

In Defense of Liberal Equality

M. E. Newhouse
University of Surrey

Abstract: In *A Theory of Justice*, Rawls concludes that individuals in the original position would choose to adopt a system of democratic equality governed by his two principles of justice. However, Rawls mistakenly defines the possibility space within which individuals in his original position must make their choice. An alternative account of the possibility space created by Rawls's original position reveals that a system of liberal equality, according to which distributive shares would be determined by market processes, would be preferred by risk-averse individuals. However, such individuals would guard against the erosion of the social bases of self-respect by including a social safety net among the basic equal liberties secured by Rawls's first principle of justice.

Key words: Rawls, *A Theory of Justice*, liberal equality, the difference principle, rational choice.

Early in *A Theory of Justice*, John Rawls (1971, 1999) considers and rejects an interpretation of his two principles of justice called “liberal equality” in favor of an alternative interpretation that he styles “democratic equality” – an otherwise similar system that substitutes his famous “difference principle” of distributive justice for distribution according to market processes. Rawls has been criticized for supposedly claiming that a person's relationship to her own attributes is morally arbitrary, and that the attributes of individuals are therefore properly considered a social resource. I argue that this is an uncharitable interpretation of the text. It is an individual's relationship to the distribution of natural attributes in her society that Rawls considers arbitrary, not her relationship to her own natural attributes. I then show that what I will call the “libertarian objection” nonetheless has some purchase as a criticism of Rawls's analysis, which misapprehends the possibility space created by the original position. I offer an alternative account of the uncertainties faced by an individual in the original position, called the Unknown Society Model. The Unknown Society Model indicates that risk-averse individuals in Rawls's original position will unanimously prefer distribution according to market processes to the difference principle and will therefore adopt a system of liberal equality rather than a system of democratic equality. Nonetheless, the Unknown Society Model implies that individuals in the original position are also likely to include a social safety net on their list of basic liberties in order to preserve the social bases of self-respect.

I. LIBERAL EQUALITY VS. THE DIFFERENCE PRINCIPLE

John Rawls's theory of justice is anchored by two fundamental principles, which apply to the “basic structure”¹ of society:

[1] The basic structure of a society is the scheme by which “the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation” (Rawls 1999, 6).

First: each person is to have an equal right to the most extensive scheme of equal basic liberties compatible with a similar scheme of liberties for others. Second: social and economic inequalities are to be arranged so that they are both (a) reasonably expected to be to everyone's advantage, and (b) attached to positions and offices open to all. (Rawls 1999, 53)

Rawls's first principle secures an equal right to a set of basic liberties for every member of society, and these rights are given priority over all other considerations of justice (1999, 55). Once the equal basic liberties are secure, the institutions comprising the basic structure should seek to implement the second principle. Rawls suggests that both clauses of his second principle are ambiguous, and he disambiguates them in a short discussion near the beginning of his analysis (1999, 57).

The principle that positions and offices should be "open to all" could be read simply to prohibit employment discrimination on the basis of irrelevant characteristics, an interpretation that Rawls calls "careers open to talents" (1999, 57). Rawls certainly endorses the requirement of non-discrimination, but he concludes that his second principle of justice must go further to ensure that individuals have genuine, as opposed to merely formal, opportunities to pursue social positions for which their natural attributes make them well-suited. He therefore settles on a broader interpretation of this part of his second principle, called "fair equality of opportunity" (Rawls 1999, 63). Fair equality of opportunity requires, in addition to non-discrimination, the provision of talent development opportunities to less privileged members of society that enable them to compete on equal terms for offices and positions in which they have the natural capacity to succeed.²

Rawls then considers two possible interpretations of his requirement that social and economic inequalities will be "reasonably expected to be to everyone's advantage" (1999, 53). One interpretation of this clause is what Rawls calls "efficiency", according to which the principle is satisfied by a competitive market economy that maximizes the total wealth in society, and in which an individual's distributive share of wealth and income will correspond to the relative productive value of her natural attributes (as developed in accordance with fair equality of opportunity), subject to her "willingness to use them" in the workplace (1999, 62-63). A system combining equal basic liberties, fair equality of opportunity, and distribution according to market processes is called "liberal equality" (Rawls 1999, 57).

Rawls characterizes the system of liberal equality as an "unstable" stopping place in his search for the best principles of justice, because "there is no more reason to permit the distribution of income and wealth to be settled by the distribution of natural assets than by historical and social fortune" (1999, 64). Once we see the distribution of natural assets in society as "morally arbitrary", Rawls thinks we will be moved to prefer an alternative interpretation of this clause according to which the principle is satisfied by institutional

2] Rawls explains, "those who are at the same level of talent and ability, and have the same willingness to use them, should have the same prospects of success regardless of their initial place in the social system" (1999, 63).

arrangements that maximize the wealth and income available to “the representative unskilled worker” whom he supposes will be among the least well-endowed “full and active participants” in an otherwise just society (Rawls 1999, 68, 84). Rawls refers to this interpretation as the “difference principle”, and he refers to the system combining equal basic liberty, fair equality of opportunity, and the difference principle as “democratic equality” (1999, 57).

II. THE LIBERTARIAN OBJECTION

Rawls’s treatment of what he refers to as “natural assets” or “natural attributes” (1999, 11, 87, 172) – “intelligence, strength, and the like” (1999, 11) – in his design and analysis of the original position is perhaps the most enduringly controversial aspect of his theory. Some critics have objected to the difference principle on the basis that Rawls has inappropriately ignored the moral significance of a person’s relationship to her own natural attributes. Such critics complain that an individual’s natural attributes are constitutive of her and therefore cannot be arbitrary from a moral point of view.³

Some language in *A Theory of Justice* can be interpreted in ways that seem to bolster this case. For example, Rawls refers to “the outcome of natural chance” and to “the outcome of the natural lottery” as morally arbitrary (1999, 11, 64). If Rawls meant to say that a person’s relationship to her own attributes is a matter of chance, then perhaps his theory presupposes an implausible metaphysical account of the nature of persons. At times, Rawls also uses language that could be interpreted to mean that the natural attributes of persons should be considered a common resource. For example, Rawls writes that in a society governed by his principles of justice, the less naturally well-endowed “view the greater abilities of others as a social asset to be used for the common advantage” (1999, 92).

However, there is at least as much textual evidence to suggest that Rawls held neither an implausible metaphysical view of the nature of persons nor so controversial a view of the moral irrelevance of an individual’s relationship to her own natural characteristics as some have suggested. Rawls’s words in context more often indicate that *an individual’s place in the distribution of natural attributes within the population* can be regarded as arbitrary. For example, he writes that “we do not deserve our place in the distribution of native endowments”, and he wishes to design a system in which “no one gains or loses from his arbitrary place in the distribution of natural assets” (Rawls 1999, 87-89). By contrast, natural attributes themselves, as part and parcel of a person, can be considered no one else’s: “To be sure, the more advantaged have a right to their natural assets, as does everyone else; this right is covered [...] under the basic liberty protecting the integrity of the person.” (Rawls 1999, 89). Rawls also writes, “The difference principle represents, in effect, an agreement to regard *the distribution of natural talents* as in some respects a

3] See for example (Nozick 1974, 215-16).

common asset and to share in the greater social and economic benefits *made possible by the complementarities of this distribution.*" (1999, 87).⁴

If this latter reading of Rawls is the correct one, then the purpose of the veil of ignorance is to eliminate the influence of an individual's arbitrary place in the social distribution of natural attributes on the selection of principles of justice. Shielding an individual from knowledge of her own attributes (which Rawls agrees are determinate⁵) is simply his means of accomplishing this goal: if a person does not know what her natural attributes are, then she cannot know what place she occupies in the larger social distribution of those attributes.

But if this is Rawls's goal, he makes a key mistake in his analysis of the choice faced in the original position, and this mistake gives the libertarian objection some purchase as a criticism of his results despite its inaccuracy as a criticism of his premises. When Rawls argues that a risk-averse person in the original position will choose principles of justice that maximize the wealth and income of the representative least-skilled worker in society, he inadvertently holds the social distribution of natural attributes constant and treats the individual chooser's attributes as variable. The difference principle is the risk-averse choice if the total wealth in society (and therefore the distribution of attributes that generate this wealth) is fixed, but the individual selecting principles of justice does not know what her own attributes are. In this situation, it is risk-averse to maximize the smallest slice of the fixed pie in case one's own attributes place one at the lowest "level of talent and ability" in this fixed distribution (Rawls 1999, 63). The possibility space presupposed by this analysis of the choice faced in the original position is inconsistent with Rawls's stipulation that an individual is supposed to know that she *has* determinate natural attributes, even if she does not know what they are (1999, 127, 152).

An individual in the original position does not need to know what her attributes are, and she must not know where these attributes place her in the larger social distribution of such attributes, since this is the arbitrary relationship we wish to eliminate from consideration. However, she is supposed to know that she *has* determinate natural attributes. Other members of her society have determinate attributes as well, but she has no way of knowing whether the attributes of others will translate into higher or lower levels of talent and ability in society relative to her own. A person in the original position therefore ought to experience the social distribution of talent and ability around her as the variable feature in the analysis, which will, after she chooses principles of justice, be settled by a "natural lottery". I will propose an "Unknown Society Model" that more accurately reflects the possibility space that remains after the original position eliminates the arbitrary influence of an individual's place in the social distribution of natural assets from consideration when principles of justice are chosen.

4] Emphasis added.

5] Rawls writes, "we must keep in mind that the parties in the original position are theoretically defined individuals." (1999, 127)

Like Rawls, I assume that individuals in the original position are non-enviously seeking to maximize the value of an index of social primary goods including wealth, income, and positions of authority and responsibility that they can expect to receive, and that positions of authority and responsibility are sufficiently correlated with wealth and income that they need not be considered separately (Rawls 1999, 83). Also like Rawls, I assume that given the conditions of deep uncertainty and high stakes that prevail in the original position, individuals will appropriately have a “high level of risk aversion” when they choose principles of justice (Rawls 1999, 144), leading them to “adopt the alternative the worst outcome of which is superior to the worst outcomes of the others” (Rawls 1999, 133).

III. THE UNKNOWN SOCIETY MODEL

The Unknown Society Model does not change the features of the original position itself. Rather, it is an alternative account of the possibility space created by the original position. It treats an individual’s natural endowments as an unknown constant rather than as a variable, reflecting the fact that individuals in the initial situation are “determinate persons” – they do not know what their specific attributes are, but they do know that they have specific attributes (Rawls 1999, 151-52). Second, the social distribution of “level[s] of talent and ability” (by which I understand Rawls to be referring to the market value of the constellation of natural traits that are useful for generating income) of members of society is treated as variable, since an individual in the original position has no idea what levels of talent and ability others in her society will have relative to herself (Rawls 1999, 63).

In the Unknown Society Model, a randomly selected individual – whom I will call Xenia from now on – must choose principles of justice knowing that her society might be comprised entirely of individuals with lower levels of talent and ability than hers, or higher. Alternatively, her level of talent and ability might fall somewhere in the middle of the distribution. Xenia also has no idea how large her society is, nor how wide the distribution of talent and ability might be. Her society might contain huge differences in levels of talent and ability, or everyone might be endowed with attributes that enable them to be roughly equally productive.⁶ Therefore, Xenia cannot choose principles of justice based on her arbitrary place in the social distribution of natural assets.

Xenia chooses principles of justice by considering the risks of the “natural lottery” that she faces: various places she might wind up occupying in the unknown social distribution of natural assets. To demonstrate the robust superiority of liberal equality to the difference principle in the face of this uncertainty, I will model Xenia’s relative expectations in terms of wealth and income in three possible worlds in which she occupies three very different places in this social distribution:

⁶ Rawls envisioned a possible future society in which its members enjoy “the greatest equal talent”, although he does not suggest that this is very likely (1999, 93).

Spartanville: In Spartanville, every other fully-participating member of society has a lower level of talent and ability than Xenia, who therefore finds herself at the very top of the social distribution of such attributes after she chooses principles of justice.

Equalitania: In Equalitania, all fully-participating members of society have equal “levels” of talent and ability (in the Rawlsian sense of market value), although those attributes may differ qualitatively.

Talentopia: In Talentopia, every other fully-participating member of society has a higher level of talent and ability than Xenia, who therefore finds herself at the very bottom of the social distribution of such attributes after she makes her decision.

Like Rawls, I assume that a system of social cooperation yields more wealth and income than the sum of what every participant might generate on her own (i.e. there are gains from trade within a society). I also adopt Rawls’s assumption that the system of liberal equality will generate shares of wealth and income that are roughly proportional to individuals’ relative levels of talent and ability in the social distribution thereof.⁷ Additionally, I suppose along with Rawls that a redistributive policy “pushed beyond a certain point weakens incentives and thereby lowers production” although there is no need to decide where, empirically, that point might be (Rawls 1999, 142).

Finally, in accordance with Rawls’s stipulation that “there is no objection to resting the choice of first principles upon the general facts of economics and psychology”, I account for the possibility of what economists call “human capital externalities” (Rawls 1999, 137). A human capital externality is the effect that a change in one person’s level of talent and ability has on the (inflation-adjusted) wealth and income of other members of her society.⁸ Economists attempting to measure human capital externalities have arrived at estimates that range from statistically insignificant to strongly positive.⁹ Xenia will deal with this uncertainty by considering her relative expectations in different possible states of the world under either assumption, and she will choose the principles of justice that maximize her expectations in terms of income and wealth in the worst possible state of the world.

In a system of liberal equality, Xenia will expect to have roughly equal amounts of wealth and income in Spartanville, Equalitania, and Talentopia if human capital externalities turn out to be insignificant. If they turn out to be significant, she will have more wealth and income in Equalitania than she will in Spartanville, and she will have more wealth and income in Talentopia than she will have anywhere else. If Xenia chooses a system of democratic equality instead of a system of liberal equality, the difference principle will increase what Xenia will expect to have in Talentopia (relative

7] That is, shares will be equal to those Rawls describes in a system of natural liberty except that they will no longer reflect inherited wealth and income due to the implementation of the principle of fair equality of opportunity (1999, 62-64).

8] Daron Acemoglu explains, “Human capital externalities arise when the investment of an individual in his skills creates benefits for other agents in the economy.” (1996, 779).

9] See for example Lucas Jr (1988), Moretti (2004), and Thönnessen et al (2013).

to her expectation under liberal equality), will decrease what Xenia will expect to have in Spartanville, and will leave her expectation in Equalitania unchanged. See Figure 1.

Figure 1: Xenia’s Expectations in Terms of Wealth and Income

	Talentopia	Equalitania	Spartanville
Liberal Equality (no human capital externalities)	a	a	a
Liberal Equality (positive human capital externalities)	a+b+c+d	a+b+c	a+b
Democratic Equality (no human capital externalities)	a+f	a	a-e
Democratic Equality (positive human capital externalities)	a+b+c+d+h	a+b+c	a+b-g

Key to Figure 1 (all values are positive)

a	Xenia’s expectation under liberal equality with no human capital externalities
b	the value to Xenia of positive human capital externalities in Spartanville
c	the additional value to Xenia of human capital externalities in Equalitania (compared to Spartanville)
d	the additional value to Xenia of positive human capital externalities in Talentopia (compared to Equalitania)
e	the value of the reduction in Xenia’s expectation in Spartanville with no human capital externalities under democratic equality (relative to liberal equality)
f	the value of the increase in Xenia’s expectation in Talentopia with no human capital externalities under democratic equality (relative to liberal equality)
g	the value of the reduction in Xenia’s expectation in Spartanville with positive human capital externalities under democratic equality (relative to liberal equality)
h	the value of the increase in Xenia’s expectation in Talentopia with no human capital externalities under democratic equality (relative to liberal equality)

For Xenia, the worst possible state of the world is Spartanville with no human capital externalities. To paraphrase Rawls, this is the world in which her enemy would place her (1999, 133). In this worst possible situation, Xenia will have more wealth and income if she has chosen to adopt the system of liberal equality (an expectation of a) than she will in the system of democratic equality (an expectation of a-e). By contrast, Talentopia with human capital externalities is the best of all possible worlds for Xenia. In this system, she

will fare even better under democratic equality ($a+b+c+d+h$) than she will under liberal equality ($a+b+c+d$).

In other words, when Xenia regards her own attributes as constant, though unknown, and the distribution around her as variable, the difference principle is a risk-loving principle – it further boosts Xenia’s already good expectations in Talentopia with human capital externalities, and it worsens her already poor expectations if she loses the “natural lottery” and is consigned to Spartanville without human capital externalities. Because Xenia is risk averse, she will choose to adopt the system of liberal equality and its principle of distribution according to market processes instead.¹⁰

IV. THE SOCIAL BASES OF SELF-RESPECT

I have shown that when the possibility space created by the original position is properly defined, individuals in the original position will maximize the share of wealth and income they can expect to receive in the worst-case scenario by choosing the system of liberal equality. However, this outcome raises a possibility that Rawls did not have to address given that his own analysis yielded a system of democratic equality instead: the possibility that very high levels of income inequality might effectively deprive individuals near the bottom of the income distribution of the social bases of self-respect. While fair equality of opportunity would certainly limit the inequalities that arise in a system of liberal equality, it cannot entirely rule out the possibility of an extreme stratification of income and wealth among mature citizens that could potentially undermine the self-respect of the relatively poor.

Rawls refers to self-respect as “perhaps the most important primary good” because “without it nothing may seem worth doing” (1999, 386). The social bases of self-respect include the resources necessary to carry out a rational plan of life in accordance with one’s conception of the good, and the ability to find a community inside the larger society “within which the activities that are rational for [the individual] are publicly affirmed by others” (Rawls 1999, 387). Rawls gestures at the lexical priority of self-respect over ordinary distributive concerns when he suggests that its importance “limits the forms of hierarchy and the degrees of inequality that justice permits” (1999, 92). I therefore believe that individuals in the original position would include a social safety net adequate to safeguard the social bases of self-respect among the basic equal liberties protected by Rawls’s first principle of justice.

10] To be sure, different natural attributes will have different market values in different societies. For example, a person whose talents would make them a particularly good hair stylist, flower arranger, or cake decorator might find her skills relatively more valuable in a rich, post-industrial society, while a person whose talents lend themselves to physical combat and wilderness survival may find themselves more marketable, on average at least, in a poorer, pre-industrial society. But these uncertainties are bi-directional, exist at all points in the distribution of natural assets, and exist to an equal degree in Spartanville and Talentopia, so they cannot change the outcome of the analysis.

V. CONCLUSION

Rawls himself never evaluates liberal equality from the perspective of the original position, having considered and rejected this interpretation of his principles before undertaking his primary analysis. Given his understanding of the possibility space created by the features of the original position, he was right not to waste any time on the analysis: if the proper methodology is to compare the expectation in terms of wealth and income of the least skilled worker in a given society under liberal equality and under democratic equality, the difference principle yields an expectation equal to or better than market processes by definition, so democratic equality will be preferred.¹¹

However, if I am correct in my account of the possibility space created by the original position, then Rawls rejected liberal equality prematurely. If the correct principles of justice are those that free and equal individuals would choose in a situation cleansed of the arbitrary influence of individuals' relative positions in larger social distributions of income and wealth as well as natural assets, then liberal equality, cushioned by the inclusion of a social safety net among the basic equal liberties, appears to be the most just available system.

Of course, those currently committed to democratic equality are not rationally required to embrace my proposed version of liberal equality even if my analysis is correct. In the process of reflective equilibrium, there is no privileged starting place. A person confronted with a discrepancy between the widely shared and highly general principles incorporated into the initial situation and her considered convictions about the requirements of justice is free to revise either the general principles or the considered convictions (Rawls 1999, 17-19). Rawlsians convinced by my analysis but likewise convinced of the unacceptability of liberal equality might prefer to respond to my critique by revising the general principles that determine the features of the original position. For example, some might decide on reflection that a system of social cooperation should literally "draw forth the willing cooperation of everyone taking part in it" rather than merely constructively doing so (Rawls 1999, 13). In that case, it might follow that a universal basic income should be adopted in preference to the more modest social safety net that I have suggested for inclusion among the basic equal liberties.

m.newhouse@surrey.ac.uk

[11] Even if it happens to be the case that market processes generate the best expectation possible for the least-skilled worker in a given society, then adopting the difference principle will result in a policy of distribution via market processes. The least-skilled worker in a given society therefore might do better, and can by definition do no worse, under democratic equality than she would under liberal equality.

REFERENCES

- Acemoglu, Daron. 1996. A microfoundation for social increasing returns in human capital accumulation. *The Quarterly Journal of Economics* 111(3): 779-804.
- Lucas Jr, Robert E. 1988. On the mechanics of economic development. *Journal of Monetary Economics* 22, (1): 3-42.
- Moretti, Enrico. 2004. Workers' education, spillovers, and productivity: evidence from plant-level production functions. *American Economic Review* 94 (3): 656-90.
- Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
- Rawls, John. 2009. *A Theory of Justice: Revised Edition*. Harvard University Press.
- Thönnessen, Rasmus, and Erich Gundlach. 2013. The size of human capital externalities: cross-country evidence. *Public Choice* 157 (3-4): 671-89.