain): if it were so adopted, some would succeed in acting on it, hence others would become victims, so unable to act on it'.[24]

These thin procedural constructions are extremely attractive from the point of view of their capacity to generate intersubjective consensus; however, they only manage to vindicate equally thin, abstract and general, normative principles – in other words, can constructivism, so thinly conceived, manage to successfully recommend or rule out any concrete course of action? We might think, for instance, that the test of universalization would rule out lying, for we couldn't possibly want lying to become a law of nature – but maybe we would not want a world in which everybody always said the truth, even when lying could save innocent lives either. Similarly, O'Neill's approach can justify the rejection of deception, coercion, and injury – but what *counts* as coercion, deception, and injury? O'Neill powerfully argues that this is a matter of applying principles through the faculty of judgment. This, however, seems again to let intuitions play a decisive role indeed without openly admitting it: it is through individual moral intuitions that we ultimately decide on courses of action, namely on whether doing x rather y would count as avoiding coercion, deception, or injury. Leaving so much space to intuitions in determining the actual content of our conduct seems problematically at odds with the very motivations of constructivism.

Finally, and for reasons that should by now be clear, thicker constructivist procedures, which construe the ideal deliberative situation of agents through moralized notions (freedom and equality, or autonomy, to name but the two most popular ones) are problematic for specular reasons: unwarranted moral concepts – whether grounded on intuitions or an unvindicated order of objective values – are introduced to specify the very conditions for intersubjective justification.

5. Constructivism Reloaded: Reflective Equilibrium and the Appropriate Role of Moral Intuitions

The discussion conducted so far has shown that, whilst intuitions constitute a shaky ground for moral justification if we take them to be the ultimate source of authority for moral principles, doing completely without them is not an option, either. Even if we wish both to formulate truly action-guiding principles and to stay clear of unwarranted value-based premises, they just keep coming back. What can be done then? This section argues that a coherentist/holistic approach towards the justification of moral principles (such as Rawls's reflective equilibrium, if embedded within an constructivist method of

[24] O'Neill, *Towards Justice and Virtue: A Constructivist Account of Practical Reasoning* (Cambridge: Cambridge University Press, 1996), p. 59.

justification by way of ideal intersubjective agreement) is the most defensible account of the role of intuitions in moral theorizing. Such an approach can (1) cater for the appeal of the grounding idea underlying the constructivist enterprise, as we have described it in the last section, while (2) using intuitions to flesh out concrete principles but (3) significantly controlling and limiting their role in moral reasoning.

The widely known idea of reflective equilibrium, applied to moral and political philosophy for the first time by Rawls, is a coherentist/holistic method, rather than a foundationalist one, in that it does not rely on undisputable premises, but rather seeks to achieve a coherent moral system by a careful balancing of our considered judgments and their generalizations in the form of normative principles. No intuitions, judgments, and principles are taken as given and sacrosanct within reflective equilibrium, there are only different levels of resilience. All 'fixed points' are 'fixed for the time being'. Reflective equilibrium proceeds by 'working back and forth among our considered judgments (some say our 'intuitions') about particular instances or cases, the principles or rules that we believe govern them, and the theoretical considerations that we believe bear on accepting these considered judgments, principles, or rules, revising any of these elements wherever necessary in order to achieve an acceptable coherence among them.'[25] Thus, intuitions do play a role, but never an all-encompassing one; any intuition can, in principle, be discounted. The role of intuitions is controlled and subject to the ultimate goal of reaching a coherent, balanced equilibrium *capable of attracting consensus*. Whereas even a rational intuitionist can in principle adopt some form of reflective equilibrium as a *technique* to find out what is to be done in concrete cases;[26] such an exercise is for the rational intuitionist only a device to track down what is independently true in complex cases where a plurality of divergent moral intuitions simultaneously apply.[27] From a constructivist perspective, instead, reflective equilibrium is much more than a technique or a device: it is the only method by which the regulative idea of intersubjective justification can come to justify sufficiently concrete and properly action-guiding principles without giving moral intuitions unlimited sovereignty. Reflective equilibrium is, from this perspective, a fully intersubjective exercise: it is the 'point of view of you and me'.[28] It is the way through which a plurality of rational agents come to justify to one another an adequate set of action-guiding principles.

When engaging in reflective equilibrium, we constantly move back and forth between considered judgments and their generalization into principles. In so doing, we are bound to find discrepancies, and the exercise of finding an equilibrium between them is the exercise of judging when either the former or the latter need to be qualified, modified, or even fully discounted. In a nutshell, it is the exercise of finding out which bullets we are prepared to bite. As a result, intuitions do play a role, for when a general principle delivers counterintuitive prescriptions it also gives us a pro tanto reason to reconsider the stability of our equilibrium. However, no intuition is ever a sufficient reason to discount a principle – the process of reassessing the equilibrium is triggered, but the outcome is open ended. Moreover, the equilibrium is not guided or limited by a set of undisputable substantive premises – the only regulative idea is that of reaching intersubjective justification.[29] Thus, what might make some intuitions more resistant

[25] Norman Daniels, 'Reflective Equilibrium', *Stanford Encyclopedia of Philosophy*.

[26] As Rawls himself concedes – see John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), pp. 95-96.

[27] Ronald Dworkin, 'The Original Position', in Norman Daniels, *Reading John Rawls* (Stanford: Stanford University Press, 1989), pp. 16-52.

[28] Rawls, *Political Liberalism*, p. 28.

[29] This, at closer inspection, renders the method of reflective equilibrium as we defend it here short of a coherentist method *all the way down*. It proceeds to justify concrete principles in a coherentist fashion, but

than others in the process (and cause us not to discount them but after a series of attempts of reconciliation) is the fact that they are widely shared, and that – *within the process of intersubjective justification* – they resist even if we test them a) against competing principles b) by making them operate in difficult scenarios and hard cases, and c) by scrutinizing whether *we* would still have them if we were in other people's shoes. Thus, whereas intuitions play an important role, and some of them may well come to serve as temporary fixed points, the reason why they do so is that they prove particularly resilient in the process of mutual justification that reflective equilibrium triggers. In so doing, reflective equilibrium does not take a stand as to whether some intuitions are particularly robust because they track independent truths or because they constitute sedimented cases of strongly intersubjective justifiable claims.

One last important remark is relevant here. By proceeding through a back and forth movement from the single case to the general, and by seeking intersubjective justifiability, reflective equilibrium cannot, and does not, resort to intuitions in a fully disembodied manner. This is the case for two main reasons.

First, as we have presented it, constructivism aims to solve a real-world practical problem: That of finding authoritative moral principles on the basis of which imperfect and finite beings like us can lead our lives together. Constructivism does not aim to discover timeless moral truths that are valid across all possible worlds. The very point of the constructivist enterprise presupposes the circumstances of human existence.[30] Were human beings very different from what they are – e.g., considerably more intelligent or powerful – and were the world different from what it is – e.g., characterized by super-abundance or extreme scarcity – constructivism would no longer make sense. What is more – and this leads us to our second point – constructivists have positive reasons to *avoid* those scenarios, since they are likely to prevent, rather than promote, inter-subjective justifiability.

In order to reach intersubjective justifiability, we must operate with scenarios that are ideal and hypothetical, but still reproduce the general and familiar features of human existence and life in a social environment. Thus, constructivist procedures make us test intuitions in unusual scenarios ('would you find x fair if you did not know whether you are a white middle class male or a black single mother?'), but always in potential real life circumstances, rather than in fully far-fetched thought experiments.

Every time we test an intuition in broadly familiar contexts, we are testing its degree of intersubjective acceptability more than its status as a *pure* moral intuition. If we do not think that intuitions are useful *qua* sources of independent moral truths, then we have no independent guarantee that, when tested in far-fetched scenarios where intersubjective justification loses its grip, they might still be useful. Our intuitions on highly far-fetched cases are wildly divergent, and we can therefore not trust those cases from the viewpoint of intersubjective justifiability. Intersubjective justification requires a sufficient degree of proximity and familiarity with the conditions of social interaction as we know it; we need at least some experience of what people tend to find acceptable and justifiable in conditions that sufficiently resemble real life.

6. Conclusion

Our aim in this paper has been to argue that a holistic approach to justification, making a constrained and controlled use of moral intuitions, is the most reliable tool to construct an authoritative moral theory in light of the problems fleshed out above. In particular, we have defended the method of reflective equilibrium. Unlike more classical defences of this approach, however, our argument has not been grounded on the fact

its ultimate justification – and hence foundation, albeit a significantly deflationary one – is the achievement of intersubjective justification.

[30] Cf. Thomas Pogge, 'Cohen to the Rescue!', *Ratio* 21(2008): 454-475.

that reflective equilibrium can avoid the counter-intuitive implications of utilitarianism while providing the systematic account and ordering of moral principles that intuitionism lacks.[31] Our main point has been, instead, that a sound moral methodology must reflect the fact that intuitions may represent an objective moral reality, but they may as well not: we simply do not know. This position of *agnosticism* accepts the possibility that there may be objective and mind-independent moral facts captured by intuitions, but holds that a *mere* possibility is insufficient to warrant the adoption of a methodology which takes moral intuitions as unquestionable fixed points in moral theorizing. Moral intuitions would carry such a binding authority only if they were proven to be revelatory of some independent moral truth. Until their epistemic and ontological status is ascertained, a sound moral method must take them into account, but in a constrained and controlled manner.

[31] Rawls, *A Theory of Justice*, pp. 30-36.