

Politics, Philosophy & Economics

<http://ppe.sagepub.com>

Contractarianism and cooperation

Cynthia A. Stark

Politics Philosophy Economics 2009; 8; 73

DOI: 10.1177/1470594X08098872

The online version of this article can be found at:
<http://ppe.sagepub.com/cgi/content/abstract/8/1/73>

Published by:



<http://www.sagepublications.com>

On behalf of:

[The Murphy Institute of Political Economy](#)

Additional services and information for *Politics, Philosophy & Economics* can be found at:

Email Alerts: <http://ppe.sagepub.com/cgi/alerts>

Subscriptions: <http://ppe.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.co.uk/journalsPermissions.nav>

Contractarianism and cooperation

Cynthia A. Stark

University of Utah, USA

abstract

Because contractarians see justice as mutual advantage, they hold that justice can be rationally grounded only when each can expect to gain from it. John Rawls seems to avoid this feature of contractarianism by fashioning the parties to the contract as Kantian agents whose personhood grounds their claims to justice. But Rawls also endorses the Humean idea that justice applies only if people are equal in ability. It would seem to follow from *this* idea that dependent persons (such as the disabled) lack claims of justice. It appears, then, that the Kantian and Humean themes in Rawls conflict. I present a reading of Rawls that resolves this tension between the Kantian and Humean themes. The first theme, I argue, allows Rawls to maintain that persons as such are owed justice regardless of their ability to engage in social cooperation. The second theme, I argue, allows him to retain Hume's connection between justice and reciprocity, but confines the reciprocity condition to relations among nondependents. I conclude that Rawls's approach permits him to rebut recent criticisms leveled by disability theorists and others who claim that his theory excludes dependents.

keywords

Rawls, reciprocity, disability, dependency, circumstances of justice

A commonly identified shortcoming of contractarian accounts of justice is that they provide a tenuous or perhaps even objectionable *ground* for people's claims to justice.¹ Because contractarians are generally committed to the idea of justice as mutual advantage, they hold that relations of justice, at least insofar as they can be rationally grounded, obtain only when each person can expect to gain from them. Individuals are subject to justice, then, only if it is to the advantage of others to cooperate with them. They do not have claims to justice on the basis of any normative considerations. They are not owed justice in virtue of being

persons or having interests and so on. John Rawls seems to avoid this shortcoming of contractarianism by fashioning the parties to the contract as Kantian persons. The parties have a special moral status in virtue of their capacity for morality, which presumably grounds their claims to justice.²

A less discussed aspect of Rawls's view is his reliance upon the Humean idea that justice applies only in circumstances where people are roughly equal in ability.³ Where there are great discrepancies in ability, so that one individual or group can subjugate or exclude others, relations of justice are not possible because they are not mutually beneficial – they are not to the advantage of those capable of subduing others. It follows that in such circumstances of inequality, the subjected or excluded have no claims of justice against their oppressors and oppressors have no duties of justice toward those whom they dominate or marginalize.

My primary aim in this article is to defend a reading of Rawls that resolves a tension in his view resulting from his adopting both the Kantian ideal of the person and the Humean account of the circumstances of justice. The issue is this: if one's status as a person is sufficient to ground one's claim to justice, then it is irrelevant to justice whether one is equal to others in ability. So, the Kantian ideal of the person seems to annul the Humean requirement of equal ability. On the other hand, if one's being equal to others in ability is necessary for one to have claims of justice, then one's status as a person is not sufficient. So, the Humean requirement of equal ability seems to subvert the role of the Kantian ideal of the person. Insofar as Rawls insists upon the equal ability condition, then, it seems that he cannot surmount the shortcoming of contractarianism in a satisfactory way; so long as people's relative abilities bear upon their claims to justice, those claims may have a tenuous or objectionable ground.

I argue that the Kantian and Humean themes can be interpreted as working in concert. Together these themes allow Rawls to propose an inclusive ground for individuals' claims to justice (personhood) while also allowing a role for reciprocity as an element of justice. By relying on the Kantian ideal of the person, Rawls can maintain, contrary to Hume, that all persons are owed justice, regardless of whether they are in relations of cooperation with one another. At the same time, the equal ability condition (or what becomes, in Rawls's later work, the assumption that all citizens are fully capable of cooperation) permits Rawls to retain Hume's connection between justice and reciprocity in circumstances in which persons *are* in relations of cooperation.

My secondary aim is to counter critics who claim that the fully cooperating assumption corrupts Rawls's theory by demoting the justice claims of the non-cooperating. I argue that the fully cooperating assumption serves to preserve the element of reciprocity in Rawls's theory, which element, I contend, is part of our pre-theoretical notion of justice. I argue further that the fully cooperating assumption does not necessarily demote the justice claims of the non-cooperating, but rather allows the issue of distributive justice for the non-cooperating to be treated separately from the issue of distributive justice for the cooperating. Such an

approach might be fruitful given the different moral problems raised by dependency relations than those raised by relations among coequals.

I begin by explaining why the equal ability condition is crucial to one kind of contractarianism, commonly known as ‘bargaining contractarianism’.⁴ I then explain why Rawls rejects bargaining contractarianism and I describe the version of contractarianism he defends.⁵ It turns out that the equal ability condition is an appendage that does no work in Rawls’s theory. Furthermore, the equal ability condition seems to conflict with a pivotal notion of Rawls: the claim that natural talents are arbitrary from a moral point of view.

Second, I note that where Rawls seems to abandon the equal ability condition, he introduces ‘the fully cooperating assumption’. This assumption, which states that all members of society are to be regarded as fully capable of social cooperation, enables Rawls, I claim, to get results that the bargaining theorist gets by means of the equal ability condition: it allows him to confine the scope of mutual advantage, and hence the scope of his principles of justice, to all and only those who contribute to the cooperative surplus.⁶ Third, I show that the fully cooperating assumption is only apparently inconsistent with Rawls’s contention that natural talents are morally arbitrary. This discussion establishes the consistency between the Humean and Kantian themes in Rawls’s account.

Fourth, I examine Rawls’s support for the fully cooperating assumption, which rests upon the claim that the problem of justice for those cooperating is more fundamental than the problem of justice for the non-cooperating. After identifying a problem with this account, I present an alternative argument for the fully cooperating assumption that is available to Rawls. Finally, I explain why the fully cooperating assumption does not necessarily discount the justice claims of the non-cooperating.

Bargaining contractarianism

Bargaining contractarianism regards principles of justice as constraints on self-interest necessary for social cooperation, which rational actors would endorse because they would be worse off in the absence of such principles.⁷ The situation of no cooperation, or the nonagreement point, is characterized as sufficiently ‘inconvenient’ that the path to one’s ends is fraught with obstacles. It is to the advantage of each, then, to agree to rules that constrain everyone’s conduct so as to realize a system of social cooperation, thereby smoothing the path to achieving one’s ends. Justice is mutually advantageous in the sense that it makes everyone better off than they would be at the nonagreement point.

The problem of arriving at mutually advantageous terms of cooperation, on this view of justice, is framed as a bargaining problem (hence the label). The parties to the bargain, on this account, are seen as equipped with different abilities and they are assumed to prefer principles that will maximize their share of the goods (rights, opportunities, material items, and so on) created by social cooperation.

The principles adopted embody the concessions each is willing to make in order to procure the cooperation of others. For example, suppose that two people, Marty and Mary, will receive a portion of US\$100 only if they can agree how to divide it up. If they cannot agree, they get nothing, so both are better off if they agree to some division or another. Suppose further that Mary is relatively indifferent to whether she gets a share. Suppose Marty, on the other hand, is desperate to get at least something. Imagine that the only division Mary will accept is US\$75:US\$25 (in her favor). She is willing otherwise to get nothing. Since Marty would rather have US\$25 than nothing, it is in his interest to accept Mary's offer.⁸ This division, then, reflects the concessions each is willing to make, given their situation, in order to improve their prospects.

Bargaining contractarians tend to follow Hume in holding that certain conditions must obtain in order for justice to arise. One of these 'circumstances of justice', which I identified above, is that people be roughly equal in ability. 'In order for human cooperation [to be] both possible and necessary', individuals must be 'roughly similar in both physical and mental powers; or at any rate, their capacities are comparable in that no one among them can dominate the rest'.⁹ It would not be in the interest of those far more superior to others to agree to principles of justice constraining their conduct because 'superiors' could simply exploit or exclude 'inferiors'.¹⁰ Conflicts would be suppressed through coercion, in the form of either domination or marginalization.¹¹ Superiors presumably would be better off under this arrangement than they would be in a system of cooperation governed by rules of conduct applying to all, for this latter arrangement would likely prohibit their exploiting or excluding inferiors. Yet they would obtain no countervailing benefit, since they would be able to avoid insecurity without rules of justice. In short, superiors, if their talents far surpass the abilities of inferiors, would not benefit from social cooperation and therefore it would not be in their interest to agree to principles setting out terms of cooperation.

For the bargaining contractarian, conditions characterized by the absence of principles of justice are not properly described as just or unjust. This is because justice is, according to contractarians, an 'artificial virtue', to use Hume's term: what people are owed in terms of justice is given by the principles to which it is in their interest to agree; there is no prior notion of justice. So, in the absence of such principles, people have no duties of justice toward one another, nor claims of justice against one another.¹²

One implication of this view is that the vulnerable have no claims of justice against those who might harm them. As Hume famously states:

Were there a species of creatures intermingled with men, which, though rational, were possessed of such inferior strength, both of body and mind, that they were incapable of resistance, and could never, upon the highest provocation, make us feel the effects of their resentment; the necessary consequence, I think, is that we should be bound by the laws of humanity to give gentle usage to these creatures, but should not, properly speaking, lie under any restraint of justice with regard to them, nor could they possess

any right of property . . . the restraints of justice and property, being totally *useless*, would never have place in so unequal a confederacy.¹³

For Hume, then, since the advantaged have no use for principles of justice regulating their relations with the disadvantaged, such principles would not apply.¹⁴ Translated into contractarian language, the point is that the vulnerable stand outside of justice because they lack the assets necessary to make it rational for the advantaged to agree with them on terms of cooperation.

There have been a number of proposals as to which principles of distributive justice rational, self-interested bargainers roughly equal in ability would choose at a specified nonagreement point.¹⁵ What is important for our purposes is not the details of these types of proposals, but their main characteristics: mutual advantage is measured against the nonagreement point; principles are the result of a bargain struck at that point; bargainers' capabilities are permitted to influence the distributive outcome; and those who lack the assets necessary to make it rational for others to strike a bargain with them stand outside of justice.

Rawls's rejection of bargaining contractarianism

Rawls rejects bargaining contractarianism outright. He denies that principles that result from a bargain count as genuine principles of justice because such principles reflect people's strategic capabilities, which are, in his view, irrelevant to justice. He says:

it is to avoid appeal to force and cunning that principles of right and justice are accepted. Thus I assume that to each according to his threat advantage is not a conception of justice. It fails to establish an ordering [on conflicting claims] in the required sense, an ordering based on certain relevant aspects of persons and their situation which are independent from their social position, or their capacity to intimidate or coerce.¹⁶

The arrangement arrived at by Marty and Mary, then, is not just, even though it is mutually beneficial and they agreed to it, for it reflects Mary's bargaining strength. 'What is lacking', Rawls tells us, 'is a suitable definition of a status quo that is acceptable from a moral point of view.' He continues, 'We cannot take various contingencies as known and individual preferences as given and expect to elucidate the concept of justice (or fairness) by theories of bargaining.'¹⁷

In order for principles resulting from hypothetical agreement to count as genuine principles of justice, Rawls tells us, it must be the case that they would be agreed upon under circumstances that are fair. Rawls provides such an initial situation, labeled 'the original position', which stands as the theoretical substitute for the nonagreement point. Rawls says, 'The original position is designed to meet the problem of the appropriate status quo.'¹⁸ He agrees with bargaining theorists that at the nonagreement point people's prospects would be poor. He describes the nonagreement point as a condition of 'general egoism' in which each person is permitted to do whatever he sees fit to achieve his aims.¹⁹ But

Rawls thinks that the nature of the nonagreement point has no bearing upon the justification of principles of justice. So, the nonagreement point plays no role in his theory except to represent what people would be 'stuck with' if the parties in the original position 'were unable to reach an understanding'.²⁰

The original position, as the appropriate status quo, is characterized, in Rawls's early work, as follows. The deliberating parties, as I stated above, represent persons understood in Kantian terms – as free and equal rational beings.²¹ They are equal in the sense that they have the same moral standing, and they are free in the sense that they are not bound by prior moral principles. They are assumed to have a sense of justice and a conception of the good.²² (Rawls later refers to these capacities as 'moral powers'.) Their rationality consists in their being concerned to advance their interests and preferring more as opposed to fewer primary social goods. Their being rational also entails that they are not moved by envy.²³ The parties are, furthermore, regarded as mutually disinterested.²⁴ Finally, they are situated behind a veil of ignorance, which conceals from them the natural abilities, social advantages, conception of the good, psychological propensities, and so on of those whom they represent.²⁵

The parties are charged with choosing, on rational grounds, principles of justice understood as mutually advantageous terms of cooperation.²⁶ In this respect, the parties are similar to the hypothetical agents of bargaining contractarianism. However, because they are ignorant of the natural abilities and social advantages of the persons whom they represent, they cannot judge how well off those persons would be at the nonagreement point nor how well off they would be under any particular principles. Naturally, therefore, the parties cannot bargain for principles that will maximize the prospects of those whom they represent relative to the nonagreement point. They must make a rational choice on some other ground.

Rawls tells us that, due to the veil of ignorance, it would be rational for the parties to agree to arrange the basic structure of society so that certain goods (rights, liberties, and opportunities) tend to be distributed equally. The parties would require that the basic structure distribute wealth equally also, Rawls says, unless the prospects of those who would have the least under an unequal distribution are improved the most by the existence of inequality. The parties would agree to terms of cooperation, in other words, designed to give to the least well off the maximal share such that, if it were attempted to give them more, they would end up with less.²⁷ Rawls calls this principle of justice the 'difference principle'. It allows inequalities in wealth as long as they are to greatest advantage of the least benefited.

The difference principle is a rational choice for the parties in the original position because it makes everyone better off than they would be under an equal division.²⁸ An equal division represents a 'suitable benchmark of comparison', according to Rawls, because it is the distribution that equal moral persons would (initially) choose in ignorance of their abilities and ideals. It is, in other words, the division that would be (initially) chosen if all and only morally relevant

factors were permitted to influence the choice of principles. One's claim to justice, then, is grounded in one having the two moral powers, and principles of justice are designed to be responsive to those facts about one and not one's strategic capabilities.

Rawls's approach appears at first blush to represent a stark contrast with bargaining contractarianism. Mutual advantage is measured against a benchmark of equality, which is invoked on explicitly moral grounds. The principles adopted are those that it is rational for moral persons to consent to in ignorance of their strategic capabilities and the distributive outcome is *not* influenced by individuals' talents.²⁹ Finally, it would seem that the disadvantaged, on Rawls's account, would not stand outside of justice, since one's claim to justice is rooted in one's personhood and does not depend on one having the assets necessary for others to include one in a bargain for justice.

Equal ability and the ability to cooperate

This contrast, though, is muddled by Rawls's official endorsement of the equal ability condition. Moreover, his adherence to this condition is somewhat mysterious because, whereas it is essential to bargaining contractarianism, it is superfluous to Rawls's contractarianism, given his formulation of the initial situation. Further yet, the equality of ability condition seems at odds with Rawls's contention that natural talents are arbitrary from a moral point of view.³⁰

Recall the close connection between the equal ability condition and the bargaining theorist's characterization of the initial situation: the highly talented (who are aware of their talents) have no rational ground for agreeing to principles of justice because they can see that no principles exist that would make them better off than they are without the principles. Rawls's renovation of the initial situation severs the connection between one's relative capabilities and the rationality of agreeing to terms of cooperation. The parties in the original position are ignorant of the extent of their abilities and they are required to choose principles in spite of that ignorance. If one does not know that one is very talented then one cannot know that one would as such be disadvantaged by rules of cooperation. So, if it is rational for one to agree to terms of cooperation nonetheless (as Rawls thinks it is), then it must be rational regardless of the fact that one is extremely talented relative to others. If it is rational for people to agree to terms of cooperation without knowing what their relative abilities are, then the rationality of choosing those terms cannot turn on their relative abilities.

By denying the parties in the original position awareness of their talents, Rawls's theory shifts the ground for claims to justice from agents' strategic capabilities to their moral capabilities. Individuals are *owed* justice, on his view, and they are owed it to the extent that they are persons in the Kantian sense. They are not merely subject to justice on the basis of the fact that they have the right degree of capability to make it rational for them to agree with others on princi-

ples of justice and to make it rational for others to agree with them. *The assumption of equal ability, then, is not needed, in Rawls's theory, for it to be rational for everyone to agree to terms of cooperation.* All that is required is that there is some rule upon which it is rational for Kantian persons who know nothing of their abilities to agree. Of course, Rawls thinks that the difference principle is such a rule.

But the equal ability assumption is not merely unnecessary to the rationality of agreeing to terms of cooperation, it appears to conflict with Rawls's claim that people's capabilities are irrelevant to the justification of principles of justice. Rawls rejects bargaining theories *precisely because* they allow people's capabilities improperly to influence the distributive outcome. So, not only does Rawls not need the equality of ability assumption, it appears he is not entitled to it: his claim that strategic capabilities are morally irrelevant to distributive justice would seem to entail that people's claim to justice must rest exclusively on the attributes of personhood and not at all on their having the requisite abilities to be parties to the social contract.

It is tempting to resolve the tension in Rawls between the logic of the initial situation and the equal ability condition by simply ignoring the latter. Rawls was mistaken, we might conclude, in adopting that feature of Hume's account of the circumstances of justice. Indeed, textual grounds for this resolution exist. In his description of the circumstances of justice in *Political Liberalism*, in contrast to his account in *A Theory of Justice*, Rawls makes no reference to equality of ability.³¹ He does, however, introduce the fully cooperating assumption. He states in the introduction to *Political Liberalism*:

Since we begin from the idea of society as a fair system of cooperation, we assume that persons as citizens have all the capacities that enable them to be cooperating members of society. This is done to achieve a clear and uncluttered view of what, for us [philosophers?], is the fundamental question of political justice: namely, what is the most appropriate conception of justice for specifying the terms of cooperation between citizens regarded as free and equal, and as normal and fully cooperating members of society over a complete life? . . . given our aim, I put aside for the time being . . . temporary disabilities and also permanent disabilities or mental disorders so severe as to prevent people from being cooperating members of society in the usual sense.³²

Where Rawls should, or perhaps does, abandon the idea that justice is possible only where people are equal in ability, he stipulates that, for the purposes of his theory, the range of difference between people's abilities is to be regarded as small enough that each person is fully capable of social cooperation.

Rawls is not entirely clear about what it means for an individual to be a normal and fully cooperating member of society over a complete life. His scattered remarks on this issue paint, I believe, the following picture. First, as the passage quoted above suggests, those with severe mental, cognitive, or physical disabilities are not fully cooperating. I assume that this group includes both persons and nonpersons in Rawls's sense. That is, the group includes individuals who lack the

two moral powers associated with personhood (the capacity for a sense of justice and for a conception of the good) and those who possess these powers, but who are otherwise severely impaired mentally, cognitively, or physically.³³

Second, I interpret Rawls's stipulation that the fully cooperating are those who are 'associated together over a complete life' to indicate that he is taking into account the normal life cycle of humans and so means to include in the category of fully cooperating children who will participate in social cooperation and elderly adults who have participated in social cooperation.

Disability theorists have stressed that whether an impairment limits one's ability to participate in society is often a matter not simply of technological and medical know-how, but of society's willingness to accommodate those with unusual needs. With this in mind, third, I understand the class of people who are not capable of social cooperation to include those whose impairments render them unable to cooperate given a society's particular level of technological advancement, even if that society is committed to making accommodations for the impaired.³⁴

In sum, I interpret the non-cooperating to include both nonpersons³⁵ and persons who are so impaired physically, mentally, or cognitively that, in spite of technological support and political will, they are not capable of participating in a scheme of cooperation and hence not capable of helping produce the benefits nor helping shoulder the burdens of cooperation.³⁶ The non-cooperating are, therefore, not able to stand in relations of reciprocity, as Rawls understands that idea, for, on Rawls's view, to be capable of reciprocity is to be capable of contributing to a scheme of social cooperation.³⁷

It turns out, then, that the fully cooperating assumption has implications for Rawls's theory that are very similar to those of the equal ability condition. First, the fully cooperating assumption serves to maintain the tension in Rawls's theory I described at the outset: the Kantian theme suggests that the capabilities of personhood are sufficient to be owed justice; the fully cooperating assumption suggests that, to be owed justice, one must have the capacity to participate fully in the institutions that create the primary social goods. Second, the fully cooperating assumption enables Rawls to align himself with the bargaining theorist in confining the scope of mutual advantage to those capable of contributing to the cooperative surplus. In turn, it permits Rawls to limit the scope of his principles of justice to those capable of contributing to the cooperative surplus. The second of these implications merits elaboration.

Confining the scope of justice to the fully cooperating

For the bargaining theorist, recall, mutually advantageous principles must benefit all and only cooperators in order for those principles to serve as the basis of rational agreement: all cooperators must benefit because it would be irrational for any contributor to agree to a principle that does not benefit her; only cooperators

must benefit because it would be irrational for any contributor to agree to a principle that extends benefits to those who add nothing to the social pie. Confining the scope of mutual advantage to the fully cooperating is entailed by the bargaining theorist's account of what would constitute a rational choice in the initial situation. Call the notion of mutual advantage that includes all and only contributors the 'strict' notion.

Now consider the difference between the bargaining theorist's description of the initial situation and that proposed by Rawls. According to the logic of the original position (prior to the addition of the fully cooperating assumption), a principle that benefits the non-cooperating as well as the cooperating *could* serve as the basis of rational agreement: if the parties are unaware of the talents of those whom they represent, they must be unaware of whether those whom they represent are capable of social cooperation. So, if there is a principle on which the parties can agree in the absence of this knowledge, it must be a principle that ignores persons' abilities to contribute to the production of social goods. In the original position, then, the rationality of adopting principles of justice turns not at all on anyone's ability to contribute to the stock of goods. Confining the scope of mutual advantage to the fully cooperating, therefore, is *not* entailed by the account of what would constitute a rational choice in the initial situation. Indeed, what would count as a rational choice in the initial situation implies what we can call the 'relaxed' notion of mutual advantage: all cooperators, but not only cooperators, must gain from principles of justice.³⁸

The scope of mutual advantage is confined to the fully cooperating, in Rawls's theory, then, simply by stipulation. Through the fully cooperating assumption, Rawls ensures that arrangements agreed to in the original position are mutually advantageous, not in the relaxed sense, as is implied by the initial description of the original position, but in the strict sense: all *and only* those who have something to offer to a scheme of social cooperation are guaranteed to benefit from the terms of cooperation as compared to the appropriate baseline. The fully cooperating assumption, then, enables Rawls to maintain that his principles of justice must be mutually advantageous, not in virtue of benefiting every *person*, but in virtue of benefiting *everyone engaged in social cooperation*.

The fully cooperating assumption also allows Rawls to limit the application of the difference principle to the fully cooperating. It transforms the difference principle from a principle that, given the initial account of the original position, *could be* rationally agreed upon by individuals regardless of their ability to engage in cooperation into a principle that *is* agreed upon (hypothetically) by only fully cooperating individuals – those not capable of social cooperation have been assumed out of the theory. The result is that the difference principle distributes the primary social goods to all and only fully cooperating individuals. Because everyone is assumed to contribute to the cooperative surplus, Rawls guarantees that the difference principle applies only where people can stand in relations of reciprocity.

While the fully cooperating assumption functions in Rawls's theory in much the way the equal ability assumption functions in bargaining theories, there is a crucial difference in the way each confines the scope of justice to those capable of participating in a scheme of cooperation. The equal ability condition, as a prerequisite for the rationality of agreeing to any principles of justice, implies that such principles can be rationally legitimated only where people are roughly equal in ability. Principles that extend a share of the cooperative surplus to those who are a net drain on the surplus lack a rational basis, and therefore, those who cannot contribute have no legitimate claims upon the benefits of social cooperation.

There are at least two ways contractarians have dealt with the unwelcome fact that justice as mutual advantage, in the strict sense, confines the scope of justice to those whose capacities fall within the normal range. First, one can claim, as Hobbes does, that *in fact* all humans are roughly equal in ability, so *in fact* the scope of justice includes all humans. Hobbes says:

Nature hath made men so equall, in the faculties of body, and mind; as that though there bee found one man sometimes manifestly stronger in body, or of quicker mind then another; yet when all is reckoned together, the difference between man, and man, is not so considerable, as that one man can thereupon claim to himselfe any benefit, to which another may not pretend, as well as he. For as to the strength of body, the weakest has strength enough to kill the strongest, either by secret machination, or by confederacy with others, that are in the same danger with himselfe.

And as to the faculties of the mind . . . I find yet a greater equality amongst men, than that of strength . . . That which may perhaps make such equality incredible, is but a vain concept of ones owne wisdom, which almost all men think they have, in a greater degree, than the Vulgar.³⁹

But, of course, it is plainly false that all humans are equal in their capabilities, the possibility of coalitions notwithstanding.⁴⁰ Severely mentally or physically impaired individuals, children, those who are injured or charged with the care of infants must depend upon others to meet their basic needs. This creates an acute, long-term (and in some cases permanent) asymmetry of power. So, Hobbes's approach buys inclusivity at the cost of implausibility.

A second approach is taken by Gauthier. Gauthier simply takes it as an unfortunate implication of an otherwise sound theory that contractarianism tends to exclude the severely impaired from the domain of justice. He says:

From a technology that made it possible for an ever-increasing proportion of persons to increase the average level of well-being, our society is passing to a technology, best exemplified by developments in medicine, that make[s] possible an ever-increasing transfer of benefits to persons who decrease that average. Such persons are not party to the moral relationships grounded by a contractarian theory.

In a footnote to this passage, he says:

The problem here is not care of the aged, who have paid for their benefits by earlier productive activity. Life-extending therapies do, however, have an ominous redistribu-

tive potential. The primary problem is care for the handicapped. Speaking euphemistically of enabling them to live productive lives, when the services required exceed any possible products, conceals an issue which, understandably, no one wants to face.⁴¹

Where Hobbes's view is arguably implausible on empirical grounds, Gauthier's is arguably untenable on normative grounds.⁴²

The fully cooperating assumption permits Rawls to avoid the pitfalls encountered by the likes of Hobbes and Gauthier, for it does not imply that no one is disadvantaged nor does it imply that the disadvantaged lack claims of justice. It merely entails that the difference principle is not responsive to the claims of justice held by the disadvantaged. The difference principle is, as it were, *silent* on the issue of what is owed to the non-cooperating. The fully cooperating assumption, then, lets Rawls admit, contrary to Hobbes, that some people are far less capable than others, without having to hold, with Gauthier, that such individuals are outside of the scope of justice.

Because Rawls is committed merely to the idea that *his principles of justice* must be mutually advantageous in the strict sense, he leaves open the possibility that there are principles of justice that govern the distribution of goods to the non-cooperating. He leaves open the possibility, in other words, that an acceptable conception of justice need be mutually advantageous only in the relaxed sense. Indeed, he characterizes the ideal of mutual advantage, which in his later work he calls 'reciprocity', as the relaxed notion. He defines it as a situation in which 'all [but not only those] who are engaged in cooperation and who do their part as the rules and procedure require, are to benefit in an appropriate way as assessed by a suitable benchmark of comparison'.⁴³ It is imperative that Rawls structure his theory to allow the scope of justice to include non-contributors because, as we saw above, he holds that the ground of one's claim to justice is one's status as a person, not one's status as contributor to the cooperative surplus.

The moral relevance of the capacity to cooperate

Two implications of my discussion so far should be noted. First, Rawls goes to great lengths to safeguard the strict notion of mutual advantage: where his original characterization of the initial situation eschews the strict idea, allowing that his principles of justice might distribute benefits to everyone, regardless of whether they contribute to the cooperative surplus, his later introduction of the fully cooperating assumption, which ensures that only productive members of society receive the benefits of social cooperation, embraces the strict idea. Clearly, then, Rawls regards the strict notion of mutual advantage as central to his account. Second, it is Rawls's effort to protect the strict idea of mutual advantage that seems to endanger his claim that native abilities are morally arbitrary. While the fully cooperating assumption gives him the strict sense of mutual advantage, it at the same time implies that productive capabilities are morally

relevant. The fully cooperating assumption, in other words, entails that the parties in the original position represent, and know they represent, people who have productive capabilities, thereby ensuring that facts about their cognitive and physical skills, and not merely facts about their capacities as persons, influence the justification of principles of justice.

So, in order to resolve the tension between Rawls's Kantianism and the fully cooperating assumption, we need an interpretation that shows that assumption to be compatible with the notion that a person's natural abilities are arbitrary from a moral point of view. I offer the following straightforward account. While Rawls holds that a person's natural abilities are morally arbitrary when they are *within the normal range*, he regards having the requisite level of ability to engage in social cooperation as *not* morally arbitrary. The following suggests that Rawls holds this view:

in establishing the fair terms of cooperation (in the case of the basic structure) the only relevant feature of persons is their possessing the two moral powers (to the sufficient minimum degree) and having the normal capabilities to be a cooperating member of society over a complete life. Features relating to social position, native endowment, and historical accident, as well as to the content of persons' determinate conceptions of the good, are irrelevant, politically speaking and hence placed behind a veil of ignorance.⁴⁴

This passage suggests that Rawls distinguishes between persons' native abilities and the abilities they must have to engage in social cooperation. On his view, having the normal capabilities to be fully cooperating is classed (along with the two moral powers) as relevant to the justification of principles of justice, while 'native endowment' is classed among those factors that are irrelevant to the justification of such principles.

An obvious objection to my interpretation of Rawls is the following. Surely Rawls cannot consistently hold that natural abilities are morally arbitrary, while the ability to engage in social cooperation is morally relevant.⁴⁵ After all, is not the ability to cooperate a kind of natural talent? The issue of what makes a characteristic morally arbitrary is tricky and I cannot here address all of the questions associated with it.⁴⁶ But I can say this: it is inconsistent to hold that a contributor's natural talents are morally arbitrary, while the ability to contribute is morally relevant only if one assumes two things. The first is that what makes something morally arbitrary is that it is undeserved,⁴⁷ and the second is that the ability to contribute to a scheme of social cooperation must be viewed as, or as akin to, a natural talent. Both of these assumptions, I contend, are mistaken.

Consider first the relation between being undeserved and being morally arbitrary. It is not clear exactly how Rawls conceives this relation. He certainly stresses the notion that our natural talents and social position are undeserved and he stresses the notion that they are also arbitrary from a moral point of view. He says, for example, 'it is one of the fixed points of our moral judgments that no one deserves his place in the distribution of natural assets any more than he

deserves his initial starting place in society'.⁴⁸ Moreover, in rejecting what he calls the 'liberal' interpretation of the difference principle, he says, 'distributive shares [on this interpretation] are decided by the outcome of the natural lottery; and this outcome is arbitrary from a moral point of view'.⁴⁹ But on only one occasion does he make claims about undeservedness and moral arbitrariness together, and even then, he does not specify the relation between the two notions. He states:

The two principles [of justice] . . . seem to be a fair basis on which those better endowed, or more fortunate in their social position, neither of which we can be said to deserve, could expect the willing cooperation of others when some workable scheme is a necessary condition for the welfare of all . . . They express the result of leaving aside those aspects of the social world that seem arbitrary from a moral point of view.⁵⁰

Nothing in these remarks implies that the fact that our natural talents and social position are undeserved is sufficient for their being morally arbitrary. To be sure, Rawls's remarks imply that if our natural talents were deserved, then they would be morally relevant.⁵¹ But this position is compatible with some undeserved things also being morally relevant. Hence, there is room in Rawls's theory for the idea that the inability to cooperate, though undeserved, is nonetheless morally relevant to distributive justice.

Consider now the idea that the ability to cooperate is or is like a natural talent. One might grant the interpretation of Rawls that allows that some undeserved things are morally relevant, but maintain that the inability to cooperate does not qualify as one of those things since the inability to cooperate is largely a result of one's natural abilities. What is significant, I counter, about the inability to cooperate is not that it is related to one's natural talents, but that it places one in a special category with respect to society, if society is regarded, as Rawls proposes, as a system of cooperation.⁵²

Suppose that Rawls is right that, from the point of view of justice, the contribution of the truck driver is no more or less valuable than the contribution of the day-care worker and so neither has a right that the basic structure be designed to reward more highly her contribution.⁵³ It is not inconsistent with *that* idea that the status of the day-care worker and the truck driver is different, with respect to distributive justice, than the status of a severely impaired individual with unusual needs. The idea that one who is dependent on others may have a right that the basic structure be designed to recognize his needs is consistent with the idea that no one who is not dependent on others has a right that the basic structure be designed to remunerate his type of contribution more highly than the contribution of others (without providing compensating benefits to the worst-off contributor). In other words, it is not inconsistent to hold, on the one hand, that contributors' natural talents should be treated as morally arbitrary (that is, that the basic structure should not be designed so that contributors' shares of the social product are indexed to their natural talents) and to hold, on the other hand, that one's lacking

the traits necessary to contribute to the social product should be treated as morally relevant to determining one's share of the social product.

Why fully cooperating?

Given that the fully cooperating assumption significantly limits the scope of Rawls's theory of justice, it is worth exploring why he proposes it. In what follows, I set out Rawls's reasoning in support of the fully cooperating assumption. His argument, I maintain, fails for reasons relating to the justificatory structure of his theory. Nonetheless, Rawls has reason to preserve the morally relevant status of productive capabilities that is implied by the fully cooperating assumption. This is because, I argue, the moral relevance of productive capabilities is linked to an aim central to social contract theory, which is to prohibit the exploitation of those participating in a scheme of social cooperation. If we wish to take this principle seriously, *and* take seriously the justice claims of the non-cooperating, we should treat the issues of justice for the cooperating and for the non-cooperating separately. We should read Rawls, then, as doing something along those lines.

In the passage quoted above, in which Rawls states the fully cooperating assumption, he claims that the problem of justice for the fully cooperating is the fundamental problem of justice, implying that the problem of justice for the non-cooperating is less fundamental. By 'fundamental', Rawls seems to intend 'clearer' in the sense of 'less philosophically complicated'.⁵⁴ In defending the fully cooperating assumption he says 'it is sensible to set aside certain difficult complications', adding, 'If we can work out a theory that covers the fundamental case, we can try to extend it to other cases later.'⁵⁵ Indeed, he identifies the question of what is owed to those incapable of social cooperation as a 'problem of extension'.⁵⁶

Rawls's remarks about whether a right to abortion is supported by public reason are instructive in this context. He says:

consider the troubled question of abortion. Suppose first that the society in question is well-ordered and that we are dealing with the normal case of mature adult women. It is best to be clear about this idealized case first; for once we are clear about it, we have a guide that helps us to think about other cases, which force us to consider exceptional circumstances.⁵⁷

We can read the fully cooperating assumption, then, as expressing a similar view about philosophical method: begin with easy cases and rely upon the insights and conclusions generated by theorizing about those cases to aid in deliberation about more complex cases.⁵⁸

If this reading is correct, Rawls sees postponing the question of justice for the non-cooperating as similar to postponing, for example, the issue of abortion for adolescents or sex-selected abortion in his discussion of abortion and public

reason. These issues (teenage and sex-selected abortion) do indeed raise difficult complications that are more easily handled after the issue of non-sex-selected abortions for adults has been settled. However, postponing the issue of justice for the non-cooperating, in contrast to postponing issues of sex selection in the abortion case, does not, I claim, as Rawls may think, merely set aside issues that are more easily handled after the issue of justice for the cooperating has been settled.

To see the difference between the two cases, we must attend to the justificatory structure of the two arguments. Consider, first, the case of abortion and public reason. In order to ascertain what policies regarding pregnancy termination can be supported by public reason, one must look to the public political culture of one's society, including its official documents (for example, its constitution), interpretations of those documents, and public practices and public discussions about those practices.⁵⁹ One looks for socially recognized values (in this case, those related to reproductive rights and responsibilities) on which there is wide consensus and which direct one toward particular policies regarding abortion. These values would include, for example, equality, liberty, fairness, reciprocity, the survival and reproduction of society over time, public health, and so on.

Once one determines what abortion policy is supported by public reason in the 'normal' case of adult women, one can make inferences from that case to more complicated cases. For instance, if it turns out that no abortion rights for adult women can be supported by public reason, then it follows that no sex-selected abortions for adult women can be supported. But it does not follow that no abortions for adolescents can be supported; perhaps pregnancy or childrearing in the case of adolescents introduces considerations that would allow abortions for some adolescents. If it turns out that adult women have abortion rights in the first trimester, it seems to follow that juveniles do as well (perhaps with certain caveats), though it does not follow that either adults or juveniles have a right to sex-selected abortions. Where the right to have a non-sex-selected abortion typically supports women's civil equality, the practice of electing abortion only if one's fetus is female tends to undermine women's civil equality. Thus the reasoning might proceed. This method allows one to avoid more complicated situations without risking error in one's reasoning about the clearer or simpler situation. The case is different, however, when it comes to legitimating principles of justice.

Rawls holds that principles of justice are legitimate only if they are justifiable to all who are subject to them, independent of subjects' particular conceptions of the good. The idea of hypothetical consent and its attendant machinery (the original position, the veil of ignorance, and so on) is designed to establish the justifiability of Rawls's principles of justice. Now, if the problem of distributive justice for the non-cooperating is postponed, the design of the basic structure, as we saw above, is determined by representatives of only fully cooperating individuals. *Those* individuals exclusively occupy the original position. But if this is the case, then the design of the basic structure is not shown to be justifiable to non-cooperating members of society who will, nonetheless, be subject to many of its rules. Because

the non-cooperating are not represented in the original position, they do not 'have a say' in the design of basic social institutions. But there is no guarantee that the social arrangements shown to be acceptable for distributing resources to full cooperators (that is, shown to be justifiable to full cooperators) are acceptable for distributing resources to non-cooperators. There is no guarantee, in other words, that these arrangements would be justifiable to non-cooperators.

In the case of principles of justice, the aim is to determine what principles rational agents, characterized in a particular way, would *choose*. What they would choose, of course, depends on how they are characterized. One cannot infer what *some* rational agents would choose from what some *other* rational agents, characterized quite differently, would choose. So, we can see that one cannot infer from what principles rational agents fully capable of productive activity would choose what principles agents lacking that capability would choose.

Though he seems to regard the postponing of the issue of distributive justice for the non-cooperating as an innocuous application of a reasonable philosophical method (one that works in the case of abortion and public reason), Rawls's postponing of that issue may have unacceptable consequences. He may end up proposing terms of cooperation that are unsuitable from the point of view of the non-cooperating, that is, terms of cooperation that would not be chosen by the non-cooperating, so are not justifiable to the non-cooperating, and therefore are not legitimate.

Although the fully cooperating assumption generates the problem of legitimacy, it also, I argued above, installs in Rawls's theory the idea that the ability to cooperate is a morally relevant feature of persons. *This* idea, I contend, is supported by a moral commitment vital to contractarianism (and perhaps to liberalism generally.) We can call this commitment the 'anti-exploitation principle'. This principle, like the ideal of equality among persons, is, for the most part, assumed rather than argued for. It counts among those ideas taken to be uncontroversial theoretical starting points. The anti-exploitation principle is expressed in the following way by Gauthier:

the contractarian sees sociability as enriching human life; for him, it becomes a source of exploitation if it induces persons to acquiesce in institutions and practices that but for their fellow-feelings would be costly to them. Feminist thought has surely made this, perhaps the core form of human exploitation, clear to us. Thus the contractarian insists that a society could not command the willing allegiance of a rational person if, without appealing to her feelings for others, it afforded her no expectation of net benefit.⁶⁰

Several clarificatory points about the anti-exploitation principle are worth mentioning. First, the principle does not presuppose that individuals are completely devoid of fellow feelings. It merely presupposes that a person's other-regarding motives ought not to provide leverage for others, *who are capable of reciprocating*, to get one to support a practice at significant cost to one.⁶¹ Second, the anti-exploitation principle does not imply that individuals incapable of social

cooperation are not owed a share of the social product.⁶² It simply requires that *if* (and *not* only if) one is helping to produce the benefits of social cooperation, one is owed a certain portion of those benefits, all things being equal. (There is a morally significant difference between withholding benefits from one who cooperates and giving benefits to one who cannot.) Third, the anti-exploitation principle does not specify what portion of the social pie participants are owed, and contractarians disagree widely on this issue. We can think of the principles of distributive justice articulated by each contractarian theory as interpretations of the anti-exploitation principle. So, in the case of Rawls, the difference principle, for example, interprets the anti-exploitation principle as mandating institutional arrangements that tend to distribute material wealth equally among participants unless every participant, especially the least well off, will gain by allowing an unequal division.

The contractarian's commitment to the anti-exploitation principle, then, requires him to treat the ability to cooperate as morally relevant because the anti-exploitation principle regards the fact of one's social cooperation as bearing upon what share of the social product one is owed; a person contributing to a scheme of cooperation, is owed, *due to her contributing*, a certain portion of the cooperative surplus (all things being equal). Not to repay her appropriately for her contribution would be to exploit her (again, all things being equal). I conclude, then, that for a theoretically deep reason, Rawls has grounds for treating the ability to engage in social cooperation as a morally relevant feature of persons.

Critics have claimed that the limitations imposed on Rawls's theory by the fully cooperating assumption render his account inadequate. Eva Kittay, for example, says:

[The] consequences [of dependency] for social organization cannot be deferred until other traditional questions about the structure of society have been settled without distorting the character of a just social order. Dependency must be faced from the beginning of any project in egalitarian theory that hopes to include *all* persons within its scope.⁶³

Kittay proposes, then, that we abandon any assumptions about people's cooperative capabilities and devise a theory that gives one set of principles for all members of society, fully cooperating or not. Indeed, she tells us how we might modify and supplement Rawls's principles to this end.⁶⁴ I would like to suggest, in opposition to Kittay, that Rawls is, in fact, on the right track. That his theory contains both the Kantian and Humean themes discussed above reveals, I believe, something significant about our pre-theoretical notion of justice. In particular, it reveals that we associate both of the following ideas with justice: the idea that it is unfair for fellow cooperators to take advantage of one another and the idea that persons as such are owed equal concern and respect.

That both of these ideas are central to our notion of justice suggests that different principles of justice should apply to relations among the cooperating,

where worries about people being taken advantage of loom large, than those that apply to relations between the cooperating and the non-cooperating, *where worries about neglect and loss of standing for the non-cooperating loom large*. The anti-exploitation principle addresses the first type of worry: it tells us that schemes of cooperation should be governed by an ideal of reciprocity. The Kantian notion of the person applies generally: it tells us that social institutions should recognize everyone's right to have their basic needs met – a condition especially important for those who cannot stand in relations of reciprocity.

The worry, if we take up Kittay's suggestion that we discard the fully cooperating assumption and craft one set of principles for all members of society, cooperating or not, is that we will run into the legitimacy problem again. On account of the different moral concerns raised by dependency relations as compared to relations among coequals, which I identified above, we may not, on Kittay's approach, be able to find one set principles justifiable to all members of society. Principles that incorporate an ideal of reciprocity that meets the anti-exploitation condition will not be appropriate for, and may not be justifiable to, the non-cooperating; principles that do not contain such an ideal of reciprocity will not be appropriate for, and may not be justifiable to, the cooperating.⁶⁵

I propose, then, that we ignore Rawls's claim that the issue of justice for the non-cooperating is subsidiary and should be addressed only after the 'fundamental' issue of justice for the cooperating has been settled. Instead, we should treat Rawls's theory as a partial theory of justice, which offers principles for only those capable of cooperating. In other words, we should read Rawls's setting aside the question of what is owed to the non-cooperating as a way of treating that issue separately from the issue of what is owed to the cooperating.

Let me stress that if treating these two issues separately is the best approach, we must regard Rawls's theory as only part of the story – the part of the story that tells us what principles should structure schemes of cooperation if we want to ensure that no participant gains at the expense of others.⁶⁶ In addition, we still need to supply the part of the story that tells us what principles should govern relations of dependency if we want to ensure that everyone's needs are met in a way that preserves their self-respect. Surely once we supply *this* part of the story, we will have to reassess the other part for points of tension or redundancy. Perhaps we will have to make drastic changes. The advantages, though, are, first, that we can make sure that our accounts meet the liberal standard of legitimacy. That is, we can make sure that principles of justice for the non-cooperating can be justified to them and principles for the cooperating can be justified to *them*. Second, we can introduce a degree of subtlety in our account that might be absent on either Rawls's or Kittay's approach. We can capture, that is, the distinct moral issues that arise when it comes to relations among coequals as contrasted with relations of dependency.

Summary

I have argued that the Kantian and Humean themes in Rawls's theory, which might appear to be in tension, are in fact compatible. Each reflects certain of our considered moral judgments. The Humean claim that principles of justice are rational only if people are roughly equal in ability reflects the idea that justice involves reciprocity, *where people are capable of reciprocity*. The Kantian notion of persons as having a special standing reflects the idea that all persons are owed justice, *regardless of whether they are capable of reciprocity*. We should read Rawls's theory as offering principles suitable for, and justifiable to, those capable of social cooperation, and as leaving room for principles, yet to be specified, suitable for and justifiable to, those not so capable.⁶⁷

notes

I am indebted to Margaret Battin, Colin Bird, Sarah Buss, Leslie Francis, Daniel Greenwood, Elizabeth Harmon, Aaron James, Chandran Kukathas, Colin McCleod, Elijah Millgram, Jacqueline Taylor, and Jeremy Waldron for helpful comments on earlier versions of this article. I also owe thanks to the Tanner Humanities Center for providing me with a fellowship during which I did research for this article.

1. See, for example, Allen Buchanan, 'Justice as Reciprocity Versus Subject-Centered Justice', *Philosophy and Public Affairs* 19 (1990): 227–52.
2. Rawls seems to interpret Kant as grounding persons' moral status in their capacity for morality. See John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), pp. 251–7; John Rawls, *A Theory of Justice*, revised edn. (Cambridge, MA: Belknap Press, 1999), pp. 221–7 (henceforth referred to as the revised edn.); John Rawls, *Political Liberalism* (New York: Columbia University Press, 1993), pp. 18, 100–4.
3. Rawls, *A Theory of Justice*, pp. 126–30; revised edn., pp. 109–12. Exceptions include Martha C. Nussbaum and Brian Barry. See Martha C. Nussbaum, *Frontiers of Justice: Disability, Nationality, Species Membership* (Cambridge, MA: Harvard University Press, 2006), *passim*; Brian Barry, *Theories of Justice* (Berkeley: University of California Press, 1989), pp. 152–63, 179–83. See also Martha C. Nussbaum, 'The Future of Feminist Liberalism', in *Varieties of Feminist Liberalism*, edited by Amy R. Baehr (New York: Rowman and Littlefield, 2004), pp. 103–32. Hume's account of the circumstances of justice can be found in David Hume, *An Enquiry Concerning the Principles of Morals* (La Salle, IL: Open Court, 1966), Section III, Part I and David Hume, *A Treatise of Human Nature*, 2nd edn., edited by L.A. Selby-Bigge (Oxford: Clarendon Press, 1985), Book III, Part II, Section II. See also, David Gauthier, 'David Hume, Contractarian', *Philosophical Review* 88 (1979): 19–22.
4. See, for example, Buchanan, 'Justice as Reciprocity Versus Subject-Centered Justice', p. 246.
5. Stephen Darwall distinguishes between contractarianism and contractualism, and he classes Rawls as a contractualist. Contractarian theories view principles of justice as

- the outcome of a bargain made by de facto equals which makes each of them better off than they would be at the 'nonagreement point'. Contractualist theories regard principles as the outcome of the deliberations of moral equals occupying an idealized perspective that allows only morally relevant factors to influence the choice of principles. For reasons that will become clear, I regard Rawls's theory as not falling neatly into either one of these categories. See Stephen Darwall (editor), 'Introduction', in *Contractarianism/Contractualism* (Oxford: Blackwell, 2003), pp. 1–6.
6. By 'cooperative surplus', I simply mean the benefits and burdens created by social cooperation. These are not confined to, nor are they even primarily, material goods, on Rawls's view, as the list of primary social goods makes clear. See Rawls, *A Theory of Justice*, p. 62; revised edn., p. 54.
 7. See, for example, Thomas Hobbes, *Leviathan*, edited by C.B. MacPherson (London: Penguin, 1968) and David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press, 1986). See also, Christopher Morris, 'Justice, Reasons and Moral Standing', in *Rational Commitment and Social Justice: Essays for Gregory Kavka*, edited by Jules L. Coleman and Christopher Morris (Cambridge: Cambridge University Press, 1998), pp. 186–207.
 8. Barry, *Theories of Justice*, p. 13.
 9. This is Rawls's rendering of the equal ability condition. See Rawls, *A Theory of Justice*, pp. 126–7; revised edn., p. 110.
 10. Hume does not put the point in terms of *agreement* upon principles of justice. He claims that under conditions of extreme inequality, conventions of justice simply would not apply: 'the restraints of justice and property, being totally *useless*, would never have place in [a very] unequal . . . confederacy'. See Hume, *An Enquiry Concerning the Principles of Morals*, p. 23. The idea that the circumstances of justice constrain agreement, rather than indicate when conventions of justice 'have a place', is a contractarian modification of Hume. What unites contractarians with Hume, according to Gauthier, is that both view mutual expected advantage as the justice-making feature of principles of justice. This justice-making feature, however, does not explain why justice is, on Hume's view, a *virtue*. See, Gauthier, 'David Hume, Contractarian', p. 18. See also Barry, *Theories of Justice*, pp. 145–73. Thanks to an anonymous referee for this journal for forcing me to make this clarification.
 11. Hume and Rawls are concerned about the possibility that the very capable are able to dominate the less capable; but it is also the case that the very capable would be able to exclude the less capable from a system of justice. See Barry, *Theories of Justice*, p. 243.
 12. Hume, *A Treatise of Human Nature*, Book III, Part II, Section I. See also, Hobbes, *Leviathan*, Part 1, Ch. 15.
 13. Hume, *An Enquiry Concerning the Principles of Morals*, Section III, Part II.
 14. For a discussion of the sense in which justice is 'useless', on Hume's account, in conditions of extreme inequality, see Barry, *Theories of Justice*, pp. 160–3.
 15. See, for example, Gauthier, *Morals by Agreement*; J.M. Buchanan, *The Limits of Liberty: Between Anarchy and Leviathan* (Chicago, IL: University of Chicago Press, 1975) (cited in Gauthier, *Morals by Agreement*); J.F. Nash, 'The Bargaining Problem', *Econometrica* 18 (1950): 155–62 (cited in Barry, *Theories of Justice*). See also Barry, *Theories of Justice*, pp. 3–139.

16. Rawls, *A Theory of Justice*, p. 134; revised edn., p. 116.
17. *Ibid.*, pp. 134–5 n. 10; revised edn., p. 116 n. 10.
18. *Ibid.*
19. *Ibid.*, p. 136; revised edn., p. 117. See also Rawls, *Political Liberalism*, pp. 278–9.
20. Rawls, *A Theory of Justice*, p. 136; revised edn., p. 117.
21. *Ibid.*, p. 252; revised edn., p. 221.
22. *Ibid.*, p. 505; revised edn., p. 442.
23. *Ibid.*, pp. 142–3; revised edn., pp. 123–4.
24. *Ibid.*, p. 129; revised edn., pp. 111–2.
25. *Ibid.*, p. 137; revised edn., p. 118.
26. *Ibid.*, pp., 14, 33, 102, 179, 510; revised edn., pp. 13, 29, 88, 156, 447; Rawls, *Political Liberalism*, pp. 16–17. In *A Theory of Justice*, Rawls uses ‘mutual advantage’ and ‘reciprocity’ interchangeably. See, for example, Rawls, *A Theory of Justice*, p. 102; revised edn., p. 88. In *Political Liberalism*, he distinguishes between justice as mutual advantage and justice as reciprocity, identifying his own view as a variety of the latter. Justice as mutual advantage encompasses views, such as bargaining contractarianism, in which advantage is ‘judged with respect to each person’s present or expected future situation as things are’. See Rawls, *Political Liberalism*, p. 17. Justice as reciprocity encompasses views in which advantage is ‘judged with respect to an appropriate benchmark of equality’. See Rawls, *Political Liberalism*, p. 17. See also, Allan Gibbard, ‘Constructing Justice’, *Philosophy and Public Affairs* 20 (1991): 264–79 and Barry, *Theories of Justice*, pp. 241–54.
27. This is Robert Nozick’s way of describing the difference principle. See Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), p. 195.
28. Rawls, *A Theory of Justice*, pp. 80, 102; revised edn., pp. 69, 88. Rawls, *Political Liberalism*, pp. 16–17.
29. Later Rawls tells us that it is determined by the interests of persons in exercising their two moral powers. See John Rawls, *John Rawls: Collected Papers*, edited by Samuel Freeman (Cambridge, MA: Harvard University Press), p. 334.
30. Rawls also describes our natural talents as ‘undeserved’. I discuss the ideas of moral arbitrariness and desert below. See Rawls, *A Theory of Justice*, pp. 15, 72, 75, 101–2; revised edn., pp. 14, 63–5, 88–9.
31. This fact about the text does not entail, however, that Rawls has abandoned the equality of ability condition, for in *A Theory of Justice*, he says, after describing the circumstances of justice, ‘For simplicity, I often stress the condition of moderate scarcity...and that of mutual disinterest’. See *ibid.*, p. 128; revised edn., p. 110. Consequently, we do not know if Rawls’s failure to mention equality of ability in *Political Liberalism* is the result of striving for simplicity or whether he in fact changed his view. Thanks to Colin McCleod and Colin Bird for their helpful feedback on the issue of Rawls and the circumstances of justice.
32. Rawls, *Political Liberalism*, p. 20.
33. Marcia Homiak and S.A. Lloyd, in presentations of earlier versions of this article, suggested to me that Rawls sees the two moral powers as sufficient for being fully cooperating. At times, Rawls seems to hold this view. He says, for example, ‘The next step is to take the two moral powers as the necessary and sufficient condition for being counted a full and equal member of society in questions of political justice. Those who can take part in social cooperation over a complete life . . . are

- regarded as equal citizens.’ See *ibid.*, p. 302. These remarks suggest that people with physical impairments, no matter how disabling, or people with serious cognitive or mental disorders that do not compromise their moral powers are classed as fully cooperating. However, other passages suggest that Rawls does not regard persons with severe impairments as fully cooperating. He states ‘I have assumed . . . that while citizens do not have equal capacities, they do have, at least to the essential minimum, the moral, intellectual and physical capabilities that enable them to be fully cooperating members of society over a complete life.’ See *ibid.*, p. 183. See also Rawls, *A Theory of Justice*, p. 505; revised edn., p. 442 and Eva Kittay, *Love’s Labor: Essays on Women, Equality, and Dependency* (New York and London: Routledge, 1999), pp. 88–93; Christine Hartley, ‘Justice for All: Constructing an Inclusive Contractualism’, PhD thesis (University of Michigan, 2005); Christine Hartley, ‘An Inclusive Contractualism: Obligations to the Mentally Disabled’, in *Disability and Disadvantage: Re-examining Topics in Moral and Political Philosophy*, edited by Kimerly Brownlee and Adam Cureton (Oxford: Oxford University Press, forthcoming).
34. As Nussbaum says, ‘we cannot prevent all disability: for some impairments will continue to affect functioning even in a just social environment’. See Nussbaum, *Frontiers of Justice*, p. 424 n. 5.
35. As Nussbaum has noted, Rawls’s appeal to the Kantian notion of the person makes it very difficult, if not impossible, for him to include human nonpersons within the scope of justice. This his critics regard as a defect of his view – one that I will not attempt to rectify. So, my claim that the Kantian aspect of Rawls’s view allows him to include the non-cooperating within the scope of justice must be understood as the claim that the Kantian aspect of Rawls’s view allows him to include non-cooperating *persons* within the scope of justice. This might seem a large concession to Rawls’s critics. However, I believe that the problem of excluding the justice claims of human nonpersons plagues all varieties of liberalism, not just Rawls’s and not just contractarian views. Hence, the fact that Rawls cannot be saved from this problem is not as pressing an indictment of his view as his critics, many of whom are liberals, seem to think. Briefly, my view is this: the exclusion of human nonpersons from the scope of justice is a problem for liberalism because it is rooted in liberalism’s standard of justification for political arrangements. This standard, which is arguably at the heart of liberalism, says that legitimate political arrangements must be justifiable to all those who will be subject to them. But to be owed a justification, one must be capable of being justified to. To be so capable, one must have a certain level of cognitive ability. Indeed, we might say that to be so capable one must be minimally reasonable and rational in Rawls’s sense and hence must be, in Rawls’s sense, a person. But if this is so, liberalism’s very standard of justification is what pushes it to confine the scope of justice to (something like) Kantian persons. So, all liberals, including those who object to Rawls’s ideal of the person, are faced with this problem of exclusion, as they are bound, as Rawls is, to embrace a notion of the person compatible with the liberal standard of justification.
36. I concur with Nussbaum when she says that the fact that the group of people not capable of cooperation might represent a small percentage of the population is no excuse for ignoring issues of justice pertaining to this group. See Nussbaum, *Frontiers of Justice*, p. 100.

37. This seems clear from his accounts of the idea of reciprocity underlying his principles of justice. See Rawls, *Political Liberalism*, pp. 16, 300. Both Nussbaum and Kittay object to Rawls's account of reciprocity, which they regard as too narrow. See Kittay, *Love's Labor*, pp. 106–9 and Nussbaum, *Frontiers of Justice*, pp. 62–3, 107, 119, 128–30, 133–5.
38. While his account of what would count as a rational choice in the original position implies the inclusive notion of mutual advantage, Rawls's definition of the least well off *precludes* the inclusive notion. The least well off representative person is the one whose contribution to the cooperative surplus receives the least remuneration. This implies that *only* cooperators must gain from an inequality in order for the difference principle to be satisfied. Notice that adding the fully cooperating assumption makes consistent Rawls's account of rational choice in the original position and his definition of the least well off. For Rawls's definition of the least well off, see Rawls, *A Theory of Justice*, pp. 97–8; revised edn., pp. 82–3. For critical discussion, see Kittay, *Love's Labor*, pp. 110–2. The fully cooperating assumption also lets Rawls counter two criticisms of his view. First, Kenneth Arrow claims that the difference principle 'implies that any benefit, no matter how small, to the worst off member of society, will outweigh any loss to the better off individual provided it does not reduce the second below the level of the first'. It follows, he says, that 'there can easily exist medical procedures which serve to keep people barely alive but with little satisfaction and which are yet so expensive as to reduce the rest of the population to poverty'. See Kenneth J. Arrow, 'Some Ordinalist-Utilitarian Notes on Rawls's Theory of Justice', *Journal of Philosophy* 70 (1973): 251. It is easy to see how this concern about expensive medical care for people in persistent vegetative states might apply to other non-cooperating individuals with severe disabilities. The difference principle would require a society to devote such a large portion of its resources to bettering the situation of severely impaired individuals that other members of society would enjoy a low quality of life, even though they would be marginally better off than the disabled individuals to whom so many resources were being devoted. A second problem, presented by Amartya Sen, concerns Rawls's idea of what is to be distributed by his principles of justice. Sen argues that primary social goods are an inadequate measure of advantage, since people will vary greatly in their ability to use these resources to develop and exercise their capabilities. To put the point slightly differently, a 'resourcist' approach to distributive justice, which focuses exclusively upon goods, ignores differences in people's ability to convert good into advantage. See Amartya Sen, 'Equality of What?' in *Choice, Welfare and Measurement* (Cambridge, MA: MIT Press, 1982), pp. 355–66. This problem is especially acute when differences among citizens range from the severely mentally and physically disabled to the highly physically and mentally capable. The fully cooperating assumption addresses Arrow's concern by ensuring that the difference principle does not apply in contexts where it might impoverish a society by diverting large quantities of resources to people with severe disabilities. It addresses Sen's concern by ensuring that citizen's needs are sufficiently similar to warrant the use of an index of primary social goods as a basis for interpersonal comparison.
39. Hobbes, *Leviathan*, Part I, Ch. 13.
40. It is not clear how coalitions are possible in the state of nature since they would

require a level of trust (or at least enforcement) conspicuously absent in the state of nature, which absence generates the need for the sovereign in the first place.

41. Gauthier, *Morals by Agreement*, p. 18.
42. See Buchanan, 'Justice as Reciprocity Versus Subject-Centered Justice', p. 229 and Nussbaum, *Frontiers of Justice*, pp. 55–6. An anonymous referee has pointed out that Gauthier's view might also be challenged on empirical grounds, as it is by no means obvious, in many cases of people with disabilities, that the 'services required exceed any possible products'. The referee cited Stephen Hawking as a case in point.
43. Rawls, *Political Liberalism*, pp. 16, 300.
44. *Ibid.*, p. 79.
45. Barry holds that Rawls's introduction of the fully cooperating assumption creates a glaring inconsistency in his view. He says:

If we start with the basic ideas [that natural endowments, social circumstances, and good fortune are morally arbitrary] as the underpinnings of the difference principle, it must seem quite bizarre to claim that people who enjoy less than rude health (mental and physical), and especially those afflicted with a disease that gives rise to "unusual and costly medical requirements", create peculiar difficulties for a theory of justice. Such conditions are, we might think, the paradigm of the kind of undeserved misfortune whose translation into actual disadvantage Rawls describes as arbitrary from a moral point of view.

See Barry, *Theories of Justice*, pp. 244–5. Rawls's reference to 'unusual and costly medical requirements' is at Rawls, *John Rawls: Collected Papers*, p. 332.

46. In *Political Liberalism*, Rawls abandons the phrase 'morally arbitrary' in favor of terms such as 'irrelevant, politically speaking'. See Rawls, *Political Liberalism*, p. 79. See also *ibid.*, p. 24, where Rawls says that the fact that one occupies a certain social position or has certain native endowments does not count as a good reason for proposing, or expecting others to accept, principles of justice that favor that position or those native endowments. This, I think, makes it clear that by 'morally arbitrary' Rawls intended 'irrelevant from the point of view of justifying principles of distributive justice'. Thanks to an anonymous referee for this journal for asking me to clarify the notion of moral arbitrariness.
47. On my reading, Rawls uses the idea of desert in at least two ways in *A Theory of Justice*. First, he uses it in an ordinary sense: a thing is undeserved when it is inappropriate for one to take credit for that thing. Second, he uses it to mean 'moral desert', which is deservingness based upon one's virtue or intrinsic worth. The ordinary sense of 'desert' is, on my reading, expressed in all the passages where Rawls says that our natural endowments and social position are 'undeserved'. The idea of moral desert is discussed in Section 48. There Rawls argues against the notion of 'justice as happiness according to virtue'. His reason for claiming that moral desert is a poor ground for distributive justice is that the relevant virtues (such as diligence) are, like natural talents and social position, undeserved. He says:

A just scheme . . . answers to what men are entitled to; it satisfies their legitimate expectations as founded upon social institutions. But what they are entitled to is not proportional to nor dependent upon their intrinsic worth. The principles of

justice that regulate the basic structure . . . do not mention moral desert and there is no tendency for distributive shares to correspond to it.

This contention is borne out by the preceding account of common sense precepts (Section 47) [where the precepts 'to each according to his effort' and 'to each according to his contribution' are rejected]. For example, in determining wages a competitive economy gives weight to the precept of contribution, but as we have seen, the extent of one's contribution (estimated by one's marginal productivity) depends upon supply and demand. Surely one's moral worth does not vary according to how many offer similar skills, or happen to want what he can produce. No one supposes that when someone's abilities are less in demand or have deteriorated (as in the case of singers) his moral deservingness undergoes a similar shift. All of this is perfectly obvious and has long been agreed to. It simply reflects the fact noted before (Section 17) that it is one of the fixed points of our moral judgments that no one deserves his place in the distribution of natural assets any more than he deserves his initial starting place in society.

Thanks to an anonymous referee for this journal for pressing me to clarify the notion of desert.

48. Rawls, *A Theory of Justice*, p. 311; revised edn., p. 274.
49. *Ibid.*, p. 73; revised edn., p. 64. See also *ibid.*, p. 72; revised edn., p. 63.
50. *Ibid.*, p. 15; revised edn., pp. 13–14.
51. For a critical discussion of Rawls's claim that our natural talents are undeserved, see Michael Gorr, 'Rawls on Natural Inequality', *Philosophical Quarterly* 33 (1983): 1–18; Nozick, *Anarchy, State and Utopia*, pp. 213–31; George Sher, 'Effort, Ability and Personal Desert', *Philosophy and Public Affairs*, 8 (1979): 361–76; George Sher, *Desert* (Princeton, NJ: Princeton University Press, 1987), pp. 22–36; Michael A. Slote, 'Desert, Consent and Justice', *Philosophy and Public Affairs* 2 (1973): 323–47; Robert Young, 'Egalitarianism and Personal Desert', *Ethics* 102 (1992): 319–41; Alan Zaitchik, 'On Deserving to Deserve', *Philosophy and Public Affairs* 6 (1977): 370–88. See also Claudia Card, 'Individual Entitlements in Justice as Fairness', in *The Idea of a Political Liberalism: Essays on Rawls*, edited by Victoria Davion and Clark Wolf (Lanham, MD: Rowman and Littlefield, 2000): 176–89.
52. Rawls, *Political Liberalism*, pp. 15–16.
53. Rawls, *A Theory of Justice*, pp. 101–2; revised edn., pp. 88–9.
54. Sometimes Rawls suggests that a question is fundamental if it is long-standing in the history of philosophy. For instance, he says, 'that this question [of the appropriate conception of justice for the fully cooperating] is indeed fundamental is shown by the fact that it has been the focus of the liberal critique of aristocracy in the seventeenth and eighteenth centuries, of the socialist critique of liberal constitutional democracy in the nineteenth and twentieth centuries, and the conflict between liberalism and conservatism at the present time'. See Rawls, *Political Liberalism*, p. 22.
55. Rawls, *John Rawls: Collected Papers*, p. 332.
56. Rawls, *Political Liberalism*, pp. 20–1.
57. *Ibid.*, p. 243.
58. It is also plausible to read Rawls as placing the question of justice for the non-

cooperating in the category of nonideal theory. See Rawls, *A Theory of Justice*, pp. 8–9, 244–8; revised edn., pp. 8, 215–8. Christine Korsgaard takes this position at Christine Korsgaard, ‘The Right to Lie: Kant on Dealing with Evil’, in *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), p. 147. This reading, however, is in tension with Rawls’s claim that the question of justice for the non-cooperating is a problem of extension; presumably, in extending justice as fairness to cover more difficult cases one is still operating in the realm of ideal theory.

59. I am drawing here on the account in S.A. Lloyd, ‘Toward a Liberal Theory of Sexual Equality’, in *Varieties of Feminist Liberalism*, edited by Amy R. Baehr (New York: Rowman and Littlefield, 2004), pp. 63–84.
60. Gauthier, *Morals by Agreement*, p. 11. See also Samuel Freeman, ‘Frontiers of Justice: The Capabilities Approach vs. Contractarianism’, *Texas Law Review* 85 (2006): 402–3. He says of Rawls’s contractarianism:

Rawls contends that each person’s rational advantage is integral to cooperation, not to emphasize that each person must benefit from every other person in order to have reason to cooperate with or benefit them. It is rather . . . when people live in a society where their rational good is not in any significant way fairly advanced, then they are being coerced, or exploited, or manipulated, or unfairly taken advantage of in some other way.
61. See Jean Hampton, ‘Feminist Contractarianism’, in *A Mind of One’s Own: Feminist Essays on Reason and Objectivity*, edited by Louise M. Antony and Charlotte Witt (Oxford: Westview Press, 1993), p. 239.
62. Although, as we saw above, this is an implication of Gauthier’s view.
63. Kittay, *Love’s Labor*, p. 77.
64. *Ibid.*, Ch. 4.
65. I take Nussbaum’s capabilities approach to count as a theory whose distributive ideals do not contain an ideal of reciprocity.
66. Rawls says the difference principle expresses this ideal of reciprocity: ‘the more advantaged are not to be better off at any point to the detriment of the less well off’. See John Rawls, *Justice as Fairness: A Restatement*, edited by Erin Kelly (Cambridge, MA: Harvard University Press), p. 124.
67. Using a Rawlsian framework, I offer an account of distributive justice for the non-cooperating in Cynthia A. Stark, ‘How to Include the Severely Disabled in a Contractarian Theory of Justice’, *Journal of Political Philosophy* 15 (2007): 125–47. See also Henry Richardson, ‘Rawlsian Social-Contract Theory and the Severely Disabled’, *Journal of Ethics* 10 (2006): 419–62 and Harry Brighouse, ‘Can Justice as Fairness Accommodate the Disabled?’ *Social Theory and Practice: An International and Interdisciplinary Journal of Social Philosophy* 27 (2001): 537–60.