

Institutions as motivational bootstrapping

Peter Dietsch

Centre de recherche en éthique de l'Université de Montréal

peter.dietsch@umontreal.ca

DRAFT

- please do not cite or circulate without permission from the author -

1 Introduction

As is the case with many debates in contemporary political philosophy, the question of what kind of institutional structure, if any, needs to be in place for a society to be just can be traced back to the work of John Rawls. The point of contention here is not primarily the content of Rawls' two principles of justice, but their domain. "Justice," one of the first sentences of *A Theory of Justice* famously states, "is the first virtue of social institutions." (1999: 3, my italics). The institutional focus of Rawls' conception of justice has proved tremendously controversial. Critics have suggested that a just society builds on principles and behaviour at the *individual* level instead. Despite the fact that both sides of this debate acknowledge that an exclusive reliance on principles of either institutional or individual justice would be misplaced, the literature has paid little attention to the question of what the interaction between these two levels does, and should, look like.

This paper aims to fill this gap. Following a brief outline of the main features of Rawls' account of institutional justice, I will set out two critiques of this aspect of the Rawlsian system. The first, formulated by G.A. Cohen (1997), questions Rawls' emphasis on a *formal* institutional structure to promote justice in society. Moreover, Cohen can be interpreted as belonging to a family of views which, contrary to Rawls, defend a *continuity* between individual and institutional principles of justice. The second critique, put forward by Liam Murphy (1999), highlights the shortcomings of the

Rawlsian framework when extended to non-ideal theory, i.e. when there exists a problem of non-compliance. My objective in analysing these critiques is not primarily exegetical, but rather to gain insight into the interplay between individual and institutional justice. Rawls' theory and the critiques by Cohen and Murphy are stepping-stones in this exercise.

Subsequently, I will use the insights from the above debate to propose an account of social justice that integrates the institutional and the individual level. *Institutions as motivational bootstrapping* suggests that for any institutional structure to find the adherence and support necessary to sustain it, the gap between the institutional demands and individual conceptions of justice may not grow too wide. This perspective, which will be spelt out in more detail below, proves particularly useful for understanding the dynamics of institutional change and their motivational underpinnings at the individual level.

In a final section, I will present some of the potential implications of this paper for the prospects of implementing a basic income guarantee.

2 Institutional justice and its critics

Once we accept Rawls' characterisation of society as a co-operative venture for mutual advantage, the need for a system of rules that governs the distribution of benefits and burdens involved in this venture becomes fairly obvious. Though the members of society share the interest of upholding the co-operation and thereby increasing their standard of living, they argue over who should get what share of the co-operative surplus. This is the classic characterisation of the "circumstances of justice," which Rawls borrows from David Hume. In order to arbitrate between conflicting interests, we need a set of principles of justice.

That said, Rawls acknowledges the existence of principles at the individual level, too. Examples include the duty not to harm the innocent, a principle of mutual respect, or a maxim of beneficence (cf. chapters 18 and 19 of *A Theory of Justice*). Yet Rawls contends that the question of distributing the benefits and burdens from co-operation needs to be addressed exclusively at the institutional level.

Importantly, Rawls complements his account of institutional justice with an equally controversial assumption about individual acceptance of the principles of justice in question. He explicitly limits his analysis to what he calls strict compliance theory (cf. Rawls, 1999, 8), that is he assumes that all members of society do in fact respect the principles of justice once they have been decided upon behind the veil of ignorance. As we shall see later, this assumption obscures some of the challenges that we face in promoting justice in the real world.

I will now turn to two criticisms that have been proposed of Rawls' focus on institutional justice. Start with Cohen. At the outset of the relevant paper, he helpfully summarises his qualms with Rawls' position in the slogan that "the personal is political," and then proceeds to spell out that he will use this slogan to argue that "principles of distributive justice [...] apply, wherever else they do, to people's legally unconstrained choices." (1997: 3) It is worth emphasising, as Cohen does, that these legally unconstrained choices obviously fall outside the realm of compliance, where the latter is defined as conformity to the law. If Cohen were correct in identifying this formally unregulated realm of choice as an important, perhaps even the principal, "site" of distributive justice, this would undermine Rawls' presentation of justice as primarily a virtue of institutions. The status of just institutions would be downgraded from being a sufficient condition to address problems of distributive justice to forming one necessary, but potentially minor, ingredient into a broader mix of factors.

Cohen's argumentative strategy is to present Rawls with a dilemma. If you want to limit the subject of justice to "the basic structure" of society, you will face the following choice. You can either restrict the notion to coercive, i.e. legal, institutions, or you can extend it to encompass the informal structure of society, which relies on convention and usage. The former, so Cohen reasons with an argument that I bracket here, results in an arbitrary notion of justice, whereas the latter inevitably incorporates the aforementioned legally unconstrained choices into the realm of justice. Cohen himself is sympathetic to this second option and suggests that formally unconstrained choices need to be informed

by an *ethos* of justice.¹ To illustrate his point, Cohen applies his argument to Rawls' difference principle, which holds that income inequalities are justified as incentive payments to the talented if they are to the benefit of the most disadvantaged group in society. In this context, the dilemma takes the following form. Either the talented do not affirm the difference principle in their market behaviour, which represents one principal instance of a legally unconstrained context of choice. In this case, the society in question would no longer qualify as just measured by Rawls' own standards. Or the talented do endorse the difference principle as well as the egalitarian ethos it encapsulates, but then it is no longer clear why the incentive payments to them are necessary in the first place.²

This tension inherent in Rawls' difference principle uncovers a general puzzle for any modern liberal society. When asked to name the defining features of such a society, many would place the negative liberty it grants people in their private and professional lives near the top of the list. Yet, at the same time, the functioning of a liberal society builds on the continued political support of its citizens for an institutional framework which, in any given instance, might well ask them to subordinate their private interest to the well-being of others in the community. To some observers, reconciling these two aspects of liberalism requires people to perform a motivational split that borders on schizophrenia. As Murphy and Nagel poignantly put it, "does it really make sense – is it psychologically coherent – for people to be acquisitive, competitive, and dedicated to advancing the interests of themselves and their families in their personal lives while being impartially concerned with the interests of all, and with reducing inequalities between families, in their political choices – choices, for example, to support taxes on themselves for the benefit of others?" (Murphy and Nagel, 2002: 72) Incidentally, this

¹ To pre-empt confusion, let me point out that the notion of a social ethos has also been used in a second, different, sense. Jonathan Wolff (1998: 104), in calling for an *egalitarian social ethos*, asserts that "there is more to a society of equals than a just scheme of distribution of material goods." What he has in mind is a plurality of egalitarian values that includes not only fairness, but also respect for the individual. Wolff is concerned with the content of the ethos, whereas Cohen emphasises the contrast to formal institutional structures.

² For an excellent illustration of the contingency of the incentives needed to elicit a certain contribution to the productive process from individuals, cf. Carens (1981). Suppose economic activities were driven less by a rationale of profit-maximisation, and more by a moral incentive to perform one's duty to society. Under such a scenario, Carens argues, our current social arrangements of the productive process would be compatible with a considerably less unequal distribution of the benefits from this process.

explains why liberals of all stripes cherish the institution of the market, in its idealized form. Under the mechanism of the “invisible hand,” as presented by Adam Smith, objectives pursued for private gain are assumed to generate public benefits. In other words, the problem of the motivational split vanishes. However, the idealized market does unfortunately not exist in this pure form in practice.

Opinions are divided about how best to conceive of morality in the face of this motivational split. Rawls and Cohen fall on opposing sides of this divide. On the one hand, the *discontinuity* view holds that there are indeed two different sets of principles of justice. Whereas individuals may be partial in their actions, the conduct of the state and its representatives has to be governed by an impartial concern for all citizens. The only constraint imposed on individuals is that they are required to support the impartial public institutions. This, essentially, is the picture we are presented in Rawls’ theory of justice.³ The *continuity* view, on the other hand, insists that the same principles of justice apply to both individuals and institutions. However, advocates of the continuity view accept that a “division of labour between social institutions and individual responsibility is the most effective way to promote the ends of morality.” (Murphy and Nagel, 2002: 71) Note that one plausible division of labour between individuals and institutions might assign the task of impartial institutional design to the latter and compliance with this institutional framework to the former. Under such a scenario, the continuity and discontinuity view become indistinguishable in terms of their practical consequences.⁴

It is not my intention here to take sides in the continuity-discontinuity debate. Instead, let me suggest that we can draw a conclusion from what seems to be the common denominator between continuity and discontinuity theorists, namely that there is and

³ Note also that, in the original position, the motivational split vanishes, since individuals are stripped of any information that could provide the foundation for partial concerns.

⁴ For completeness’ sake, I should mention that the continuity-discontinuity debate has also been conducted at a different, perhaps even more fundamental, level. Ronald Dworkin’s (1990) case for continuity has the ambitious goal of finding an ethical foundation for liberalism. Similar to Cohen, the target of Dworkin’s arguments is also John Rawls, in this case his project of defending liberalism as a political doctrine (which later finds expression in his *Political Liberalism* (1993)). This debate is enriched by adding, and giving centre stage to, another variable in the political process, namely the possibility of moral disagreement between members of the community. My inclination is to side with Rawls, but we cannot pursue this question further.

needs to be a division of labour between individuals and institutions in the promotion of distributive justice. Since institutions are not independent entities, but are made up of individuals, this division of labour is subject to the motivational split we identified before. Now, it is plausible to assume that when this motivational split becomes too wide and hence too demanding, people's allegiance to the institutional framework will give way. This brings us back to the aspect of Cohen's position we started off with, namely the idea of a social ethos. Such an ethos can be instrumental in narrowing the motivational split in two ways. Suppose, to make its implications more tangible, a social ethos that requires people not to walk past a homeless person on the street without giving them some change. For this maxim to be a true social ethos, members of society have to internalise its content at least to some degree, which will furnish *internal* reasons to give change to a homeless person when they see one. At the same time, the social ethos functions as a device of institutional justice itself, in that the social stigma attached to violating it provides people with *external* reasons to conform. People will not walk past a homeless person because it triggers too costly a social reaction from others. Both sets of reasons taken together, it would seem, increase the likelihood of people to either comply with or even vote for *formal* institutional structures which encapsulate the spirit of the ethos. In our example, this might be taxation to fund accommodation and reintegration projects for the homeless.

These considerations reveal two things.⁵ First, the difference between formal and informal institutional structures is really only one of degree.⁶ Formal institutional structures like laws *tend* to rely on external reasons to exert compliance, whereas informal institutional structures like a social ethos *tend* to rely on internal reasons for their effectiveness. Yet, it is quite clear that our legal structure would collapse if people did not, at least to some extent, identify with it. Correspondingly, the more a social ethos draws on the undesirable consequences that ensue in case one violates it, the more it

⁵ A third aspect, which I ignore here, could be to reassess Cohen's critique of Rawls' difference principle in light of the above considerations. For a defence of Rawls in this vein, cf. Andrew Williams (1998).

⁶ Cohen implicitly acknowledges this possibility in the endnote to his article (1997), by conceding that behaviour is constitutive not only of informal social structures, but of formal or coercive ones, too. For a contribution to the literature which, similar to my argument here, softens the distinction between institutional framework and ethos from a categorical one to one of degree, cf. Joshua Cohen (2002).

looks like a law and becomes itself subject to a motivational split. When the support for them crumbles, both laws and a social ethos either change or collapse.⁷

Second, institutional justice is well advised to draw on both its formal and its informal elements. In fact, from a public policy point of view, the more we can rely on informal structures and the internal reasons they furnish, the better. They provide the “glue” that promotes and maintains social cohesion.

Let us now briefly turn to the second critique of Rawls’ institutional focus. Liam Murphy (1999) describes Rawls’ approach as *dualistic*, given its bifurcation of the normative realm into principles governing institutions versus ones directed at personal conduct.⁸ Murphy, like Cohen, concludes that limiting the subject of social justice to the institutional level is inappropriate, yet his reasoning to arrive at this conclusion is different. Murphy’s principal worry is that a theory with different standards of justice for the institutional and individual level “will yield an implausible account of what people should do in *nonideal* circumstances.” (1999: 279, my italics) Murphy acknowledges that Rawls explicitly limits the scope of his theory of justice to ideal theory,⁹ but he rightly insists that the value of this ideal theory will stand and fall depending on how informative it proves in tackling problems of injustice in the real world. Given an unjust situation, Murphy questions the Rawlsian idea that we should prioritise the moral demand to push for institutional reform over the moral demand to alleviate the injustice in a more direct way. Murphy gives the example of people living in extreme poverty in developing countries. Rather than adopting the abstract and distant goal of promoting just

⁷ This paragraph also draws our attention to the possibility that the content of a social ethos, much like that of a law, may *conflict* with justice. Such a situation is particularly likely to occur under an unjust *status quo*, with the ethos in question only being endorsed by a subgroup of the population, namely those who benefit from the injustice. This comes close, I believe, to Cohen’s interpretation of the role of incentives in Rawls’ difference principle. Can the *attitude* or *ethos* of the privileged to demand incentive payments to make their contribution really be justified from a perspective of justice? I think Cohen has a strong point here.

⁸ As far as I can see, the dualism *versus* monism debate significantly overlaps with the continuity *versus* discontinuity debate.

⁹ Cf. “Obviously the problems of partial compliance theory are the pressing and urgent matters. These are the things we are faced with in everyday life. The reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems.” (Rawls, 1999: 8)

international institutions, one should consider giving my resources to humanitarian aid agencies instead (cf. 1999: 281).

My principal reservation towards Murphy's position is the following. The individual in Murphy's example is motivated to alleviate poverty, and wonders about how best to achieve this aim. This seems like a highly idealistic assumption, especially under an unjust *status quo*. The more unjust the distributive *status quo*, the wider the motivational split between the personal and impersonal viewpoints, and the more of a gamble it becomes to rely on individual motivation to do something about the injustice. In fairness to Murphy, I should emphasise that he is acutely aware of this problem. His way of expressing it is to say that in non-ideal theory, i.e. when some members of society are not complying to the principles of justice, the moral demands on the other members will be excessive (cf. 2000, and 1999: 291) But instead of acknowledging the need for an institutional solution to tackle the problem of non-compliance, Murphy focuses his attention on modifying the moral demands on those who do comply. I think he is putting his money on the weaker horse.

In sum, and contrary to Murphy, I believe that as we move further away from ideal theory, to wit, as the distributive structures of a society become more and more unjust, institutional change becomes more rather than less important.

3 Motivational bootstrapping

The preceding discussion shows that justice does not have to choose between individuals and institutions for its promotion. It necessarily relies on both.¹⁰ I will now suggest that the interaction between the two levels can be pictured as a process of *motivational bootstrapping*.

¹⁰ Incidentally, the criticisms presented in the previous section notwithstanding, I believe Rawls was well aware of this. Cf. for instance: "An institution may be thought of in two ways: first as an abstract object, that is, as a possible form of conduct expressed by a system of rules; and second, as the realisation in the thought and conduct of certain persons at a certain time and place of the actions specified by these rules." (1999: 48) Besides, his notion of a *sense of justice* can be interpreted as a concept very similar to that of a social ethos.

In order to unpack this metaphor, recall the conclusion from the debate between Rawls and Cohen that the notion of institutional justice should be broadened to encompass informal social structures.

One of the conclusions from this exercise has been to broaden the notion of institutional justice to encompass informal social structures. We saw that informal social structures, which can for instance take the form of a social ethos, are likely to be a preferable vehicle to promote social cohesion. The internal reasons that a social ethos speaks to, open the door to the motivation of individuals, and by extension to their resistance to change, in a way the external reasons relied upon by laws rarely do. Here is the snag. Under an unjust *status quo*, it is unlikely that such a social ethos exists, and its absence represents the principal obstacle in transforming the unjust structure in question.

The natural question to ask at this point is whether we can find ways to *establish* a social ethos of the required kind. Let me suggest that our starting point in this exercise should not be the shortcomings of institutional justice, but rather what it can achieve. Few people would deny that even formal structures like laws can have a significant formative influence.¹¹ In fact, this feature is co-extensive with our assertion that, though laws *tend* to provide external reasons, internal ones are not out of their reach as a matter of principle. On the other hand, there always exists a tension between people's personal motives and the duties asked of them for the benefit of others in the community. Given that even the authority of existing institutions, be they formal or informal in nature, will be undermined if this tension grows too strong, it is not surprising that the motivational split acts as a severe constraint on the feasibility of new institutions.

We can now define *motivational bootstrapping*. The idea is to pursue incremental institutional change, which delivers the benefits of formative influence without overstretching the motivational split, thereby gradually expanding the realm of the

¹¹ Unsurprisingly, this idea is prominent in the thought of John Rawls. He finds a poignant formulation for it when he says that "what sort of persons we are is shaped by how we think of ourselves and this in turn is influenced by the social forms we live under." (1975: 300) Joshua Cohen (2002: 377-80) provides a tangible illustration of the policy implications that taking the formative influence of institutions seriously is likely to have. For instance, if a consensual as opposed to majoritarian form of democracy delivers better results for the least advantaged in society, should we not have a preference for the former from the perspective of justice?

politically feasible. The secret of transforming unjust distributive structures, according to this view, lies in “managing” the motivational split. Sure enough, the privileged will resist sweeping reforms that would erase their privileges overnight. Yet, might we not be able to find policies that will be both palatable to them and, over time, lead them to accept more fundamental change?

Let me add a few comments on this notion of institutions as motivational bootstrapping. First, the advantage of conceptualising institutional justice in this way is that it captures both its power and its limits. Institutional design, if conducted carefully, indeed represents a potent instrument to foster social consensus. The transformation of an unjust structure is particularly unlikely to succeed without capitalising on this feature. At the same time, we have seen that the feasibility of institutional reform is bound by the motivational split. The latter must not be overstretched for institutional justice to be effective. If we do the boot up too tightly, i.e. if the demands of the formal and informal institutions in society are too far removed from people’s personal motives, the lace will rip.

Second, the notion of motivational bootstrapping suggests that what people perceive as acceptable policy in the name of justice varies with circumstances. In a society with a social ethos of sorts, the set of feasible policies to promote economic justice will be larger than in a comparable society that lacks this ethos. This, one might object, is evidence that motivational bootstrapping is not a moral concept, but essentially a pragmatic one. One way to meet this criticism is to accept a personal viewpoint as a constitutive feature of our moral make-up (cf. for instance Nagel 1991).

Third, consider a couple of admittedly very general examples for motivational bootstrapping in practice. It has been successfully applied in Scandinavian countries, for instance, where citizens express a preference for the relatively large welfare state operated by their governments. Attempts to significantly gear up the welfare state overnight in other countries, where no corresponding social ethos exists, would be an uphill struggle, because it would overstretch the motivational split. Moreover, consider the economic transition of Eastern European countries during the 1990s. Their teething

troubles suggest that the functioning of the market requires both people's acquaintance with, and their support of, its institutions – features that clearly take time to develop.

This leads me to a final remark. Unfortunately, the social glue that motivational bootstrapping tries to create is much like a ligament. Ripping it is a matter of an instant, rebuilding takes time. The management of this long-haul process is not made easier by a political system that is governed by election-cycles rather than political leadership.

4 Implications for Basic Income

I see two main, and related, challenges in implementing a basic income guarantee. The first consists in overcoming the lack of political will for the project. Yet, this obstacle could be characterised as a mere symptom. After all, the political will would most likely emerge if the political representatives sensed a willingness among the electorate to support such a measure.

I believe that any successful attempt to promote more widespread acceptance of a basic income guarantee will have to work a descriptive and a normative lever at the same time. On the descriptive level, scepticism towards basic income is partly based on the uncertainty that surrounds the practical consequences that such a scheme would entail. What are the fiscal implications for different groups within the population? Will people reduce their work effort? Will there be resentment towards those who choose to do so? Implementing a basic income guarantee on a sufficiently large scale, in a country of five to ten million people, say, could help to allay fears about potential negative side-effects. In addition, it may reveal weaknesses that would have to be addressed if the scheme were to be applied more comprehensively. Whereas speculation about the empirical ramifications of basic income based on opinion surveys can point in very different directions, such a real life test of the scheme promises to provide more reliable insights. That said, transferring these insights from the specific circumstances of the test to other contexts may still encounter difficulties.

As to the normative lever, the proposal of a basic income guarantee is up against the strong sense of entitlement to one's pretax income that characterises most societies operating in a market economy. A tax designed to fund a basic income guarantee that is

independent of one's contribution to the productive process conflicts with this notion of entitlement. In order to resolve this conflict in favour of a higher level of political acceptance towards a basic income guarantee, people need to be prepared to question this entitlement to pretax income. One recent argument in this generally underexplored direction has been presented by Liam Murphy and Thomas Nagel in their book *The Myth of Ownership*. They point out that in the current debate, "[t]he natural idea that people deserve to be rewarded for thrift and industry slides into the much broader notion that all of pretax income can be regarded as a reward for those virtues." (Murphy and Nagel, 36) Murphy and Nagel point out that, instead of uncritically accepting this move, the differences in pretax income have to be part of our normative deliberation. As they poignantly put it, "[e]valuation must decide how "mine" and "yours" ought to be determined; it cannot start with a set of assumptions about what is mine and yours." (Murphy and Nagel, 75)

In the context of this paper, this latter, normative challenge that needs to be negotiated in introducing a basic income guarantee can be expressed as follows. Pushing through an institutional basic income guarantee given the current configuration of conceptions of justice in modern market economies runs a considerable risk of overstressing the motivational split. In order to make a basic income guarantee politically feasible, individual convictions about what constitutes a just distribution of income need to change. In particular, there needs to be increased awareness of the fact that pretax incomes are to some extent conventional, and that these conventions are subject to normative scrutiny. If basic income can dispel the belief in a strong entitlement to pretax income, its chances of implementation will increase dramatically.

5 Conclusion

- This paper has tried to defend the Rawlsian focus on institutional justice against two recent critiques. Against Cohen, it can be pointed out that the informal institutional structures he advertises are in fact relevantly like their formal cousins in that they also crumble once support from individual convictions breaks away.

Against Murphy's critique, we can question the hypothesis that the pursuit of institutional justice becomes a lesser priority in circumstances of *injustice*.

- That said, the criticisms of institutional justice help us gain a better understanding of the constraints under which it operates. I have proposed that institutional justice, in order to be effective, needs to ensure that the motivational split between institutional demands and individual convictions does not grow too wide. At the same time, exploiting the formative influence of institutions can help us to move the boundaries of institutional change over time. Hence the characterisation of institutions as a process of motivational bootstrapping.
- I have suggested that there is one conviction in particular that is likely to act as a stumbling block for any attempt to introduce a basic income guarantee. The strong sense of entitlement to pretax income, which is prevalent in contemporary market economies, needs to be questioned.

6 Bibliography

Joseph H. **Carens** (1981), *Equality, Moral Incentives, and the Market*, The University of Chicago Press.

G.A. **Cohen** (1997), “Where the Action Is: On the Site of Distributive Justice”, *Philosophy and Public Affairs* 26/1, 3-30.

Joshua **Cohen** (2002), “Taking People as They Are?”, *Philosophy and Public Affairs* 30/4, 363-86.

Ronald **Dworkin** (1990), “Foundations of Liberal Equality”, in: Grethe B. Peterson (ed.), *The Tanner Lectures on Human Values*, vol. XI, 1-120.

Liam **Murphy** (2000), *Moral demands in nonideal theory*, Oxford University Press.

Liam **Murphy** (1999), “Institutions and the Demands of Justice”, *Philosophy and Public Affairs* 27/4, 251-91.

Liam **Murphy** and Thomas **Nagel** (2002), *The Myth of Ownership*, Oxford University Press.

Thomas **Nagel** (1991), *Equality and Partiality*, Oxford University Press.

John **Rawls** (1999), *A Theory of Justice*, 2nd edition, Oxford University Press.

John **Rawls** (1993), *Political Liberalism*, Columbia University Press.

Andrew **Williams** (1998), “Incentives, Inequality, and Publicity”, *Philosophy and Public Affairs* 27/3, 225-47.

Jonathan **Wolff** (1998), “Fairness, Respect, and the Egalitarian Ethos”, *Philosophy and Public Affairs* 27/2, 97-122.