



Motivating Justice

Vittorio Bufacchi¹

Department of Philosophy, University College Cork, Cork, Ireland.

E-mail: v.bufacchi@ucc.ie

This article challenges the received view on the role of motivations in contemporary theories of social justice. Neo-Kantians argue that a theory of justice must be rooted in moral motivations of reasonableness, not rationality. Yet reasonableness is a demanding motivation, stipulating actions that people may not be able or willing to perform. This opens egalitarians like Rawls to the accusation of prescribing a political philosophy that is not ‘followable’. The aim of this article is to explore the benefits for egalitarian theories of justice of sanctioning both rational and reasonable motivations. Being vulnerable to self-regarding passions, people often fail to act upon the principles of fairness or impartiality, even when they are motivated by their sense of justice. The institutions of justice must therefore find a way to fortify and encourage the development of our sense of justice. This can be achieved by giving people a rational incentive to act upon and therefore engender their sense of justice. Finding ways to motivate people to act justly is the most pressing challenge facing egalitarian justice today.

Contemporary Political Theory (2005) 4, 25–41. doi:10.1057/palgrave.cpt.9300109

Keywords: social justice; motivations; sense of justice; Rawls; Scanlon; Hume

‘However that may be, after prolonged research on myself, I brought out the basic duplicity of the human being. Then I realized, as a result of delving in my memory, that modesty helped me to shine, humility to conquer, and virtue to oppress’. (Albert Camus, *The Fall*, 62–63)

Introduction

The basic duplicity of the human being lies in our ability to be good citizens, even being capable of supererogatory acts, while all the time indulging in self-pity, self-love, and ultimately self-regarding sentiments. Jean-Baptiste Clamence, the anti-hero of Camus’ *The Fall*, is the epitome of virtue and hypocrisy. He shamelessly reminds us: ‘too many people now climb on to the cross merely to be seen from a greater distance’ (Camus, 1963, 84). Yet just when the reader feels morally superior to the villain in Camus’ masterpiece, it is he, the villain, who has the last laugh. At the end of the story, the portrait he has painted of himself becomes a mirror, he passes from the ‘I’ to the ‘we’: ‘When I get to ‘This is what *we* are’, the game is over’ (Camus, 1963, 103, emphasis added).



Contemporary political philosophy has, to its great loss, forgotten Camus' lesson, and neglected Jean-Baptiste Clamence's insights. While seeking legitimacy in theories of moral psychology and moral motivations, the vast literature on social justice has failed to capitalize on the duplicity of the human being. In an effort to compensate for this lacuna in the literature on social justice, this article sets itself two aims. Firstly, to challenge the received view on the role of motivations in contractarian views of social justice, and secondly, to explore the benefits for egalitarian theories of justice of sanctioning the type of motivations that reflect the duplicity of the human being. Endorsing a range of conflicting motivations has potentially profound implications for the way we think about social justice. All of us have the capacity for moral actions, although we often act immorally. Similarly, our behavior is often determined by our self-interest, when instead we ought to be more other-regarding. This is the essence of our duplicity. The biggest challenge for a theory of justice is to prescribe an institutional context that will turn self-interest into benevolence.

Social Justice and its Motivations

A motive is what induces a person to act. The Latin root of the word is *motivus*, meaning 'to move', hence the idea of movement, or initiating action. A motive gives a person a reason to perform an action. It follows that a moral motivation may be defined as what induces a person to act morally, or what reason a person has to act morally.

Moral motivations are a crucial component of any theory of justice. Brian Barry (1995) hits the nail on its head when he reminds us that any theory of justice must address the question: 'What is the motive for behaving justly?'² Two opposing contractarian models dominate the contemporary debate on social justice: neo-Hobbesian *vs* neo-Kantian. These two rival archetypes are distinguished by the way the above question is answered. According to the neo-Hobbesian theory of justice as mutual advantage, people are moved by prudence, a desire to secure individual advantage. On the other hand, according to the neo-Kantian theory of justice as impartiality, the motivation behind the pursuit of justice is the belief that the well-being of others matters intrinsically, therefore justice is essentially about giving equal consideration to everyone.³

In recent years, the debate on moral motivation and social justice has converged on two key terms: rationality and reasonableness. Following Allan Gibbard (1990) we may want to define rationality in terms of two components: instrumentality (the appropriate means to whatever goal one may have) and formal coherence (one's preferences form an ordering, and one always does what one most prefers).⁴ This conception of rationality, which is now orthodox



among economists and decision theorists, is fully endorsed by David Gauthier (1986, 1997) and other advocates of mutual advantage, who equate rationality with self-regarding moral motivations.

Reasonableness, on the other hand, supposedly captures certain moral qualities. According to John Rawls (1980, 528), the concept of reasonableness is autonomous and independent from the notion of rationality, and ought to be defined on different grounds than rationality: the reasonable expresses a conception of fair terms of cooperation, while the rational expresses a conception of each participant's rational advantage. It follows that according to Rawls (1989, 236), the reasonable (unlike rationality) has a moral underpinning.

Of course, Rawls (2001, 186) does not deny the appeal or veracity of rationality. On the contrary, he explicitly and repeatedly reminds his readers that as a liberal political conception, justice as fairness views its citizens as both reasonable and rational: 'A liberal conception of political legitimacy aims for a public basis of justification and appeals to free public reason, and hence to citizens viewed as reasonable and rational'.⁵ At the same time, he also states in no uncertain terms that the reasonable and the rational do not stand on equal rank, instead the reasonable presupposes and subordinates the rational. In *Political Liberalism* (Rawls, 1993, 50) and *Justice as Fairness: A Restatement* (Rawls, 2001, 6–7), Rawls affirms that rational agents lack the moral sensibility that underlies the motivation to engage in fair cooperation, they lack a sense of justice and therefore they fail to recognize the independent validity of the claims of others. Reasonable persons desire a social world in which they, as free and equal, can cooperate with others on terms all can accept.

After many years of debating on the virtues and short-comings of rationality and reasonableness, of mutual advantage and impartiality, the literature on social justice seems to have reached an impasse.⁶ Today, the debate between contractarian theories of justice has degenerated into a shouting match: if you believe that we are moved by reasonableness and a sense of justice, then you will probably endorse something along the lines of a theory of justice as impartiality. Alternatively, if you believe that people are moved by rationality and self-regarding interests, then you will probably champion justice as mutual advantage. Advocates of justice as impartiality condemn Gauthier and his followers of misappropriating the language of morals to defend a position that has nothing to do with justice, while on their part those who champion justice as mutual advantage frown on the likes of Rawls, Barry and Scanlon for making demands in the name of justice that are morally unwarrantable.

In fact, this debate between neo-Hobbesians and neo-Kantians is fundamentally misguided. There are two different contexts, or levels, in which a motivation enters the domain of social justice: at a lower-order level when acting upon a certain conception of justice, and at a higher-order level when



constructing a theory of justice.⁷ In neo-Hobbesian and neo-Kantian theories of justice, motivations perform two radically different roles, therefore for all means and purposes the two dominant approaches to social justice we have discussed so far operate on different levels. For neo-Hobbesians, the starting point is the way people behave, and any construction of a theory of justice must conform to the psychological motivations that social agents happen to have. Alternatively, neo-Kantians prioritize the idealized motivations that contractual agents are required to have in the process of ‘constructing justice’, which in turn informs the way social agents behave.

In what follows, I will not engage with neo-Hobbesians on their approach to social justice. From an egalitarian point of view, the inadequacy of their project is all-too familiar, and its defects have been exposed many times before.⁸ Instead, my concern rests with neo-Kantian theories of justice, and in particular with the account of the role of motivations in the works of Rawls and Scanlon.

The Role of Motivations in Rawls

Arguably, the most common misinterpretation of Rawls’ theory of justice as fairness, as it appeared in 1971 in *A Theory of Justice*, has to do with the difference between the parties in the original position, and how people actually behave in certain situations. It is a common mistake, but still a mistake, to confuse the motivations of the contracting parties in Rawls’ original position with the motivations that Rawls supposes people to have in real life.⁹ It is no doubt in an effort to avoid further misunderstandings that in *Justice as Fairness: A Restatement*, Rawls (2000, 81) returns on this point with renewed vigor:

Despite the similarity between familiar arguments in economics and social theory and the argument from the original position, there are fundamental differences. One difference is that our aim is not to describe and explain how people actually behave in certain situations, or how institutions actually work. In describing the parties [in the original position] we are not describing persons as we find them. Rather, the parties are described according to how we want to model rational representatives of free and equal citizens.

The parties in the original position are what Rawls (2001, 87) calls ‘artificial persons’. They are the citizens’ representatives in the process of constructing justice. Unlike you and I, these artificial persons are not moved by envy and spite, they are not risk-takers, and they have no inclination to exercise power over others. Instead, because they operate in a context where ‘the reasonable



has priority over the rational and subordinates it absolutely', (Rawls, 2001, 82) any agreement made by the parties in the original position is fair.

This is fine as far as it goes, but there is still work to be done. Rawls has told us how we, through our representatives, come to agree on fair terms of agreement. However, what reassurance do we have that the citizens will go along with what the representatives have agreed? In other words, what if we fail to live up to the expectations of the artificial persons? After all, our motivations are different from those of our representatives. So it is not inconceivable that the motivations of the actual citizens will undermine all the good work done by the 'artificial persons' in the original position.

The retort to all these preoccupations lies in the idea of a sense of justice. As Barry (1995, 57) reminds us, 'Rawls's conclusions are intended to appeal to his readers' sense of justice, not their self-interest'. The desire to act justly is intrinsic to the sense of justice, which Rawls (1972, 567) defines as 'an effective desire to apply and to act from the principles of justice and so from the point of view of justice'. It follows that a well-ordered society is 'effectively *regulated* by a public conception of justice' (Rawls, 1972, 5, emphasis added). Of course, Rawls is right when he argues that we have a sense of justice, but this is still not a satisfactory solution. After all, this sense of justice may not be sufficiently strong to motivate our moral actions. I am not denying that we have a sense of justice, but we need to be reassured that our sense of justice will not succumb to our self-regarding interests. Unlike the artificial persons in the original position, you and I cannot assume that whenever our moral reasonableness clashes with our self-interested rationality, the former will always come out on top.

Even Barry (1995, 100), who on the issue of the sense of justice sees eye-to-eye with Rawls, is aware of the high expectations and uncompromising demands of justice upon us: 'the most important and at the same time perhaps the most elusive of the circumstances of impartiality is a motivational one: the willingness to accept reasonable objections to a proposal regardless of the quarter from which they come'. We should take note of the *elusive* nature of the motivations of impartiality. What is to become of justice as impartiality if people in a given society do not share the required motivation? What if our sense of justice, or our motivation to be impartial, is trumped by our self-interest? There is a risk that Rawls' justice as fairness, for all its merits, is deemed unfollowable and therefore will not be endorsed.

Scanlon and The Problem of 'Followability'

One of the most enduring and prevalent influences on Rawls' work, especially on the question of reasonableness as a dominant moral motivation, is the



moral theory of Thomas Scanlon. In *What We Owe to Each Other*, Scanlon does not deal specifically with political questions or even with issues of social justice, nevertheless because of his influence on Rawls it is instructive to see what Scanlon has to say about the nature of moral motivations. My impression is that on the role of motivations Scanlon's moral theory runs into similar trouble as Rawls' theory of justice.

Scanlon is famous for formulating the most original and erudite contractarian moral theory in recent memory, which is captured by the following formula: 'An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced general agreement' (Scanlon, 1998, 153). The most interesting aspect in this contractarian theory, indeed what makes it specifically Scanlonian, is the motivational basis of this agreement: 'people have reason to want to act in ways that could be justified to others' (Scanlon, 1998, 154).

Readers who are well versed in the literature will have noticed that there is a small but significant discrepancy between the above account of moral motivation and Scanlon's famous original formulation of the same from 1982: 'the desire to be able to justify one's actions to others on grounds they could not reasonably reject' (Scanlon, 1982, 116). The crucial difference between these two accounts of contractarian motivation is that 'desires' have been dropped, and replaced by 'reason'. This is, in fact, a monumental switch. In a philosophical move reminiscent of Kant's critique of Hume, Scanlon rejects the claim that desires can be the sources of motivation. Instead, he now takes the idea of reason as primitive. Reasons, Scanlon (1998, 35) tells us, are the *only* motivating factors.

Scanlon replaced 'desires' with 'reasons' in his account of moral motivations after 1982 because the idea of desires created special problems for his conception of practical reasoning. To be precise, Scanlon (1998, 7) was having trouble explaining how someone who lacks the desire to justify one's actions to others has any reason to be moral: 'Many people pressed me to say whether, on my view, a person who lacked this desire would have any reason to avoid acting wrongly, and to explain how I would account for the fact that lacking this desire is a particularly serious fault'. The way Scanlon solved this problem is by reassessing the relationship between reasons and desires. While in 1982, he believed that having a reason can be understood in terms of the notion of desires, he now thinks that it is the other way around, and the notion of desire needs to be understood in terms of the idea of taking something to be a reason.

There is a lot to be said for Scanlon's argument that reason is primitive and desires derivative (rather than the other way around), yet the switch from 'desires' to 'reason' may not be the solution to all Scanlon's problems. Recall that what got Scanlon in trouble in 1982 was the fact that he could not account



for those who lack the desire to justify one's actions to others. The suspicion is that the idea of reason may run into similar problems, although for different reasons. It is not so much a question of *lacking* a reason that presents an obstacle, since the nature of reasons is such that we can work on and develop them in a way that would not apply to desires. Instead, the potential problem is that even if we have a reason, we may not do anything about it. In other words, having a reason for being good is a necessary but not a sufficient condition for doing good.

There is a suspicion that the entire project of building accounts of motivation into conceptions of practical reason may be unworkable. In assessing contemporary theories of justice and virtues, Onora O'Neill (1996, 7) has expressed reservation along similar lines. The problem with most accounts of moral motivation is not only that the subject matter is 'among the most confused and uncertain domains of philosophical inquiry at present' but, what is even more worrying, a conception of practical reasoning built upon an account of motivation 'will lack import for those for whom it is not followable'. In other words, the problem with reasons is not that they may be *lacking* (as is the case with desires), but that they may be *unfollowable*. I will refer to this as the problem of followability.

A word about 'non-followability'. My understanding is that 'not followable' points to a cleavage between our moral beliefs and our moral actions. A moral motivation is not followable if it endorses actions that people may not be able or willing to perform. That is to say, someone may hold certain beliefs about what is right and wrong, but somehow this person may be unable to follow through on their beliefs. The inability to translate a belief into an action is what makes a moral motivation 'not followable'. The challenge posed by Onora O'Neill is serious, and needs to be confronted. That certain philosophical assumptions about our moral motivations 'are not followable' by everyone is especially damning for neo-Kantian egalitarian theories of justice. Once again, the suspicion arises that neo-Kantians are out of step with ordinary thinking about moral motivations. One aim of this article is to rescue egalitarian theories of justice from the accusation of 'non-followability'.

Rawls on Moral Education

I have argued that neo-Hobbesian and neo-Kantian theories of justice are to be distinguished at the level of motivations, and that lower-order motivations of actual citizens should not be confused with higher-order motivations of the citizens' representatives in the original position. For neo-Kantians like Rawls, the motivations of the artificial persons in the original position are such that the reasonable has priority over the rational and subordinates it absolutely. As



for the actual citizens, although they do not have the same motivations as their representatives, they still have a sense of justice, which supposedly will make them act upon the principles of justice as fairness.

While neo-Kantians are right to reject the crude moral psychology championed by their neo-Hobbesians adversaries, based on a one-dimensional conception of self-interest, there are serious problems with this neo-Kantian approach. I am not disputing the view that actual citizens have a sense of justice, but having a sense of justice is no guarantee that one will act upon it. For neo-Kantians, the absolute source of our moral conduct is our moral motivation. Our reasons for action are specifically moral, they represent moral requirements. Yet, the problem with reasons is not that they may be lacking, but that they may be unfollowable.

Rawls' theory of justice faces a problem at the operational level. After defining the idea of a well-ordered society, and the idea of the basic structure, grounded on fair terms of cooperation and impartial principle of justice, what remains to be shown is that justice as fairness is followable. The gap between what the artificial persons in the original position have agreed upon, and what actual people are capable of, needs to be addressed, otherwise Rawls' theory of justice (and neo-Kantian egalitarian theories in general) will not have an impact outside the walls of a university lecture hall. Rawls is aware of this issue, and to his credit has addressed it in his work.

Rawls tackles this problem from two different angles. Firstly, he points out that the parties in the original position must take the strains of commitment into account. That means that the parties must be reasonably confident that the person each represents will be able to honor it: 'The parties must weigh what we may call the strains of commitment. They must ask themselves whether those they represent can reasonably be expected to honor the principles agreed to in the manner required by the idea of an agreement' (Rawls, 2001, 103). Secondly, Rawls suggests that efforts must be made to kindle people's capacity for a sense of justice. Thus, in *A Theory of Justice*, Rawls makes the following claims:

Moreover we are to *encourage* certain traits of character, especially a sense of justice. (Rawls, 1972, 327, emphasis added)

the sense of justice it [the institutions of a well-ordered society] *cultivates* and the aims that it *encourages* must normally win out against propensities towards injustice. (Rawls, 1972, 454, emphasis added)

[a conception of justice] is seriously defective if the principles of moral psychology are such that it fails to *engender* in human beings the requisite desire to act upon it. (Rawls, 1972, 455, emphasis added)



The choice of terms used by Rawls leaves no room for ambiguity: ‘to encourage’, ‘to engender’, ‘to cultivate’ clearly indicate a determination to help citizens develop their latent sense of justice.

More recently, in *Justice as Fairness: A Restatement*, Rawls takes an even more explicit line. He now refers to ‘the educational role’ performed by the basic institutions of a just society:

If citizens of a well-ordered society are to recognize one another as free and equal, basic institutions must educate them to this conception of themselves, as well as publicly exhibit and encourage this ideal of political justice. (Rawls, 2001, 56)

Certain features of a political conception importantly affect the political sociology of the basic institutions that realize it. More exactly, we must consider how that sociology may be affected by the educational role of a political conception of justice such as justice as fairness with its fundamental ideas of person and society. (Rawls, 2001, 146)

The idea behind the educational role of a political conception of justice suitable for a constitutional regime is that by being embedded in political institutions and procedures, that conception may itself become a significant moral force in a society’s public culture. (Rawls, 2001, 147)

Rawls’ idea of moral education is undoubtedly important, and ought to be pursued further.¹⁰ In fact, it is both surprising and disappointing that this aspect of Rawls’ work, particularly Part III of his *A Theory of Justice*, has not received the attention it clearly deserves, especially when compared to the mountains of critical literature devoted to other aspects of his theory, such as the original position and the difference principle. Yet, as a solution to the problem of followability, Rawls’ appeal to moral education simply does not go far enough. By introducing the idea of moral education, Rawls is only pointing us in a certain direction; he does not tell us how to get to our destination. It is to this latter question that I turn to next.

Motivating Justice

In what follows, a solution to the problem of followability will be advanced. What I am going to suggest should not be seen as an alternative to Rawls’ idea of moral education, but only as an elaboration of it. The starting point of this elaboration is an account of the motivations of the ordinary citizens that are asked to act upon their sense of justice. This account is neither neo-Kantian nor neo-Hobbesian, but (for lack of a better term) neo-Humean.¹¹



There are two crucial differences between the Humeans and the Kantians on the question of moral psychology.¹² Firstly, Hume begins with psychology, and sees ethics as an elaboration of it. Kant, on the other hand, starts with ethics, and works backwards to psychology, making the latter fit his ethical theory. Secondly, while Kant grounds moral motivations on rational moral beliefs and moral judgments, Hume argues that our motivations stem from desires or inclinations, therefore there is a component in human motivation which can be understood independently of ethics.

Of course, Hume is not the first to emphasize the role of desires in our psychology. According to Hobbes, we have one dominant passion, namely, self-preservation. Famously, Hobbes (1994, 32) has no time for other-regarding emotions, since he believes that every emotion can be reduced to a self-regarding perspective, as his definition of ‘pity’ in *Leviathan* demonstrates beyond doubt: ‘grief for the calamity of another is pity, and ariseth from the imagination that the like calamity may befall himself’. Clearly, Hobbes’ account of emotions is inadequate, being overly simplistic and crude. Hume, on the other hand, appreciates the complexity of human psychology. He is prepared to accept that other-regarding emotions are also part of our moral nature, therefore apart from self-regarding ‘self-love’ we are also capable of what he calls ‘sympathy’ or ‘general benevolence’, namely the tendency to enter into the joys and sorrows of others.

This is not the place to discuss in any detail the validity of Hume’s analysis of passions. The aspect of Hume’s philosophy I want to emphasize here is specifically his belief in our mixed motivations, or in other words the belief that we are moved by both self-regarding and other-regarding emotions, by both ‘self-love’ and ‘sympathy’. However, Hume is not simply making the obvious point that we are neither completely absorbed by self-love or by sympathy. More specifically, Hume thinks that the motivation of sympathy, while present, is weaker than the motivation of self-love, and more often than not sympathy gets trumped by other passions of a more self-regarding nature, such as avoiding pain and promoting pleasure. To use Rawls’ terminology, Hume’s position could be described as the rational by and large having priority over the reasonable and subordinating it. This is a crucial aspect of Hume’s theory. It is precisely because sympathy is weaker than self-love that it requires buttressing by self-interest if it is to be sufficiently powerful to resist conflicting claims on our motivation stemming from self-love.¹³

Hume’s attempt to reconcile sympathy with self-love warrants further discussion. Hume suggests ways in which self-regarding motivations can fruitfully interact with other-regarding motivations. There is an important lesson here that neo-Kantians, and advocates of egalitarian justice in general, would be well advised to learn. It is unrealistic demands on our motivations that have turned egalitarian justice into an ideal that most people deem



unfollowable. By being labeled unfollowable, egalitarian justice is running the risk of being dismissed as a not serious contender for social reform, being therefore relegated to the league of intellectual pies-in-the-sky with the likes of utopianism and anarchism. Endorsing Hume's view of mixed motivations could have profound implications for a theory of social justice.

Acknowledging our mixed emotions, and the range of motivations that goes with it, forces us to rethink the scope of justice. We are all capable of both other-regarding and self-regarding behavior, of being moved by reasonable and rational motives, and the context within which we operate will often determine which side will prevail.¹⁴ It follows that a just society is not simply defined in terms of institutions that assign rights and duties, or distribute benefits and burdens.¹⁵ Instead just institutions must also perform another function, namely: *influencing the rational incentive structure of each individual citizen with a view to foster their sense of egalitarian justice.*

In the *Nicomachean Ethics*, Book II, Aristotle argues that moral virtue comes about as a result of habit. If we think of a sense of justice as a moral virtue, it follows that in order to engender or cultivate our sense of justice, we must first get into the habit of acting justly. This is where rationality takes center-stage: rational incentives can encourage the habits of a sense of justice. There are two types of rational incentives: negative and positive. Rawls dismisses negative incentives, such as sanctions,¹⁶ but there is no reason why positive incentives, of a psychological nature (status and reputation) or even a material nature (resources) should not be allowed. With the help of positive rational incentives, actual people¹⁷ may get into the habit of acting upon their sense of justice, which is the first step towards the kind of moral education Rawls is advocating.

There is no guarantee that people will embrace other-regarding motivations simply on the force of conviction. However, sympathy can translate into moral action with the support of self-interested motivations. It is one of the functions of just institutions to provide the added rational incentives that invite people to be egalitarians, in both word and deeds.¹⁸ The reason why egalitarian theories of justice must find a way to integrate aspects of rationality within its ethical framework is that the morality of reasonableness is not immune from the logic of collective action. The fact that the public good may be of a moral nature does not invalidate the rationality of free-riding. This point is made by Gross (1997, 11) in his study of ethical activism:

There is no reason why moral interests are structurally any different from economic or material interests, or why shared moral interests should be sufficient to precipitate moral action. As a collective-action problem, the chief obstacle to the realization of any moral interest is not insufficient moral energy but very pedestrian free-rider problems and lack of appropriate incentives.



The rational tendency to free ride thwarts successful action, and individuals must usually be motivated by incentives other than the benefits provided by the public good. There is no reason why moral interests should function any differently and also succumb to the logic of collective action. As a result, shared moral interests, even those supported by a mature sense of justice and autonomy, may prove to be insufficient.

Contrary to what neo-Kantian moral philosophers like to believe, we cannot assume that reasonableness is the dominant moral motivation. Rationality is a reality that we neglect at our peril. There is a clear discrepancy between what the egalitarian moral philosophers want to believe, and how actual people behave. It is time for neo-Kantian egalitarian moral philosophy to take rationality seriously, and use it to their advantage. If having a reason to act morally is a necessary but insufficient condition for guaranteeing moral action, then we must give people a rational incentive to act upon their sense of justice. For as much as we do not want to admit it, unlike the artificial persons in the original position, actual people will act upon their sense of justice only if triggered by rational incentives.

A Brief Elaboration

Even to suggest that social, economic and political institutions could appeal to our rationality in order to engender our sense of justice is highly controversial, as one can easily fall prey to accusations of social engineering. Describing how these institutions may work in practice would require a detailed analysis which is beyond the scope of this article, yet for the sake of clarification, it may be useful to elaborate on one example of how this process might operate.

In order to engender our sense of justice, the institutions of justice must find a way to encourage us to act according to the principles of fairness and impartiality, even if it means appealing to our rational interests, such as the desire to have a good reputation. Personal reputation is very dear to each individual citizen. Reputation is a faithful indicator of a person's social capital, therefore what is generally said or believed about a person's character will have a major impact on a person's life. A positive reputation will make our life easier within our community, just as a bad reputation will make our life harder: an honest shopkeeper will gain from having a noble reputation as much as a dishonest shopkeeper will suffer from having a ruthless reputation. Our task is to think of creative ways in which the institutions of justice can use the stick of a bad reputation or the carrot of a good reputation as incentives to get citizens to embrace the principles of fairness and impartiality. The ultimate goal of course is that by giving citizens a (rational) incentive to be fair and impartial, citizens may act accordingly and eventually internalize the norms of fairness and impartiality.



One important tool at the disposal of the institutions of justice is to enforce the rule of publicity. There is a long tradition in the history of philosophy endorsing this rule, which includes among its advocates both Kant and Bentham.¹⁹ The rule of publicity says that each citizen ought to justify their actions to others, in the open, under public scrutiny. Thus, Robert Goodin (1992) praises the moral powers of publicity, suggesting that publicity will work to make people behave more morally, while Gutmann and Thompson (1996) argue that the principle of publicity is a fundamental requirement of deliberative democracy. Even US President Woodrow Wilson could not deny the moral force of publicity: ‘publicity is one of the purifying elements of politics. Nothing checks all the bad practices of politics like public exposure’.²⁰

What is it about publicity that brings out the best in us? It is not that it awakens our latent moral virtues, instead it is the fear of gaining a bad reputation, or the anticipation of a good reputation, which turns us into good citizens. In the last analysis, we value our reputation for prudential reasons. It is only when actions are made public that reputations are established, therefore everyone has a reason (on prudential grounds) to be on their best behaviour. As Goodin (1992, 124) explains, ‘publicity will work to make people behave more morally only insofar as they would be embarrassed at being seen to behave indefensibly or immorally’. Through the rule of publicity, the institutions of justice can appeal to our prudential concern for our reputation to deter any inclination we may have to forego the principles of fairness and impartiality.

What I am suggesting is not as radical or blasphemous as it may seem. Before damning accusations of social engineering are made, we should recall that J.S. Mill was advocating something along these lines in 1861, and Mill can hardly be accused of dictatorial tendencies. In *Considerations on Representative Government*, Mill (1991, 362) famously endorses the idea of public voting, on the grounds that ‘external inducements’ are necessary to make a person perform a duty to the public: ‘The best side of their [the electorate] character is that which people are anxious to show, even to those who are no better than themselves. People will give dishonest or mean votes from lucre, from malice, from pique, from personal rivalry, even from the interests or prejudices of class or sect, more readily in secret than in public’. While there are good reasons why we may want to keep voting secret, there may be a case for forcing public figures involved in democratic decision making to justify their actions in public. For example, the elected representatives in a democracy ought to publicly declare how much they earn privately, as well as reveal the source of their personal incomes. Similarly, political parties should declare publicly where their funding comes from. Being concerned with their reputation, the rule of publicity will induce politicians and political parties to act according to the principles of justice.



This is only a brief elaboration of how, in one specific case, the institutions of justice (via the rule of publicity) may target our rational incentive structure (reputation) with a view to foster our sense of egalitarian justice. Clearly, much more work will need to be done before these suggestions are presented as serious social reforms. The point here was not to suggest a specific public policy, but only to give an idea of the practical implications of embracing what I have called a neo-Humean approach to motivations and justice.

Conclusion

For too long neo-Kantian egalitarians have argued that a theory of social justice must be rooted in moral motivations of reasonableness. This has opened neo-Kantians like Rawls to the charge of prescribing a political philosophy that is not 'followable'. For egalitarianism in general to be taken seriously, it is imperative for it to endorse Hume's idea of mixed motivation. Our actions are the result of our feelings of sympathy and self-love, although these two motivations are not equally balanced. The former (sympathy) is usually weaker than the latter (self-love). If Hume is right, and I suspect he is, then the sooner egalitarians are prepared to work with rational motivations, the better it will be for the long-term prospects of egalitarian justice.

Combining sympathy with self-love means that we do not have to choose between a theory of justice grounded on rationality, which is morally abhorrent but realistic, and a theory of justice grounded on reasonableness, which is morally attractive but unfollowable. Coming to terms with our mixed motivations, and taking on board self-regarding motivations as well as other-regarding motivations, has profound implications for egalitarian justice. Above all, it challenges us to rethink the concept of justice in vogue at the moment. A broader understanding of social justice is needed, which includes promoting social, economic and political institutions that work on the incentive structures of citizens in order to give 'actual people' rational incentives to act according to the principles of fairness and impartiality, and therefore foster their sense of egalitarian justice.

The first step towards promoting an egalitarian society is to accept the reality that individuals are not always motivated by other-regarding feelings of fairness and impartiality. We all fall prey to self-regarding passions, therefore if left to our own devices, actual people will almost inevitably under-achieve. The second step is to give people a positive rational incentive to act upon the principles of fairness and impartiality. Having an egalitarian sense of justice is not sufficient. People must also be given a reason to endorse and act upon their sense of justice. Finding ways to motivate people to act justly is arguably the most pressing challenge facing egalitarian justice today.



Date submitted: 4 November 2002

Date accepted: 13 May 2003

Notes

- 1 Earlier versions of this article were presented at Yale University; University of Colorado, Boulder; University of York; University College Dublin, Trinity College Dublin; and at the APSA Meeting (Boston, September 1998). I am grateful to Iseult Honohan, Attracta Ingram, Susan Mendus, Claudia Mills, Philippe Van Parijs, Federico Varese, Andrew Williams, and the editors of *Contemporary Political Theory* for their comments on earlier versions of this article. I am especially grateful to Dario Castiglione, who read many drafts of this paper, and saved me from several blatant errors. Regrettably the responsibility for the article's remaining faults rests squarely on my shoulders.
- 2 To be precise, Barry (1995, 46) argues that a theory of justice may be characterized by its answer to three questions: 'First, what is the motive (are the motives) for behaving justly? Secondly, what is the criterion (are the criteria) for a just set of rules? And thirdly, how are the answers to the first two questions connected?'
- 3 For an account of justice as mutual advantage *vs* justice as impartiality, see Barry (1989). For a comparison of neo-Hobbesian and neo-Kantian social contract theories, see Bufacchi (2000).
- 4 Gibbard calls this the Hume–Ramsey theory of rationality, which is the Humean idea of instrumentality combined with Ramsey's emphasis on formal coherence of preferences.
- 5 The fact that Rawls allows for both rationality and reasonableness suggests perhaps that he is not indifferent to a Humean position. His endorsement of Hume's conditions of justice is also indicative. On the affinity between Rawls and Hume on justice, see Barry (1989, 145–178). See also Rawls' chapter on Hume in Rawls (2000).
- 6 For a more detailed analysis of 'rationality' and 'reasonableness', see Bufacchi (1998).
- 7 The distinction between lower and higher order is taken from de Wijze (2002, 175), who identifies 'a political notion of reasonableness understood as a higher order concept with its own normative and epistemological content'.
- 8 Moore (1994) rightly chastizes Gauthier for committing the naturalistic fallacy. See also Barry (1995), Buchanan (1990), and Lehning (1990).
- 9 On this point, see Barry (1995, 56–57).
- 10 Political philosophers have always put faith on moral education to solve their problems. For an account of moral education in Enlightenment thought, from Locke to Helvétius, see Parry (2000).
- 11 My views on what a neo-Humean theory of justice may look like have been significantly influenced by Simon Blackburn's (1998) outstanding book *Ruling Passions*.
- 12 For a direct comparison between Kant and Hume on moral motivations, see Thomas (1988) and Nagel (1970).
- 13 On this point, see Nagel (1970, 10).
- 14 On the human potential for both good and evil, explored within the context of the rescue of Jews from Nazi-occupied Europe, see Geras (1998).
- 15 This is of course Rawls' (1972, 2–3) famous definition of social justice, which has almost universally been endorsed in contemporary debates on justice.
- 16 'The explanation of why [justice as fairness] is practicable must be of the right kind. The problem of stability is not that of bringing others who reject a conception to share it, or to act in accordance with it, by workable sanctions, if necessary, as if the task were to find ways to impose that conception once we are convinced it is sound' (Rawls, 2001, 186).



- 17 'Actual people' refers to you and I, not the 'artificial persons' that represent us in the original position.
- 18 That social institutions have a role to play in fostering beliefs in justice and equality is a view that is slowly gaining favor among egalitarians. For example, Scheffler (1992, 137 and 138) argues that neither the prevalence of the relevant psychological structures nor the degree of conflict between morality and the agent's interests is fixed and immutable, instead 'both are strongly influenced by social institutions and practices that are not themselves unchangeable'. Furthermore, 'social institutions have a role to play in fostering those beliefs that are in turn required if certain basic sorts of human relationship are to flourish'.
- 19 See Kant (1970) and Postema (1989).
- 20 Quoted in Goodin (1992, 126).

References

- Barry, B. (1989) *Theories of Justice*, Berkeley, CA: University of California Press.
- Barry, B. (1995) *Justice as Impartiality*, Oxford: Oxford University Press.
- Blackburn, S. (1998) *Ruling Passions*, Oxford: Oxford University Press.
- Buchanan, A. (1990) 'Justice as reciprocity versus subject-centered justice', *Philosophy and Public Affairs* 19(3): 227–252.
- Bufacchi, V. (1998) 'Reasonable agreement: a defence', *Imprints* 2(2): 223–240.
- Bufacchi, V. (2000) 'The Enlightenment, Contractualism and the Moral Polity', in N. Geras and R. Wokler (eds.) *The Enlightenment and Modernity*, Basingstoke: Macmillan.
- Camus, A. (1963) *The Fall*, Harmondsworth: Penguin.
- de Wijze, S. (2002) 'The political limits of reasonableness', *Imprints* 6(2): 171–186.
- Gauthier, D. (1986) *Morals by Agreement*, Oxford: Oxford University Press.
- Gauthier, D. (1997) 'Political contractarianism', *The Journal of Political Philosophy* 5(2): 132–148.
- Geras, N. (1998) *The Contract of Mutual Indifference: Political Philosophy After the Holocaust*, London: Verso.
- Gibbard, A. (1990) *Wise Choice, Apt Feelings*, Oxford: Oxford University Press.
- Goodin, R. (1992) *Motivating Political Morality*, Oxford: Blackwell.
- Gross, M. (1997) *Ethics and Activism: The Theory and Practice of Political Morality*, Cambridge: Cambridge University Press.
- Gutmann, A. and Thompson, D. (1996) *Democracy and Disagreement*, Cambridge, MA: Harvard University Press.
- Hobbes, T. (1994) *Leviathan*, Indianapolis: Hackett.
- Kant, I. (1970) 'Perpetual Peace', in H. Reiss (ed.) *Kant's Political Writings*, Cambridge: Cambridge University Press.
- Lehning, P. (1990) 'Right constraints? an analysis of Gauthier's reasoning about morals', *Acta Politica* XXV (January): 3–36.
- Mill, J.S. (1991) *On Liberty and Other Essays*, Oxford: Oxford University Press.
- Moore, M. (1994) 'Gauthier's Contractarian Morality', in D. Boucher and P. Kelly (eds.) *The Social Contract from Hobbes to Rawls*, London: Routledge.
- Nagel, T. (1970) *The Possibility of Altruism*, Princeton: Princeton University Press.
- O'Neill, O. (1996) *Towards Justice and Virtue*, Cambridge: Cambridge University Press.
- Parry, G. (2000) 'Education Can Do All', in N. Geras and R. Wokler (eds.) *The Enlightenment and Modernity*, Basingstoke: Macmillan.
- Postema, G. (1989) 'Bentham on the public character of law', *Utilitas* 1: 41–61.
- Rawls, J. (1972) *A Theory of Justice*, Oxford: Oxford University Press.



-
- Rawls, J. (1980) 'Kantian constructivism in moral theory', *The Journal of Philosophy* 77(9): 515–572.
- Rawls, J. (1989) 'The domain of the political and overlapping consensus', *New York University Law Review* 64(2): 233–255.
- Rawls, J. (1993) *Political Liberalism*, New York: Columbia University Press.
- Rawls, J. (2000) *Lectures on the History of Moral Philosophy*, Cambridge, MA: Harvard University Press.
- Rawls, J. (2001) *Justice as Fairness: A Restatement*, Cambridge, MA: Harvard University Press.
- Scanlon, T.M. (1982) 'Contractualism and Utilitarianism', in A. Sen and B. Williams (eds.) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.
- Scanlon, T.M. (1998) *What We Owe to Each Other*, Cambridge, MA: Harvard University Press.
- Scheffler, S. (1992) *Human Morality*, Oxford: Oxford University Press.
- Thomas, L. (1988) 'Moral motivations: Kantians versus Humeans (and evolution)', *Midwest Studies in Philosophy* XII: 367–383.