



AARHUS UNIVERSITY



# Coversheet

---

**This is the accepted manuscript (post-print version) of the article.**

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

**How to cite this publication**

Please cite the final published version:

Hartz, E., & Nielsen, C. F. (2015). *From conditions of equality to demands of justice: equal freedom, motivation and justification in Hobbes, Rousseau and Rawls*. *Critical Review of International Social and Political Philosophy*, 18(1), 7-25. DOI: 10.1080/13698230.2014.995498

## Publication metadata

**Title:** From conditions of equality to demands of justice: equal freedom, motivation and justification in Hobbes, Rousseau and Rawls  
**Author(s):** Emily Hartz & Carsten Fogh Nielsen  
**Journal:** Critical Review of International Social and Political Philosophy  
**DOI/Link:** <http://dx.doi.org/10.1080/13698230.2014.995498>  
**Document version:** Accepted manuscript (post-print)

**General Rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## **From conditions of equality to demands of justice**

### **Equal freedom, motivation and justification in Hobbes, Rousseau and Rawls**

Emily Hartz<sup>1</sup> and Carsten Fogh Nielsen<sup>2</sup>

#### **Introduction**

There are at least two ways to conceptualize the notion of justice as equality. On the one hand equality can be seen as simply a purely formal requirement on what it means for a state of affairs to be just, where equality is commonly defined along the lines of “treating like cases alike”. According to this conception justice as equality simply means something like “equality before the law” or “equal treatment by the law”, where this formal requirement for equal treatment is completely indifferent to the actual content of the law.

However justice as equality can also be interpreted in another, more substantial way. Equality can be seen as not simply a formal requirement of justice but as the foundation or normative framework *from* which and *within* which the demands of justice are to be justified and derived. According to such a conception equality is a substantial normative category, which should (at least partly) define the content of justice itself.

In both of these two conceptions equality is claimed to be a basic and fundamental part of justice. To make sense of this claim, however, we need to be much more specific about *how* and *why* justice and equality are related. The formal conception of justice as equality seeks to avoid or preempt this question by simply stipulating that equality, in the sense of equal treatment, *is* a formal requirement of justice. But simply stipulating that something is the case does not answer the question of *why* it is so or *why* we should accept this particular stipulation.

The substantial conception of justice as equality on the other hand, according to which equality is the normative foundation for justice, obviously needs an account of how and why equality obtains normative force. More precisely it needs to explain how and why a certain *condition*, a condition of equality, can be (or be transformed into) a *demand* of justice.

It is this second set of questions, and thus the substantial conception of justice as equality, which is the subject of this paper. More precisely the aim of the paper is to clarify and briefly discuss how the contractualist tradition, in casu Hobbes, Rousseau

---

<sup>1</sup> Associate professor, Department of Law, University of Southern Denmark, Denmark. Email: eml@sam.sdu.dk

<sup>2</sup> Postdoc, Department of Culture and Society, Aarhus University, Denmark. Email: filcfn@hum.au.dk

and Rawls, have approached the twin questions of how and why a condition of equality translates into a conception of justice. By asking what reasons free and equal individuals in a (hypothetical) state of nature could have for entering into a political community, contractual theories posit the notion of equal freedom as the foundation from which a conception of justice must be developed. By examining and briefly discussing how different contractualist positions understand this equality-condition and what arguments they offer to explain the transformation of the condition of equality into a fully fledged conception of justice might help us to better understand the systematic presuppositions and aspirations entailed by a conception of justice as equality.

The primary focus of the paper will be the relationship between *motivation* and *justification* within contractualist theories. Section 1 outlines the overall structure of the contractualist approach and argues that such approaches often conceptualize the relationship between motivation and justification as inherently problematic; as a conflict between *self-interested* and *normative* concerns. . To elucidate and elaborate this claim sections 2 and 3 then provide a brief sketch of two very different ways of developing the basic contractualist argument. Section 2 looks at Hobbes, who bases his entire argument for the political commonwealth on human self-interest. Section 3 focuses on Rousseau who rejects appeals to self-interest as possible justifications for the political order and instead provides an argument explicitly based on a normative conception of freedom. Section 4 introduces Rawls' contemporary version of the contractualist approach, and argues that Rawls' account of moral development in part three of *A Theory of Justice* is an attempt to show that the relationship between motivation and justification *need not* and *should not* be conceived as necessarily conflictual. In the conclusion we argue that what our brief overview of the contractualist position reveals is that the condition of equality cannot, indeed should not, be conceived in purely non-normative terms.

## **I The contractualist approach**

In Book 1 of *The Social Contract* Rousseau provides a succinct formulation of not merely the overall aim of his own book, but of social contract theories more generally:

“To find a form of association that will defend and protect the person and goods of each associate with the full common force, and by means of which each, uniting with all, nevertheless obeys only himself and remain as free as before.” This is the fundamental problem to which the social contract provides the solution.

Rousseau 1997b, Book 1, ch. 8

Rousseau's formulation brings out two distinct and important aspects of the contractualist approach. On the one hand this approach has a clear, pragmatic, self-interested focus. A stable political order established through a social contract which define both the duties of the state and the rights of the citizens provide a much more efficient and stable way of protecting the interests of the citizens than anything possible in the state of nature. On the other hand the contractualist approach also has an (implicitly or explicitly) normative aim: to explain how, why and to what degree it is legitimate to restrict the condition of equal freedom, which characterizes the state of nature prior to the establishment of the political community.

These two aspects of the contractualist approach are closely intertwined. Most contractualists thus argue that what justifies and motivates restrictions on the natural state of equal freedom are the pragmatic, self-interested advantages made possible by and through such restrictions. And it is precisely this move by the contractualist approach which transforms a factual or natural *condition* (of equality) into a normative *demand* (of justice). Contractualist theories thus typically measure the legitimacy of a political order by whether and to what extent the restrictions on natural liberty can be justified by appeals to enlightened self-interest. As Locke formulates it:

The only way whereby any one devests himself of his Natural Liberty, and *puts on the bonds of Civil Society*, is by agreeing with other Men to joyn and unite into a Community, for their comfortable, safe, and peacable living one amongst another, in a secure Enjoyment of their Properties, and a greater Security against any that are not of it.

Locke 1994, *Second Treatise*, ch. VIII, § 95

There is however an inherent (although not always explicit) tension between the two aspects of the contractualist approach. Because is it really the case that appeals to self-interest can be used to justify restrictions on liberty? More generally: are appeals to self-interest the right sort of arguments to justify the authority and legitimacy of the political order? It is not all clear that prudential self-interest and considerations concerning liberty and freedom belong within the same normative vocabulary. Arguments based on self-interest seem an obvious starting place for these kinds of discussions because they seem to explain why human beings should be *motivated* to accept certain restrictions on their natural liberty. But is self-interested *motivation* and normative *justification* really the same?

To expand upon this problem consider the way in which contractualist approaches typically define the natural condition of equal freedom, namely as a sort of negative freedom; freedom *from* external control and restrictions. More precisely: natural liberty,

equal freedom in the state of nature, is defined as the absence of external obstacles and constraints which might impede the individual's self-interested pursuit of whatever needs, wants and desires he or she might happen to have. But if freedom is interpreted in terms of freedom from external constraints, including normative constraints imposed by a political community, then it seems reasonable to ask, why human beings should ever be motivated to accept and act in accordance with such principles and norms? Why should anyone be motivated to accept restrictions on their ability to pursue their own, self-interested ends and needs?

This *problem of motivation* is inherent in the very structure of the contractual argument and is a challenge, which every contractual theory of justice must at some point face.<sup>3</sup> And it is this challenge which the contractualist appeal to (enlightened) self-interest is intended to solve. If the contractualist can show that entering into a political community and thus accepting certain restrictions on our natural liberty, is the best and most efficient way to further our self-interested ends and purposes in the long run, then this would seem to explain why people would be motivated to accept such restrictions.

This however does not resolve the basic tension outlined above. The problem is that an *explanation* of why people are *motivated* to do something does not automatically translate into a normative *justification* of that which they are motivated to do. To use a well-known example (Hart 2012, p. 20): The fact that a bank robber threatens a bank teller with a gun might very well explain why the teller is *motivated* to hand over the money in the cash drawer. It even seems to provide a *pragmatic* or *prudential justification* for doing so. But it does not *morally justify*, in the sense of *legitimizing*, the practice of robbing banks.

This tension between self-interested *motivation* and normative *justification* lies at the very heart of the contractualist approach. To illustrate this, and in order to further elucidate the presuppositions and implications of this tension, we now turn to three of the most prominent contract thinkers in modern times: Hobbes, Rousseau and Rawls.

## II Hobbes

Hobbes defines the (hypothetical) state of nature as a state of both equality and liberty. It is a state of equality because “nature hath made men so equall in the faculties of body,

---

<sup>3</sup> As we have noted elsewhere the problem of motivation can be articulated in a number of different ways, each of which raises a distinct set of philosophical challenges. For a discussion of the (relationship between) different articulations of this problem see Nielsen and Hartz, "Why be just? Hegel and Rawls on the problem of motivation". This article is currently in review; please contact the authors for further information.

and mind [...] that the difference between man, and man, is not so considerable” (Hobbes 1985, p. 183). All human beings, according to Hobbes, are more or less alike in their physical and mental constitutions, and no one is so (physically and mentally) preeminent that they are automatically able to rule over others.

Liberty according to Hobbes is “the absence of external Impediments; which Impediments may oft take away part of man’s power to do what hee would” (Hobbes 1985, p. 189). The state of nature contains physical and natural impediments to human action: physical objects might restrict our ability to move as we want and other human beings can oppose our plans and intentions. But, so Hobbes argues, in the state of nature there are no moral, legal or social restrictions on our movements; no normative constraints on our pursuit of our wants or desires (Hobbes 1985, p. 188; p. 202). The state of nature is thus a state of liberty in the sense that it contains no *social* or *normative* restrictions on human action.

Hobbes’ project in the first and second book of *Leviathan* is an attempt to explain why human beings would ever choose to leave the state of nature. Why would anyone willingly give up their freedom to do whatever they want? Why, to use Hobbes’ own terminology, would anyone agree to relinquish

...the Liberty each man hath to use his own power as he will himselfe for the preservation of his own Nature; that is to say, of his own Life; and consequently, of doing anything which, in his own Judgment and Reason, hee shall conceive to be the aptest means thereunto.

Hobbes 1985, p. 189

In the first book of *Leviathan* Hobbes argues that it is *rational* for human beings to leave the lawless state of nature and enter into a political community with other human beings. Hobbes’ argument for this is well known. According to Hobbes what moves human beings to act are simply our self-interested desires and wants. Unfortunately human wants and needs are insatiable whereas the material and social resources necessary to fulfill these needs are limited. This inevitably leads human beings to compete against each other for power, honor and riches (Hobbes 1985, pp. 184-185). And since the state of nature is defined by the complete and utter lack of any normative restrictions on human action this competition necessarily takes the form of “a warre of every man against every man” (Hobbes 1985, p. 188).

Hobbes admits that people might not actually be fighting in the streets in the state of nature. But because of the way he has described this hypothetical situation the state of nature necessarily implies that there is a constant risk that someone will try to take

advantage of you in order to further his or her own ends. In Hobbes' memorable formulation:

In such condition there is no place for Industry, because the fruit thereof is uncertain: and consequently no Culture of the Earth; no Navigation, nor use of the commodities that may be imported by Sea; no commodious Building; no Instruments of moving and removing such things as require much force; no Knowledge of the face of the Earth; no account of Time; no Arts; no Letters; no Society; and which is worst of all, continuall feare, and danger of violent death; and the life of man, solitary, poore, nasty, brutish, and short.

Hobbes 1985, p. 186

The state of nature is thus such a miserable place that everybody has an interest in leaving it in exchange for something better. More precisely: since human beings naturally seek to preserve their own existence, they have very good reason to leave the state of nature where no one can ever be certain that they will not be robbed, raped or killed. For Hobbes it is thus, “a precept, or generall rule of Reason, *That every man ought to endeavour Peace, as far as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of Warre*” (Hobbes 1985 p. 190).

From this “first and fundamentall Law of Nature” Hobbes then derives a number of consequences, including a number of secondary laws of nature. These laws all take the form of rational recommendations outlining the necessary conditions for establishing a political commonwealth, in which everybody give up their natural liberty in order to create a sovereign power invested with the power to protect the life and interests of everybody against everybody else (Hobbes ch. 14-15).

This however, is not the end of the story. Though Hobbes claims that there is a rational argument based on self-interest for leaving the state of nature, he does not believe that this argument in and of itself is sufficient to *actually* motivate people to accept and comply with the “Lawes of Nature” and whatever requirements and restrictions follow from these laws.

For the Lawes of Nature (as *Justice, Equity, Modesty, Mercy*, and (in summe), *doing to others as wee would be done to,*) of themselves, without the terrour of some Power to cause them to be observed, are contrary to our natural Passions, that carry us to Partiality, Pride, Revenge, and the like.

Hobbes 1985, p. 223

The problem is that the laws of nature are rational requirements for how to best ensure the overall fulfillment of our interests in the long run on the condition that everybody else adheres to these laws. As such these requirements will not in every particular instance facilitate the fulfillment of whatever particular wants and desires we might happen to have. This gives rise to the free-rider problem. From the point of view of the individual accepting a *general* system of restrictions on liberty might very well be the most rational course of action on most occasions. But it might *also* be rational to occasionally disobey these restrictions when doing so will promote our own self-interested desires (Hobbes 1985, p. 215).

Unless the restrictions on natural liberty imposed by the commonwealth is backed up by something more than appeals to long term prudential concerns, the commonwealth will thus be inherently unstable. Free-riders will use the legal and political restrictions on natural liberty strategically to further their own interests, thus destabilizing and ultimately undermining the initial agreement (Hobbes 1985, p. 190-191), which forms the basis of the commonwealth.

This is where the sovereign enters Hobbes' story. What is needed to ensure that the members of the commonwealth comply with the basic principles of justice is a credible threat of sanctions for non-compliance; sanctions which are so severe that they overpower our natural passions and desires. "[T]o make their Agreement constant and lasting" Hobbes explains, we need "...a Common Power to keep them in awe, and to direct their actions to the Common Benefit" (Hobbes 1985, p. 227). Such a power can only be established if everybody agrees to confer all their power on one man (or an assembly of men); an individual who "hath the use of so much Power and Strength conferred on him, that by terror thereof he is enabled to forme the wills of them all..." (Hobbes 1985, p. 227).

Hobbes thus discusses the problem of motivation at two distinct but intimately related levels. He first argues that human self-interest provides the basis for a rational argument for leaving the state of nature. Simply put: Hobbes believes that the benefits of entering a political community far exceeds the disadvantages of giving up our natural liberty. In so far as human beings are motivated by prudential concerns (and human beings according to Hobbes *are* motivated by prudential concerns), they should therefore be *rationally motivated* to become members of a commonwealth. Rational motivation however is not, at least according to Hobbes, enough to ensure that people will *actually* comply with the restrictions on natural liberty imposed by the commonwealth. If the commonwealth is to be stable, the basic principles of justice need to be backed up by a credible threat of sanctions for non-compliance.

Hobbes thus supplements his discussion of rational motivation based on considerations of how to best facilitate our overall, long term interests with a discussion of how to motivate people through the use (or credible threat) of force. It is the combination of these two approaches, which transforms the initial non-normative *conditions* of equality and liberty (the state of nature), into an explicitly normative field, the commonwealth, in which moral and legal demands and requirements can be raised and rationally assessed.

Hobbes thus does not distinguish between *motivation* and *justification*. For him arguments based on (enlightened) self-interest are what explains and justifies both the *legitimacy* and the *stability* of the political order. As rational, self-interested creatures human beings can and should be able to see, that it is to their own advantage to enter into a commonwealth, where all power is in the hands of a sovereign, whose decrees are accompanied by a credible threat of sanctions for non-compliance. The restrictions imposed by the commonwealth thus derive their normative force solely from the self-interested concerns of those on which these restrictions are imposed.

Hobbes probably would not see this as a problem. In fact he would probably deny that there is any meaningful distinction to be drawn between self-interested motivation and normative justification, when we are discussing the transition from the state of nature to the political commonwealth. In the state of nature, Hobbes argues, “...nothing can be unjust. The notions of Right and Wrong, Justice and Injustice have here no place” (Hobbes 1985, p. 188). Substantial normative concepts and moral considerations only have meaning within the context provided by a political community, and thus cannot be used to argue for the legitimacy of such a community. In the state of nature only appeals to human self-interest have any kind of rational, normative force. Only appeals to human self-interest can thus justify the political order.

### **III Rousseau**

Rousseau agrees with Hobbes (and, one might add, Locke) that there are pragmatic, self-interested benefits to be gained by entering into a political association with other human beings. Rousseau also agrees that these benefits might be what initially motivate people to accept certain restrictions on the natural condition of equal freedom. In his discussion of the family as “the first model of political societies” in *The Social Contract* Rousseau thus claims that “all being born free and equal, alienate their freedom only for the sake of their utility” (Rousseau 1997b, Book 1, ch. 2).

However, although Rousseau thus officially subscribes to the standard view of the contractualist approach he actually deviates from this approach in significant and interesting ways. To begin with Rousseau is deeply suspicious of the idea that pragmatic, self-interested considerations can justify limitations on individual freedom.

In his *Discourse on the Origin of Inequality* Rousseau openly mocks those who employ such arguments to explain and defend the necessity of political institutions which restrict and limit human freedom. Appeals to enlightened self-interest, according to Rousseau, are nothing but rhetorical devices by which the rich and the powerful intend to convince others to voluntarily give up their freedom (Rousseau 1997a, Part II, p. 173–178). And Rousseau acidly compares those who accept such arguments and think it expedient or beneficial “to sacrifice one part of their freedom to preserve the other” to “a wounded man [who] has his arm cut off to save the rest of his Body” (Rousseau 1997a, Part II, p. 173).

Arguments from self-interest might thus very well *motivate* people to accept restrictions on their natural liberty, but such argument cannot, according to Rousseau, *justify* such restrictions. The only arguments which could possibly justify restrictions on natural freedom are arguments based on and emanating from freedom itself. When Rousseau compares the state of nature with the civil state, the political community established by the social contract, his primary focus is thus not the prudential advantages of this state (which he openly admits exists, see e.g. Rousseau 1997b, book II, ch. 4), but rather the ways in which entering this society changes and contributes to human freedom.

More precisely; Rousseau argues that the civil state should not merely be regarded as a (necessary and prudentially beneficial) limitation on natural freedom, but also and more importantly as an enabling condition for new forms of freedom.

What man loses by the social contract is his natural freedom and an unlimited right to everything that tempts him and he can reach; what he gains is civil freedom and property in everything he possesses. [...] To the preceding one might add to the benefit of the civil state moral freedom, which alone makes man truly master of himself, for the impulsion of mere appetite is slavery, and obedience to the law one has proscribed for one self is freedom.

Rousseau 1997b, Book I, ch. 8

For Rousseau the social contract is not simply a means to advance the long term interests of self-interested individuals in a (hypothetical) state of nature. The contract also represents a crucial step in the moral development of human beings. It is not merely the way human beings relate to each other which changes with the introduction of the social contract. Human beings *themselves* are fundamentally and radically changed, when they enter into the civil state.

This transition from the state of nature to the civil state produces a most remarkable change in man by substituting justice for instinct in his conduct,

and endowing his actions with the morality they previously lacked. Only then, when the voice of duty succeeds physical impulsion and right succeeds appetite, does man, who until then had looked only to himself, see himself forced to act on other principles [...] his faculties are exercised and developed, his ideas are enlarged, his sentiments ennobled, his entire soul is elevated [...] and out of a stupid and bounded animal [is] made an intelligent being and a man.

Rousseau 1997b, Book 1, ch. 8

Rousseau rejects arguments based on self-interest as justifications of civil society for at least two reasons. First of all letting yourself be determined by brute, uncultivated self-interest, by whatever physical appetites and desires you happen to have, is, according to Rousseau, simply another form of slavery. If your actions are governed by desires and wants over which you have little or no control, then you are not truly free; you are still being determined by forces outside your own control. Justifying civil society through appeals to self-interest thus cannot in any meaningful sense be claimed to promote or secure human freedom. And arguments from freedom are the only arguments which can legitimately be used to restrict or curb freedom.

Secondly, entering the civil state so radically and fundamentally changes human beings that self-interest and natural desires, the kinds of motivation which define the state of nature, are no longer the most important and basic forms of human motivation. To justify the civil state by appeals to self-interest is to fundamentally misunderstand the nature of the political community, whose primary purpose, according to Rousseau, is precisely to *replace* self-interest as the primary source of motivation.

Rousseau *does* believe that there are good prudential reasons for entering into civil society, reasons which might very well *motivate* people to agree to the social contract. But these reasons are for Rousseau not what *justify* the restrictions on natural liberty imposed by civil society. What justify these restrictions are *moral* reasons, reasons grounded in freedom. More precisely restrictions on natural freedom (the “unlimited right to everything” that tempts us) are justified only because and insofar as they are necessary to ensure and enable other, more complex and adequate conceptions of freedom: civil freedom (legally enforced protection of my life and my property against others) and moral freedom (freedom from the tyranny of uncultivated desires and wants). This is what motivates Rousseau’s infamous remarks that in the civil state “whoever refuses to obey the general will shall be constrained to do so by the whole body; which means nothing other than that he shall be forced to be free” (Rousseau 1997b, book I, ch. 8).

What *justify* the civil state are thus arguments based on *freedom*. What defines the *form* of the civil state however is an argument based on *equality*. The basis of the social contract, Rousseau argues, is “the total alienation of each associate with all his rights to the whole community [...And] since each gives himself entirely, the condition is equal for all” (Rousseau 1997b, book 1, ch. 7). What the social contract requires is that *everybody* agrees to hand over *all* their rights to *everybody else*, thus making absolute normative equality the basis of the civil state. Only if equal (civil and moral) freedom is the foundation of the political order, so Rousseau argues, can tyranny and despotism can be avoided.

Rousseau thus views the civil state as an explicitly *normative* field, which can only be justified through appeals to explicitly normative arguments based on conceptions of freedom and equality, which are themselves normatively loaded. More precisely: the conception of equal freedom, which structure the civil state, is so different from the natural conditions of liberty and equality in the state of nature, that the former cannot be derived from the latter simply through appeals to (enlightened) self-interest.

This means that Rousseau conceives of the relation between justification and motivation in a radically different way than e.g. Hobbes. For Rousseau the question of how to justify the political community (the civil state) calls for a *moral* or *normative* justification, which is not based on self-interest. For Rousseau questions concerning the normative *justification* of the political community are thus clearly distinct from questions concerning the self-interested *motivations* individuals might have for entering such a community. Whatever motivational force such substantial normative justifications might have (a question which Rousseau admittedly does not explicitly discuss in his political philosophy) thus cannot be derived from human self-interest in any simple and straightforward way.

#### **IV Rawls**

Rawls' contractual argument is an explicit attempt to "present a conception of justice which generalizes and carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau, and Kant" (Rawls 1999, §3). Just like these classic contractual theorists, Rawls defines a condition of equality as the starting-point of his contractual argument. In Rawls' argument this condition of equality is the original position, a “position of equality [that] corresponds to the state of nature in the traditional theory of the social contract” (Rawls 1999, § 3).

The original position, Rawls explains, is “a purely hypothetical condition characterized so as to lead to a certain conception of justice” (Rawls 1999, § 3). It is *not* a condition in which individuals can actually be thought to live, but rather a specific deliberative point

of view. The participants in the original position deliberate behind a "veil of ignorance", which ensures "that no one knows his place in society, his class position or social status, nor does any one know his fortune in the distribution of natural assets and abilities, his intelligence, strength and the like" (Rawls 1999, § 3).

What the veil of ignorance is intended to achieve is to enable us to deliberate from a position which "leave aside those aspects of the social world, which seems arbitrary from a moral point of view" (Rawls 1999, § 3). And the point of formulating such a deliberative position or point of view, according to Rawls, "is to represent equality between human beings as moral persons, as creatures having a conception of their good and capable of a sense of justice" (Rawls 1999, § 4).

At a first glance therefore, Rawls' original position seems to constitute a much more abstract starting point than the state of nature from which the classical contractual theories depart. But a closer look reveals that in one important respect the original position is much more closely tied to our concrete experience as members of a political community than e.g. Hobbes' conception of the state of nature. Because what the original position is supposed to do is *not* to provide us with a pre-social or non-normative perspective, from which to deliberate about the basic principles of justice. The explicit purpose of the original position is rather to explicate a set of conditions, which at one and the same time reflect and enable us to critically evaluate our deepest and most fundamental moral beliefs; beliefs which, at least in part, is the result of living in a particular political community (Rawls 1999, § 4).

This point mirrors Rousseau's approach in so far that Rawls, like Rousseau, insists that the justification of the political order must incorporate and take as its starting point those substantial normative conceptions of freedom and equality, which is made possible through and define us as members of a political community. However, while this point leads Rousseau to explicitly reject the normative force of the classical contractualist argument from self-interest, Rawls holds on to the intuition that the principles of justice "...are the principles that free and rational persons *concerned to further their own interest* would accept in an initial position of equality" (Rawls 1999, § 3, emphasis added). On at least some level Rawls thus seems to accept the very thought that Rousseau explicitly denies: that appeals to self-interest can be used to justify the basic principles of justice.

At first glance Rawls might here appear similar to Hobbes. Both Rawls and Hobbes take as their starting-point the prudential reasoning of self-interested individuals in an initial condition of equal freedom. Now, Rawls is quite explicit that he in no way conceives of the participants in the original position as self-interested Hobbesian individuals whose

primary intention is to further their own ends. In fact the way Rawls interpret the original position is explicitly aimed to either avoid or overcome the strong individualistic form of self-interest, which defines and structures the Hobbesian version of the social contract. Participants behind the veil of ignorance, Rawls explains, are not egoists. They are merely “mutually disinterested”. And when mutual disinterest is combined with the veil of ignorance the result is something very much like benevolence: “For this combination of conditions forces each person in the original position to take the good of others into account” (Rawls 1999, § 25). Furthermore, by rendering the original position completely abstract Rawls aims to impose a procedural constraint on his argument, which ensures that the choices made behind the veil of ignorance are those that would be made if we adopted a universal perspective, *sub specie aeternitatis* (Rawls 1999, § 87).

However the very fact that that the veil of ignorance and the other procedural limitations imposed by Rawls on the original position are specifically designed to avoid narrowly self-interested forms of deliberation, shows that prudential reasoning *is* the driving force behind Rawls argument. Otherwise there would be no need to explicitly tailor the original position the way he does.

One important consequence of these procedural restrictions is that although Rawls, like Hobbes, includes a motivational presumption of self-interest (albeit in a very weak form) in his definition of the initial condition of equality, Rawls *unlike* Hobbes does not have to solve the problem of motivation. The question of why people in the original position are motivated to accept the restrictions on liberty, imposed by whatever principles of justice chosen in the original position, simply does not arise. Everyone in the original position, so Rawls argues, is necessarily motivated to choose the principles that would be best for all, because that is how the original position is defined (Rawls 1999, § 40).

However, while the problem of motivation is not of immediate importance in the original position, Rawls explicitly acknowledges that this problem *is* important at the level of practical applicability. Because if the basic principles of justice chosen in the original position are inherently at odds with the motivations and interest of ordinary people, then a society based on those principles will be fundamentally unstable. And this fact, Rawls argues, is something, which should be taken into consideration in the original position:

It is evident that stability is a desirable feature of moral conceptions. Other things equal, the persons in the original position will adopt the more stable scheme of principles. However attractive a conception of justice might be on

other grounds, it 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 1999, § 69

Although the problem of motivation at first glance appears to be completely absent (defined out of existence one might say) from the original position, it thus necessarily re-emerges in Rawls' theory at a later stage. Because even if all the participants in the original position are necessarily motivated to accept and adopt a certain set of principles as the mutually binding basis for fair social cooperation, the same might not be true of the members of any concrete political community. "The motivation of the persons in the original position" Rawls notes "must not be confused with the motivation of persons in everyday life [...] In practical affairs an individual does have a knowledge of his situation and he can, if he wishes, exploit contingencies to his advantage." (Rawls 1999, § 25). If the members of a political community have little or no "sense of justice", i.e. if they are not motivated to endorse the basic principles of justice which define the basic structure of their community, then that community will be inherently unstable (Rawls 1999, § 76). And considerations concerning social stability *are* relevant for the deliberations among the participants in the original position.

This means that Rawls' contractualist argument for his favored set of basic principles of justice (the principle of equal freedom and the maximin-principle, see Rawls 1999 § 11 & 46), cannot be carried out purely in terms of what goes on in the hypothetical original position. Rawls also needs to explain that and how members of a particular political community can come to be motivated by these principles. If he cannot provide a convincing account of this, then the argument from the original position fails, because rational and self-interested individuals cannot and will not agree on a set of principles, which renders the political community inherently unstable. Somewhat surprisingly Rawls' response to the problem of motivation thus turns out to be the crucial glue needed to keep his contractual argument together.

As we have seen Hobbes' solution to the problem of stability (which is simply a version of the more general problem of motivation) is the sovereign. As long and in so far as the sovereign represent a credible threat of sanctions for non-compliance, fear of punishment will motivate even the potential free-rider to adhere to the basic principles of justice. However, this solution to the problem of motivation is not acceptable to Rawls (at least not in a strict Hobbesian version) for a number of reasons. For our purposes the most important of these reasons is that Rawls, as we have seen, explicitly designed the original position as a deliberative framework which incorporates and

allows us to critically reflect upon our most basic and fundamental moral beliefs. And as members of a democratic and well-ordered society our most basic and fundamental understanding of morality and justice is very difficult, probably impossible, to reconcile with the notion of an absolute Hobbesian sovereign. (For Rawls' own discussion of Hobbes' solution see Rawls 1999, § 38 & 76).

Rawls thus needs another solution to the problem of motivation, and as it turns out most of the third part of *A Theory of Justice*, almost 200 pages, is dedicated to precisely this problem. What Rawls here proposes is something quite different from an explanation of how purely self-interested individuals can be motivated to accept the rules of a political community. Instead Rawls argues that members of particular political communities are in fact *never* purely self-interested in the Hobbesian sense. On the contrary, one of the effects of living in a political community, or so Rawls argues, is that by partaking in such a community we gradually come to identify with the basic principles of justice which defines it.

In the third part of *A Theory of Justice* Rawls therefore tackles the problem of motivation through a developmental account of how human beings acquire a sense of justice; "...a normally effective desire to apply and to act upon the principles of justice, at least to a certain minimum degree" (Rawls 1999, § 77). The purpose of this account is to show, that it is possible, perhaps even probable, that an individual simply by living in a well-ordered society will gradually become motivated to accept, endorse and act in accordance with the basic principles of justice as articulated in the original position.

According to Rawls (and very much in the spirit of Rousseau) the political community thus cannot be adequately understood as simply a means to advance the long-term interests of self-interested individuals. Participation in the many different associations and partnerships that make up the social and moral world of a political community expands and re-orientates the motivational constitution of human beings in fundamental ways. In other words: Rawls believe that the political community plays a vital and integral role in the moral development of human beings (Rawls 1999, § 79).<sup>4</sup>

On the one hand Rawls thus seems to accept the basic Hobbesian thought that prudential deliberations among self-interested individuals in an initial state of equal freedom can *justify* a particular set of basic principles of justice. Rawls however, as opposed to Hobbes, defines this initial condition of equality in such a way that the question of why the participants in the original position are *motivated* to accept the restrictions imposed by these basic principles simply does not arise. However the problem of motivation

---

<sup>4</sup> See Nielsen and Hartz (in review) for a more detailed account of Rawls' discussion of moral development in the third part of *A Theory of Justice*.

emerges at another level of Rawls' theory. Because to ensure that a society based on the basic principles chosen in the original position is stable Rawls needs to show that and how human beings in a non-hypothetical, concrete social world can actually come to be motivated by these principles.

On the other hand Rawls thus also seems to agree with Rousseau that the political order cannot and should not be viewed simply as a social system which enables self-interested individuals to better pursue and fulfill their needs and desires. Membership of a political community and a well-ordered society fundamentally alters the motivations and desires of human beings, in ways which are both relevant and necessary for the justification of the basic principles of justice. And this raises the question of whether and to what extent the original position is or can be thought of as independent of, distinct from and normatively prior to the political community whose basic principles it is supposed to justify.

## Conclusion

The question we set out to explore in this paper was the question of how and why a certain *condition*, a condition of equality, can be (or be transformed into) a *demand* of justice. More precisely we have tried to explicate a) how Hobbes, Rousseau and Rawls *justify* the legitimacy of a political community which imposes restrictions on natural liberty, and b) how these thinkers conceive of the relationship between questions of *justification* and questions of *motivation*.

As we have seen Hobbes argues that a stable and legitimate political order can be justified solely through arguments based on enlightened self-interest. For Hobbes appeals to self-interest both *justify* and *motivate* why human beings move from the state of nature to the political commonwealth and explain the conditions under which human beings are willing to accept the restrictions on natural liberty imposed by this commonwealth.

Rousseau on the other hand acknowledged there are good prudential reasons for self-interested individuals to establish a political community; reasons which might very well explain why people are *motivated* to do so. But he seems to deny that such self-interested reasons can or should be used to legitimize the political order. According to Rousseau the transition from the state of nature to the civil state produces such a radical moral change in human nature that appeals to pre-social or pre-political notions of self-interest are rendered irrelevant as possible justifications for the legitimacy of the political order. What justifies the civil state in Rousseau's view are substantial

normative considerations concerned with the new and complex forms of human freedom, which are made possible by and through the establishment of such a political order.

Rawls' contractualist position appears to hover, perhaps somewhat uneasily, between Hobbes and Rousseau. While Rawls agrees with Hobbes that prudential, self-interested reasoning in an initial state of equal freedom can justify a set of basic principles of justice, he also seems to agree with Rousseau that membership of a political society changes human nature in fundamental ways, which puts into question the validity and relevance of justificatory arguments based on or derived from self-interest.

What all of this suggests is that the different equality conditions, which define the starting-points of these three theories, cannot and should not be viewed as a non-normative foundation or normatively neutral premise in the contractualist argument. On the contrary the way the initial condition of equality is explicated at least partly determines how the transition from the conditions of equality to the demands of justice takes place. And conversely; the normative content of the conception of justice, which is supposed to be the conclusion of the contractualist argument, seems to influence, perhaps even define, how the initial equality condition is conceptualized.

In Hobbes' theory for instance the very definition of the state of nature, a condition free from all normative restrictions and constraints, naturally entails a very particular conception of freedom: freedom from external normative constraints. And this conception of freedom, combined with the Hobbesian view of human beings as rational, self-interested individuals, sets the stage for the way Hobbes explicates and attempts to justify the transition from the state of nature to the political commonwealth. For Hobbes this justification takes the form of two questions concerning human motivation. First: why would human beings ever be motivated to enter a political commonwealth and thus accept restrictions on their natural liberty? And second: how must the political commonwealth be structured in order to ensure that each member of this commonwealth is actually motivated to follow the basic principles of justice?

Rousseau on the other hand moves in the other direction. Because of his strong emphasis on the inherent normativity of the civil state and the way in which this state enables new and more complex forms of freedom Rousseau explicitly denies that the self-interested arguments of free and equal persons in an initial state of nature can ever justify the political order. Such arguments might very well be what motivate people to enter a political community, but only normatively loaded conceptions of freedom and justice can justify the legitimacy of such a community.

As for Rawls the procedural restraints which define the original position are not normatively neutral. Instead they are designed so as to impose a particular normative structure on the deliberations and choices of the participants in the original position; a structure which fundamentally influences which conceptions of justice these participants can rationally commit to. Whereas Hobbes' state of nature is purposefully cleansed of normative content Rawls intends the original position to reflect and incorporate certain normative considerations and constraints.

More precisely: Rawls explicitly structures the original position so that it reflects a particular normative ideal of equal freedom; an ideal which Rawls believe we already, implicitly, accept and endorse. At the very end of *A Theory of Justice* Rawls thus asks “why should we take any interest in it [the original position], moral or otherwise?” The answer, Rawls believes, is that “the conditions embodied in the description of this situation are ones that we do in fact accept. Or if we do not, then we can be persuaded to do so by philosophical considerations of the sort occasionally introduced [in the book]” (Rawls 1999, § 87). And the reason we accept and endorse these conditions Rawls argues is because they reflect and respect our status as free and equal rational beings. “...when we knowingly act on the principles of justice in the ordinary course of events,” Rawls thus explains, “we deliberately assume the limitations of the original position. One reason for doing this, for persons who can do so and want to, is to express their nature as free and equal rational beings” (Rawls 1999, § 40).

This might be yet another reason why Rawls spends the last 200 pages of *A Theory of Justice* on explaining how individuals in a well-ordered society gradually come to internalize the values and norms of that society. Because if the very structure of the original position depends on values and norms, which we already implicitly accept, then it becomes crucially important to explain how and why we have come to these values. This would seem to not only further distinguish Rawls from Hobbes. It would also bring Rawls much closer to Rousseau - closer perhaps than Rawls himself might like to admit.<sup>5</sup> Because if it is only as members of a political community that we are able to develop a moral perspective, and if it is only as moral beings equipped with such a perspective that we are qualified to deliberate about the principles of justice in the original position, then it becomes quite unclear exactly what the justificatory work of the original position consists in. Which, of course, was precisely Rousseau's point when he criticized the idea that appeals to prudential self-interest could ever be used to justify the political order.

---

<sup>5</sup> Or, as we have argued elsewhere, to Hegel. See Hartz & Nielsen (in review).

Although Rawls explicitly claims that the original position is supposed to bring to a higher level of abstraction the equality condition (the state of nature) of the social contract tradition, it would seem that Rawls has in fact altered this condition in ways that sets his theory apart from e.g. Hobbes' in important ways. To once again return to the problem of stability: despite certain similarities this problem actually presents itself quite differently in Hobbes' and Rawls' theories. For Hobbes the problem is to show how the self-interested desires of basically non-moral individuals can be kept in check by fear of the sovereign. The challenge for Rawls on the other hand is to show that the basic principles of justice can in fact become a determining motivating factor in the lives of ordinary persons living in a well-ordered society.

In conclusion: It seems reasonable to be skeptical about whether the equality condition can ever be articulated in morally neutral and strictly non-normative terms. A more promising way to develop a viable and interesting conception of justice as equality would be to explicate the substantial normative structures inherent in our existing political communities, and show how and to what extent these structures support and enable particular normative conceptions of equality and freedom. Rousseau (and to some extent Rawls) seems to have been well aware of this point. This is one reason why their theories continue to be of interest today.

## **Bibliography**

Hart, H. L. A. (2012): *The Concept of Law*, Third Edition, Oxford UP, Oxford, UK.

Hartz, Emily & Nielsen, Carsten (in review): "Why be just? Hegel and Rawls on the problem of motivation"

Hobbes, Thomas (1985): *Leviathan*, ed. C. B. MacPherson, Penguin Books, London, UK.

Locke, John (1994): *Two Treatises of Government*, ed. Peter Laslett, Cambridge UP, Cambridge, UK.

Rawls, John (1999): *A Theory of Justice, Revised Edition*, Oxford UP, Oxford, UK.

Rousseau, Jean-Jacques (1997a): *Discourse on the Origin of Inequality*, pp. 111-188 in *The Discourses and other early political writings*, ed. and tr. Victor Gourevitch, Cambridge University Press, Cambridge, UK.

Rousseau, Jean-Jacques (1997b): *The Social Contract*, pp. 39-152 in *The Social Contract and other later political writings*, ed. and tr. Victor Gourevitch, Cambridge University Press, Cambridge, UK.