

The Priority of Liberty: Political not Kantian

Greg Whitfield[†]

May 2009

In A Theory of Justice John Rawls orders his principles of justice in a famously strong manner: equal basic liberties are *lexically prior* to the considerations of fairness treated by the second principle, and these considerations in turn are organized so that fair equality of opportunity precedes concerns about socioeconomic inequality and justice precedes efficiency. Thus each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberty for others, and only then may parties in the original position move on to address concerns about the availability of offices and responsibilities to those with similar talents and motivations, the fair worth of liberties, and the degree to which social and economic inequalities are arranged to everyone's advantage.

These are demanding conditions imposed on the basic structure of a society. Indeed, the priority which Rawls gives to liberty has been a major concern among diverse critics, who might agree on little else than the unjustified stringency of Rawls's first priority rule. The criticism is important not simply because of the diversity and persistence of its champions, but also because of the clear centrality of the priority of liberty to justice as fairness: “the force of justice as fairness would appear to arise from two things: the requirement that all inequalities be justified to the least advantaged, and the priority of liberty” (*TJ* 220). Rawls thus viewed the priority of liberty as an essential and distinctive element of justice as fairness, and it seems that critics have agreed with him on this centrality.

In a recent contribution to this line of criticism, Robert Taylor has argued that while Rawls's arguments support a *high* priority for liberty, this priority does not amount to *lexical* priority.¹ To sustain the stronger ordering, Rawls needs, and is best read as adopting, a commitment to Kantian autonomy stronger than has generally been supposed. According to Taylor's argument, then, the Kantian interpretation of justice as fairness is not simply a more charitable alternative interpretation, it is the correct one. But this interpretation raises at least one very serious issue for Rawlsians. The turn to the overlapping consensus approach in Political Liberalism is in tension with the Kantian approach to the priority of liberty, since we cannot reasonably expect that a value as comprehensive and contentious as autonomy will be the subject of wide agreement in a society under conditions of reasonable pluralism.

But we are saved from the problems posed by Kantian autonomy, as Taylor's argument gives short shrift to the strains of commitment and concerns about stability, both critical elements of Rawls's view. Indeed, I will show that the other arguments Rawls offers for the priority of liberty, though important, are secondary to the work being done by the argument from

[†] Department of Political Studies, Queen's University, Canada. Email: greg.whitfield@queensu.ca

¹ Rawls uses 'lexical priority' to refer to a method for ordering principles in a way akin to alphabetic ordering: an ordering “which requires us to satisfy the first principle in the ordering before we can move on to the second, and the second before we consider the third, and so on. A principle does not come into play until those previous to it are either fully met or do not apply. A serial ordering avoids, then, having to balance principles at all; those earlier in the ordering have absolute weight, so to speak, with respect to later ones, and hold without exception” (*TJ* 38).

the strains of commitment and the concern for stability in his account of justice. This, I hope to show, should obviate the need for the appeal to Kantian autonomy which Taylor thinks must be doing the heavy lifting in Rawls's argument for the first principle and its lexical priority over the second. Rejecting the Kantian interpretation in this way is an attractive reading of Rawls, since the Kantian privileging of autonomy is a reasonably contestable premise which will be unacceptable or simply not endorsed from the perspective of many reasonable comprehensive views. That is, given the burdens of judgment and the fact of reasonable pluralism, the Kantian interpretation gives up on the hope that diverse reasonable comprehensive conceptions of the good might be brought into an overlapping consensus about justice. Now whether or not this move to overlapping consensus that Rawls makes between A Theory of Justice and Political Liberalism is desirable or convincing is hotly debated, but for those of us who worry that appeals to the good of autonomy will be (at least) problematic in societies characterized by reasonable pluralism, avoiding the necessity of turning to the Kantian interpretation rescues much of what we find distinctive and appealing in the Rawlsian edifice.

In his paper, Taylor seeks to show that the lexical priority of the first principle over the second is such a demanding condition that it requires a special form of justification. He suggests that Rawls courts a fallacy of inference, failing to recognize the distinctiveness of lexical priority when he argues as though the “lexical priority of the basic liberties can be inferred from the *high* priority of the interests they serve” (Taylor 2003, 248). On Taylor's account, Rawls offers three separate arguments for the lexical priority of liberty: “the Self-Respect Argument,” the “Equal Liberty of Conscience Argument,” and the “Hierarchy Argument” (Taylor 2003, 248). He claims that both the self-respect argument and the equal liberty of conscience argument exhibit the ‘inference fallacy’, and that only a more strongly Kantian ‘hierarchy of interests’ argument can salvage Rawls's ordering scheme for the principles of justice.

I want to challenge both Taylor's arguments concerning the inference fallacy, and more generally, his characterization of the arguments Rawls makes for the lexical priority of liberty. I will first consider Taylor's criticisms of the so-called ‘equal liberty of conscience’ and ‘self-respect’ arguments, showing that his Kantian approach seems to miss much of what is truly distinctive about Rawls's position. I will then challenge the implied separateness of the arguments for the priority of liberty, showing that both arguments originate from a shared premise which grounds the core argument for the lexical priority of the basic liberties in the strains of commitment and the stability of a well-ordered society. I think this repositioning of the priority of liberty and the centrality of stability in justice as fairness not only obviates the need for the Kantian interpretation that Taylor and others favour (cf. O'Neill 2003a; 2003b), but also shows a deep underlying unity between his works that other commentators (e.g. Wenar, Freeman) have affirmed.

The Equal Liberty of Conscience Argument

Taylor quotes the following passage from Rawls's discussion of the importance of the interest people have in their religious commitments and how the equal liberty of conscience protects that interest:

“[the parties in the original position] regard themselves as having moral or religious obligations which they must keep themselves free to honour... They cannot take chances with their liberty by permitting the dominant religious or moral doctrine to persecute or to suppress others if it wishes... To gamble in this

way would show that one did not take one's religious or moral convictions seriously" (*TJ* 180-181).

Taylor claims that this may be interpreted as an instance of the inference fallacy. He suggests that "one might "highly value [an interest in subscribing to particular moral and religious beliefs] yet still endorse small sacrifices of equal liberty of conscience if such sacrifices were necessary to advance *other* highly valued interests" (Taylor 2003, 252).

But the statement about "other highly valued interests" leaves open which specific interest(s) might count highly enough for the parties to endorse sacrifices of liberty. On Rawls's account, given the priority of fair equality over the difference principle and justice over efficiency, interests in efficiency and welfare are not acceptable candidates for trading in the sense Taylor means. Indeed, as Rawls's critics have noted, this is an enormously restrictive account of liberty, since even quite small restrictions of liberty which result in very large benefits to the least well off will be prohibited. On the other hand, Rawls clearly admits that particular liberties may be made less extensive so long as the less extensive liberty strengthens the total system of liberties shared by all, and those with the lesser liberty find the trade-off acceptable (*TJ* 266). So in (at least) one sense, Rawls is not against the sort of trade-offs Taylor seems to be considering. Here justice as fairness seems to permit the weighing and trading of different interests within the context of equal liberty. But this type of trade-off – one consistent with the priority rules – does not constitute a counterexample to Rawls's arguments for the lexical priority of liberty, since balancing liberties against one another still gives priority to the scheme of liberties as a whole.

There is another interpretation of the religious obligations passage from Rawls which Taylor considers. This is the only point in his argument where Taylor seems directly to consider the strains of commitment and the priority of liberty, suggesting that we might best interpret the equal liberty of conscience argument as an argument from the strains of commitment. Noting that this is an especially powerful interpretation when applied to the religious case, he writes:

"If the parties in the original position agree to anything less exacting than lexical priority of equal liberty of conscience, they may emerge from behind the veil of ignorance to discover that their own religious beliefs and practices have been put in jeopardy by discriminatory legislation and that they are psychologically incapable of abiding by such legislation due to an overriding fear of supernatural punishment" (Taylor 2003, 252).

Taylor suggests that the strains-of-commitment argument is not merely "especially strong" when applied to the religion example; he suggests that the strains-of-commitment argument *only* provides support for the lexical priority of liberty when applied to the religion example – that it does not extend to philosophical and moral commitments, as Rawls claims. Taylor's reasons for this are puzzling, as they do not seem to take seriously the demands on representatives seeking to formulate principles of justice behind the veil of ignorance.

Part of the reason Taylor offers for the claim that the strains of commitment argument only works for certain types of basic liberties, namely religious freedom and integrity of the person, is that these are common and deeply important elements of liberty of conscience, while moral and philosophical liberty in relation to freedom of speech are not so common or so important. Specifically, he claims that the historical rarity of people incapable or unwilling to abide by laws that discriminated against their philosophical and moral (as opposed to religious) beliefs means that the strains of commitment argument is not in force for non-religious philosophical and moral beliefs. He offers Socrates and Galileo as examples of reasonable

persons “incapable or unwilling to abide by laws which discriminate against their philosophical and moral beliefs” (Taylor 2003, 253), suggesting that the rarity of such examples is an argument for not ensuring their equal liberty, rather than a counterexample to his point that philosophical and moral beliefs are less important to equal liberty of conscience than are religious beliefs. Thus it seems that Taylor’s claim against the strains-of-commitment argument turns on whether we accept his assertion that the rarity of actual examples should determine (at least in part) the convincingness of a strains-of-commitment argument.

Derogating moral and philosophic commitments in this way exhibits a two-fold problem of considering non-ideal circumstances in an argument against ideal principles. First, we must ask whether capability and willingness to abide by laws that discriminate against particular beliefs should reasonably constitute a test for the strength of a strains of commitment argument. But this is a complicated test. There are deep psychological differences between a person in a society who considers whether she can endure the particular rights regime under which she lives, and a person behind the veil of ignorance considering the same thing. Status quo bias, high discounting rates and weak introspection all contribute to a greater probability that a person will accept a lesser system of equal rights while out from behind the veil of ignorance, than will that same person if she is considering the predicament from the original position. So the test is not entirely inappropriate, and we must keep firmly in mind the stipulation that parties in the original position look to endorse principles which they can willingly live under, regardless of how their values might shift over a complete life.

Second, when Taylor appeals to the actual presence and distribution across time of citizens who treat philosophic and moral beliefs as many others treat religious commitments, he makes a move that is unacceptable from the standpoint of Rawlsian justice, and not a move accessible to citizens in the original position. The purpose of the veil of ignorance is to render moot morally irrelevant considerations when deciding on principles to regulate the basic structure of society. Thus we must ask whether the actual historical presence – and indeed, presence in sufficient numbers, on Taylor’s account – of people who would find the strains of commitment pushed beyond tolerance in a society in which they lack the liberty to fully value their philosophical and moral commitments, is a morally relevant consideration.

I think we must answer no: that the presence or proportions of people who hold such reasonable beliefs are not morally relevant considerations when formulating principles of justice. Indeed, rendering moot such morally irrelevant considerations is one of the main features of the original position argument. The bare fact that such persons *can* exist and reasonably hold their beliefs commits parties in the original position to ensure the equal liberties of such people, compatible with like liberties for all. The original position is intended to model Rawls’s idea of society as a fair system of social cooperation, such that anything a co-operator could not in good faith commit herself to over a complete life cannot be considered a fair demand for society to make on her. That is, the distinct possibility of coming out from behind the veil of ignorance and finding oneself with the values of Socrates or Galileo precludes agreement to principles which Socrates or Galileo could not abide by. Such an agreement, if made, could not be viewed as a good faith agreement, since the parties recognize that certain conceptions of the good rule out good faith agreement to principles they would be in deep tension with.

So the historical presence or absence of a particular sort of valuer is not a relevant test for how compelling a strains-of-commitment argument is. Nor is the particular or general importance assigned to a given value. Rather, the relevant test is whether a reasonable person might hold such beliefs, and whether they might reasonably hold them to such a degree that

curtailing their freedom to act and think on these values would violate the strains of commitment. If we admit that Socrates is an example of someone who reasonably values truth and justice to the degree that he would rather die (or leave society) than live in a world where he was not free to form, hold and value these beliefs, then we admit that moral and philosophical beliefs are of the same importance as religious beliefs, and must be afforded the protections of equal liberty of conscience.

Thus Taylor is mistaken to insist that equal liberty of conscience is merely something we highly value, and thus may be traded off “to advance other highly valued interests” (Taylor 2003, 252). Equal liberty of conscience in fact must be given strict lexical priority, and thus gives strong support to the lexical priority of the basic liberties generally.

Here it may be objected that while lexical priority seems correct when applied to the core of a particular right (eg. the right to worship proper) in the case of more marginal elements of the right (eg. right to display religious symbols for those occupying roles in the state) our interest is weaker, and insufficient for establishing a strains of commitment claim in the favour of lexical priority. But this seems to me entirely consistent with the stipulation in justice as fairness that liberties may be made less extensive so long as the less extensive liberty strengthens the total system of liberties shared by all (*TJ* 266). This is almost certainly the case with religious displays by public officials, and I believe is similarly the case with any other example of a right or prohibition marginal to the core of the liberty at issue.

However, we may still be left with the troubling (for lexical priority) idea that the more marginal a right is to the core liberty it supports the less reason we have to insist on its lexical priority over other important interests, like equality and efficiency. Even restrictions on speech or the exercise of religion may be organized along a spectrum moving away from the core liberty of conscience. At some point along such a spectrum, we may want to say, with Taylor, that efficiency or equality trump. Against this worry, I suggest that there are two factors weighing against its force. First, once again it will be the case that many (if not all) such tradeoffs will be justifiable through appeal to other liberties. Restrictions on certain forms of speech and religious expression are likely to be justified vis a vis self-respect or other basic liberties, rather than the forbidden appeals to fair equality or efficiency. And second, the line across which we would begin to think that restrictions to rights would be justified via appeal to interests other than liberty is truly distant from the core liberty we seek to protect. The enjoyment of the basic liberties could be reliably secured through lexical priority while still acknowledging the sensibility of limits around the margins. This is because we can distinguish between the core aspects of a liberty and aspects sufficiently marginal to its full exercise as to not be straightforwardly necessary – hence admitting that while the former demands lexical priority, the latter is not essential in supporting our higher-order interests.

The Self-Respect Argument

Self-respect for Rawls is the psychological attitude or disposition that “includes a person’s sense of his own value, his secure conviction that his conception of his good, his plan of life, is worth carrying out,” as well as his confidence in his ability to succeed in carrying out that plan of life (*TJ* 386). Absent confidence in either the ability to arrive at a plan of life worth carrying out, or the ability to actually carry out a plan of life, “nothing will seem worth doing, or if some things have value for us, we will lack the will to strive for them” (*Ibid.*). It is plain to see why Rawls at several places refers to self-respect’s social bases—the institutional arrangements which foster self-respect in citizens—as perhaps the most important primary good, since without it the development of the two moral powers will seem an empty and baseless endeavour and the

other primary goods will serve no purpose at all. Thus for Rawls, that a society or set of institutions better supports self-respect is a strong or compelling reason for the parties to the original position to accept that society over others (Ibid.).

Rawls writes that “in a well-ordered society...self-respect is secured by the public affirmation of the status of equal citizenship for all,” and the best way to achieve equal citizenship is to “support the primary good of self-respect as far as possible by the assignment of the basic liberties that can indeed be made equal, defining the same status for all” (TJ 478). So as it relates to self-respect, Rawls’s justification of the priority of liberty is as follows:

When it is the position of equal citizenship that answers to the need for status, the precedence of the equal liberties becomes all the more necessary. Having chosen a conception of justice that seeks to eliminate the significance of relative economic and social advantages as supports for men’s self-confidence, it is essential that the priority of liberty be firmly maintained (TJ 478).

The intuition here is that for people in a poorly-ordered society, status is an important source of self-respect, often manifesting as a competition for material means, with the wealthy or those possessing high positions attaining status at the expense of others, since it then amounts to a scarce resource. By ensuring that status is instead linked to equal citizenship, Rawls seeks to change status and self-respect from a zero-sum game into one where it is possible for all to have a sufficient level of self-respect. Since equal citizenship – cashed out in terms of equal political rights and freedoms – answers this need for status and self-respect, any tradeoffs against the liberties of equal citizenship will necessarily result in an unequal distribution of political liberties, a position which would be “humiliating and destructive of self-esteem” (TJ 477). Thus the lexical priority of liberty is required to maintain self-respect.

Taylor claims that this is another instance of Rawls committing the inference fallacy: “Rawls tries to derive the lexical priority of the basic liberties from the central importance of the interest they support – in this case, an interest in securing self-respect for all citizens” (Taylor 2003, 250). Why, he asks, is lexical priority necessary to protect our interest in self-respect? Might a merely very high level of priority be sufficient? Would small restrictions on the basic liberties in favour of a pareto improving efficiency gain, or a more just distribution of social and economic goods—such as an absolute increase in the position of the worst off—not be accepted by parties in the original position? Here we find the thrust of Taylor’s argument against Rawls’s justification of lexical priority: “the only way to justify something as strong as lexical priority for the basic liberties is to justify *lexical priority for the interest they support*...securing self-respect for all citizens might be such an interest, but Rawls’s arguments do not show why” (Taylor 2003, 251).

Consider, however, another passage from Taylor: “why...would very small restrictions on the basic liberties threaten the social basis of self-respect, so long as they were applied equally to all citizens?” (250). While it may be claimed that a restriction limiting a certain type of speech applies equally to everyone in the sense that no one may engage in the prohibited actions, this is not a relevant metric for assessing whether a restriction might be acceptable from the point of view of representative citizens in the original position. The problem lies in assuming that liberties hold equal value to all people. This is not the case under conditions of reasonable pluralism. The devout religious believer who finds her good in living a life according to the revealed truths of scripture and values sternly enforced adherence to its teachings will likely not place a great deal of weight on rights to free speech, perhaps even finding their exercise anathema to salvation. A more extreme case has the stalwart ascetic who, eschewing material

possessions and pleasures and even political participation, places no value on political liberties, even finding the idea of social status to be incoherent from the standpoint of his metaphysical views.

If these two types of people live in a society which places merely *high* priority on the basic liberties, and that society chooses to limit certain basic liberties, then the effects of this restriction will be hardly felt (if at all) by the pious and the ascetics. However, this same change would be keenly felt by anyone holding the opinions, ideas, or viewpoints which the change would forbid or restrict. The urban activist and the pornographer, for example, would feel quite keenly a whole slate of restrictions to political liberty that the ascetics would be utterly oblivious to. The activist and the pornographer stand to have their life plans interrupted, made more difficult, or even stymied entirely by restrictions on free speech or peaceful assembly, while the ascetics life plan requires neither of these liberties. Behind the veil of ignorance parties are (in part) concerned with their capacity to pursue, develop, and realize their comprehensive conceptions of the good. A small restriction of liberty, even if nominally applied to everyone, will only limit the capacity of a particular set of people to fully develop the moral powers, since a given restriction to basic liberties will affect those holding certain comprehensive doctrines more severely than others. Thus a restriction on the basic liberties cannot reasonably be said to apply equally to everyone *in the relevant sense*, regardless of the rise in efficiency and the sum of advantages which might be the result of such a restriction.

To this Taylor might respond that if the increases in benefits resulting from a restriction to basic liberties are made to accrue, in high proportion, to those negatively affected by it (the activists and the pornographers, say), then the trade-off should be acceptable even to those who have lost some amount of political liberty in the exchange. This reply seems to me to be mistaken. By trading liberties against some change in the distribution of wealth (even one characterized by a Pareto improvement in overall wealth), we run the risk of reintroducing the problem of relative socioeconomic status as a determinant of self-respect. Since in this case, while the citizens who lose a measure of liberty in the deal may still endorse it, there will be a set of citizens who, in absolute terms, will be unaffected by the trade-off in the sense and to the extent that they never cared to possess the liberties in question (recall the pious and the ascetics), but who will end up behind in terms of relative socioeconomic benefits once the activists and pornographers are bought off. That is, while compensating the citizens whose treasured liberties are being curtailed may entice those citizens to accept the restriction, a problem arises from the other direction when the set of citizens who didn't care about the restriction are then at a socioeconomic disadvantage (all else being equal). Even if the restriction results in a Pareto improvement in socioeconomic benefits, the net result would be an increase in the disparity of how these benefits accrue across society.

Important here is the point that when considering status, we are not concerned with benefits and burdens in absolute terms, or even Pareto improvements in total advantages, since both measures obscure the *relative* character of status. It is this which Rawls seeks to account for in assuring that, in a well-ordered society, the need for status is answered by the position of equal citizenship (*TJ* 478), thereby guaranteeing equal status (across this dimension) and therefore self-respect for everyone, while preventing the appeal to socioeconomic status which tends to be determinate of self-respect in societies that are not well-ordered. Socioeconomic status, as Rawls argues, is a poor and indeed intolerable determinant of self-respect in a just society, since certain features of the market and political organization make full equality along this dimension either impossible or undesirable. Equal citizenship through equal basic liberties,

on the other hand, is fully and equally realizable in a just society, and so lacks the self-respect problems raised by socioeconomic status.

So, given full knowledge of the effects of supporting a trade-off of basic liberties for socioeconomic advantages, the citizens who would be on the lower end of the benefit might veto the restriction of liberty because it has the potential to result in an unequal distribution of self-respect, due first to unbalancing equal citizenship and second to replacing fully equal citizenship with socioeconomic inequalities. That is, the citizens who, like the pious and the ascetics, value the liberties of equal citizenship in such a way that the restriction at hand would not effect them adversely (eg. any group which sees the liberty in question as unnecessary or even counter to the good life and thus need no recompense for its restriction) and hence receive a relatively lower benefit from the trade-off than do the burdened, who are compensated for the extra costs they bear from the restriction. Such citizens would be rightly concerned to see the benefits of a Pareto improvement accruing in relatively high proportion to a particular group or set of groups purely because of the nature of the comprehensive conceptions of the good to which they subscribe (cf. Van Parijs 1991, 105-108).

Thus, in answer to Taylor's question about why *lexical* priority is required, rather than merely a very high degree of priority, we may say that since the concept of equally distributed costs resulting from a restriction on liberty is not a coherent idea, anything but lexical priority will result in the unequal provision of the basic liberties of equal citizenship, thus granting the potential for socioeconomic status to re-emerge and supplant equal citizenship as the central determinant of self-respect. Since Taylor agrees that "Rawls convincingly argues that citizens in a just society would never consent to less than equal basic liberties" (Taylor 2003, 250), and I have shown equal provision of political liberties requires lexical priority, we have good reason to dismiss Taylor's assertion about the inference fallacy.

Thus far I have established that even the quite small tradeoffs implicit in granting only very high priority to liberty will be unacceptable to persons behind the veil of ignorance. Taylor grants that it may be the case that such minor restrictions on equal basic liberties will be ruled out as too risky, but he insists that even if this were so, "we would still need to ask why self-respect is of such overwhelming importance that its social basis, an equal distribution of the basic liberties, must be given lexical priority" (Taylor 2003, 250). His answer is that self-respect either rests sufficiently high in a hierarchy of interests that it may not be traded off against any other interest, regardless of the rate of exchange, or doesn't, and the lexical priority of liberty is unjustified (250-251). But this answer misrepresents what we ought to be concerned with when establishing principles of justice. What is at play when parties in the original position reject a potential principle or ordering rule is *not* considerations about a hierarchy of values which justice as fairness seeks to enshrine: since the parties behind the veil of ignorance are blind to the particular complex value judgements which individuals might use to construct personal value hierarchies, it seems unlikely that they might have the resources required to formulate a hierarchy for the people they represent that would be of sufficient complexity and specificity to serve Taylor's purpose. What the parties do have is knowledge that allows for the identification of citizens' "fundamental interests" (*JF* 85) in advancing their determinate conceptions of the good and in developing the two moral powers. While this may indeed be a type of interest/value hierarchy, with the fundamental or higher-order interests of persons resting at the top, self-respect in particular is *not* part of the hierarchy. Instead, self-respect is an instrumental good necessary for the development and security of the fundamental interests. Thus Taylor is wrong to suggest that it is self-respect itself that ranks highly on a hierarchy of values. His *quam* might

instead be with Rawls's arguments from moral personality and the fundamental equality of citizens that establish the fundamental status of the moral powers. But whatever Taylor's position on this might be, he doesn't pursue such a challenge, so I will turn back to his argument from the original position.

To reformulate Taylor's question, we should ask ourselves why it is that some principles are rejected in the original position. An answer to this should give us a clearer picture of what might be acceptable to people behind the veil of ignorance. I think the answer must be that a principle is rejected if, for some set of representative people, it imposes excessive, and indeed intolerable, strains of commitment. Principles of justice and their ordering rules are accepted by those in the original position if and only if the commitments, benefits, and burdens arising from a particular principle are of a sort that they can agree to in good faith and with a reasonable expectation of fulfilling their commitments over a complete life (*TJ* 153). So in determining the priority (lexical or otherwise) assigned to particular liberties we should not look to value hierarchies or the hierarchies of interests which these might suggest.

Thus the reason that lexical priority for the basic liberties is selected as a principle in the original position is that if the basic liberties were not given this exact degree of priority – their equal provision ensured above all things – then a significant number of real people would not accept the strains of commitment: they would reject any conception of justice which did not ensure the lexical priority of liberty. And since each party to the original position has an effective veto over principles, such a conception of justice would never reach the second stage in the four stage sequence. In this way the lexical priority of liberty functions to ensure that the parties in the original position “run no chance of having to acquiesce in a loss of freedom over the course of their life for the sake of a greater good enjoyed by others, an undertaking that in actual circumstances they might not be able to keep” (*TJ* 154).

It may be objected that the reliance on the strains of commitment as the ultimate source of lexical priority ignores the reasons people have for deciding whether and to what degree a particular principle might strain the bonds of commitment, and that this, at bottom, is where value hierarchies come into play and thus what Taylor attempts to avoid. However, when I wrote above that the parties in the original position lack the resources to be meaningfully concerned about the kind of value hierarchies and reasons Taylor is concerned with, I put forward a partial defence against this charge. Here I take up this objection in greater detail.

The Strains-of-Commitment Argument and Stability for the Right Reasons

To review: the self-respect argument states that the lexical priority of liberty is justified because to satisfy the demands of status without resorting to social and economic inequalities, Rawls turns to equal citizenship, established through equal basic liberties. In this way, Rawls pegs self-respect to the equal status of citizens in a society. Thus, self respect serves such a deeply fundamental interest that its assurance through equal basic liberties must be given priority. Since without such priority, it seems that the parties to the original position have failed in their duty to represent the fundamental interest of the citizens they represent. The equal liberty of conscience argument, on the other hand, holds that certain sets of beliefs are so important to people that they are unable to endorse the principles of any society which doesn't guarantee their liberties to form, hold and value those beliefs. Such societies impose excessive strains of commitment and conscience on the parties and cannot be the subject of good faith agreement behind the veil (*TJ* 153-154). Taylor further argues that only religious beliefs properly fit this description.

This is Taylor's picture of Rawls's two attempts at justifying the lexical priority of liberty. His central criticism of each is that they seek to establish absolute or lexical priority for liberty by appealing to the merely high priority of the interest the liberties support. I have argued that Taylor is mistaken in his characterization of Rawls; that Rawls is not best read as fallaciously bootstrapping liberty to lexically prior status. Instead, I have suggested that appealing to the strains of commitment and the role of self-respect serves to vindicate Rawls's confidence that the basic liberties must be lexically prior to the other elements of justice as fairness. In discussing each of Taylor's arguments against the priority of liberty, it has become clear that the argument from the strains of commitment plays a special role in Rawls's thought. To illustrate the centrality of this concept and to further elaborate its importance, I now turn to briefly consider the role of stability in Rawls's thought, and how this concept relates to the argument from the strains of commitment.

Stability, for Rawls, is the state of social cooperation where 1) the rules of social cooperation are regularly complied with and generally acted upon; and 2) deviations away from justice are met by corrective forces which tend to keep further deviations from occurring and to restore justice (*TJ* 398). He later clarified the concept by adding that stability is not sought simply for the sake of peace and order. Rather, the relevant sort of stability is a peaceful and orderly system that is stable *for the right reasons* (*PL* 392). This stability is based not on a modus vivendi constructed out of bare aversion to suffering as a result of oppression or collective action problems, but on people's sense of justice, whereby "those taking part in these [just] arrangements acquire the corresponding sense of justice and desire to do their part in maintaining them" (*TJ* 398).

Showing the central importance of stability on this account is relatively simple when we view the above two conditions for stability from the point of view of the original position. This way, instead of viewing stability as a descriptive state of societies, we see it to be a condition on the parties to the original position which forces them to select principles which individuals can and will willingly comply with over complete lifetimes, come what may (Freeman 2007, 182-183). We can see, then, that stability is related to the argument from the strains of commitment. Since 'the right reasons' for stability turn on the wide exercise and proper development of the moral power to form and act from a conception of justice, the highest-order interest we have in developing this moral power is tied up with our interest in stability.² This moral power is in turn protected as a highest-order interest in the original position by appeal to the strains of commitment. Thus the parties to the original position are bound to select principles which generate a stable society because instability indicates a lack or failure of reasonableness and as such does not properly respect our nature as reasonable and rational agents. Of course a stable society must still admit that either trembling hands or bouts of egoism may occur and result in deviations away from reasonableness and justice; in any collective action some individuals will invariably defect or free-ride. One point of stability and stability for the right reasons, then, is that citizens will develop (various) senses of justice which don't see them reacting to defection with excessive distrust or a loss of faith in cooperation altogether. Rather, Rawls's hope is that the desire to cooperate and to support just cooperation-facilitating institutions would be sufficiently robust to withstand the inevitable shocks—that the principles selected in the original position would be both feasible and enduring (*CP* 487).

² The first moral power is the capacity to be reasonable, the capacity to form and act on a determinate conception of justice. The second is the rational capacity to form, revise and pursue a coherent conception of the good life (Freeman 2007, 54).

But it may be that this position leaves my account open to a final response from critics like Taylor. They may claim, as I have, that it is a fundamental interest of citizens that they be able to pursue the development of their moral powers in a stable cooperative society, and that it is this interest in developing the moral powers which results in the need for stability. If this assertion is correct, then there may be some reason to think that something like Taylor's hierarchy-of-interests argument is correct. It may be that an appeal to the importance of peoples' interest in the development and exercise of the moral powers is what gives force to the lexical priority of the basic liberties. However, it was never my intention to eliminate interests from consideration entirely in ordering the principles of justice. Any theory of justice which fails to account for the interests and basic moral psychology of people is not only likely to be unconvincing, but most certainly will not pass the test of stability if established in the basic institutions of society. Instead, I sought to ground lexical priority in a less metaphysical interest than Taylor does when he pegs it to Kantian autonomy. So while development of the moral powers is certainly an interest in the normal sense of the term, we must also recognize that a prior condition for the realization of the moral powers is a stable society where no excessive strains of commitment or conscience are imposed in which to exercise these powers. *This*, not human flourishing or autonomy or some other comprehensive theory of the good, is the fundamental purpose of justice on Rawls's account: principles which enable people to exist in cooperative society and to share in the benefits and burdens of cooperation in ways which all reasonable persons can endorse. This makes clear the importance of stability in an account of justice. A society that is stable *for the right reasons*, that is, one where the basic institutions are endorsed in good faith by all who live under their coercive apparatus, is a necessary precondition for flourishing or autonomy or any number of other purposes to which others have cast justice, and is thus suitably neutral and therefore political in a way which Kantian autonomy is not.

Works Cited

- Freeman, Samuel. Rawls. New York: Routledge, 2007.
- O'Neill, Onara. "Constructivism and Contractualism," Ratio 16.4 (2003): 319-331
- "Constructivism in Rawls and Kant," The Cambridge Companion to Rawls. Ed. Samuel Freeman. Cambridge: Cambridge University Press, 2003.
- Rawls, John. Theory of Justice. Revised ed. Cambridge: Harvard University Press, 1999.
- Collected Papers. Ed. Samuel Freeman. Cambridge: Harvard University Press, 1999.
- Political Liberalism. New York: Columbia University Press, 1993.
- Justice as Fairness: A Restatement. Cambridge: Belknap Press, 2001.
- Taylor, Robert S. "Rawls's Defense of the Priority of Liberty: A Kantian Reconstruction" Philosophy & Public Affairs 31.3 (2003): 246-271.
- Van Parijs, Philippe. "Why Suffers Should be Fed: The Liberal Case for an Unconditional Basic Income" Philosophy & Public Affairs 20.2 (1991): 101-131.