

## **Non-social human beings in the original position**

*Author:* Terence Rajivan Edward, University of Manchester.

*Abstract.* This paper argues that Rawls must commit himself to non-social human beings to defend his original position procedure.

In this paper I contribute to answering the question, is the philosopher John Rawls committed to the possibility of human beings who are not social beings? The answer depends on what exactly is meant by ‘not social beings.’ Perhaps it also depends on which of Rawls’s texts one focuses on. I argue that in order to defend proposals made in his classic book *A Theory of Justice*, Rawls needs to take on a certain commitment, and it makes sense to describe this as a commitment to non-social human beings. An argument has already been made for a similar conclusion (Sandel 1984: 86-87), an argument which has been contested (Edwin Baker 1985: 900), but my argument below is quite different.

In *A Theory of Justice*, Rawls asserts that a society should be just (1999: 3-4). Here we can take ‘just’ as another word for fair. Rawls also presents a method for determining which principles should be implemented by the major social institutions of a society, in order for the society to be just (1999: 118). The method is to imagine a certain situation, called the original position. We are to imagine some individuals coming together to form a society and deciding on an agreement. The agreement concerns how the major institutions of that society will distribute rights, duties and the benefits of social cooperation. But the individuals we are to imagine are very different from us in certain respects. Each individual does not know various features of themselves, features which if known might lead them to try to make an agreement that favours individuals with those particular features. One feature that individuals in the original position do not know is their talents. If an individual knew that they have a

certain talent, for writing with perfect grammar say, they might promote an agreement which favours individuals with this talent, for instance one which says that only individuals with this talent can be members of the government. Apart from talents, there are other features of themselves that individuals in the original position do not know: their sex, their race, their income, and more. Rawls describes them as being behind a veil of ignorance. By considering what such individuals would agree to, we can realize what would be a just agreement for how the major social institutions should distribute rights, duties and the benefits of social cooperation. The agreement would be just because these individuals lack knowledge of features that might lead them to be biased.

This in broad outline is Rawls's method. There are a lot of details and qualifications that can be added to the outline. I shall focus on one particular detail. Each individual in the original position does not know their views about what would be a good life for them (1999: 118). For example, if one individual believes that a pudding after dinner is essential to a good life for them, they do not know that. If they knew that, they might try to promote an agreement which ensures that the society is always well-stocked with pudding ingredients.

From this example, it may seem that Rawls is right to imagine individuals in the original position as lacking knowledge about their views regarding what would be a good life for them. But are there not some very general views regarding what would be a good life that are shared by all individuals, or at least all adult individuals without severe psychological disorders? The individuals we are to imagine in the original position are adult human individuals.<sup>1</sup> If you gave a human being a choice between living completely alone, cut off from all other human beings, and living as a member of society, presumably a sane adult would prefer to live in a society, so long as they are not subject to especially brutal treatment. And they would prefer this partly, maybe even wholly, because they want communication or

---

<sup>1</sup> In this paper, I shall use 'person' to mean human being, although I do think there can be non-human persons.

interaction with other human beings. The kind and degree of communication or interaction that one person wants and that another person wants may not be the same. But presumably they would all want some kind of communication or interaction with other human beings and would regard this as essential to a good life for them. (I am not sure whether there can be non-interactive communication or non-communicative interaction, so I have written of ‘communication or interaction.’)

Contemporary political philosophers use the term ‘perfectionists’ to label those who allow for a view of the good life to be involved when deciding government policies. Since some such views are quite moderate and since the term ‘perfectionist’ sounds as if it refers to an extreme type of person, this term does not seem very suitable. But this is the terminology in place and I shall work with it. The proposal above is a proposal for a perfectionist version of the original position. It allows for individuals in that position to appeal to the view that communication or interaction with other human beings is essential to a good life.

This perfectionist version may well have different outcomes to Rawls’s version. If individuals in the original position know that they are forming a society partly, or wholly, because of desires for communication or interaction, it makes sense for them to agree to policies that help protect against a life in which these things are absent or too little. The agreement Rawls recommends provides no protection against this.

How can Rawls defend his version of the original position from a perfectionist variation<sup>2</sup> on it? Here is the simplest answer I can think of: “Given that there are hermits, it is possible that there are some people who value being part of society but only because of certain material benefits it brings, such as security and the opportunities to obtain food. They have no special interest in communication or interaction. If they could somehow get the same benefits in another way, then they may well choose to leave society, but at present there is no

---

<sup>2</sup> There is another perfectionist version, which only allows for the view that interaction or communication is essential for a good life. It omits the ‘with other human beings’ component. For some individuals that component may be essential, but this perfectionist version does not assume it to be essential for all.

other way to get these benefits. If we allow individuals in the original position to work with the view that communication or interaction is essential to a good life for them, we would be biased against such people. We should not be biased against them, so we should work with my version of the original position, which does not allow this view.”

I think Rawlsians will prefer a slightly moderated version of this answer, according to which it is plausible that there could be human beings who fit the description above, sufficiently plausible that we have to avoid bias against such individuals, even if we are not fully sure that they can exist.

There are two objections that I can anticipate towards this defence. First, it may be objected that the kind of person described above is impossible. But Rawls can say that this judgement is based on a very controversial claim about human nature, the kind of claim that it is not appropriate to rely on when determining which principles should be implemented by society. We must derive these principles from appropriate foundations, foundations that are sound and which citizens can be expected to accept, and such a controversial claim is not an appropriate foundation.<sup>3</sup>

The second objection is that a person who does not value communication or interaction as essential for a good life is suffering from an extreme psychological disorder, and so we should discount their view about what would be a good life for them when designing the original position. This does not mean we should ignore the view in all contexts, but we can deal with it at a later point, rather than when working out which principles the major social institutions should implement.

Much the same response can be given to the second objection as was given to the first. It is very controversial to claim that a person who values being part of society because of certain material benefits it brings, but does not value communication or interaction as

---

<sup>3</sup> I do not wish to claim that the debate ends here. Perhaps the objector can challenge this response. See the next footnote.

essential to a good life for them, must be suffering from an extreme psychological disorder. This is not a claim that we can rely on when trying to determine which principles the major social institutions of society should implement.<sup>4</sup>

Before concluding this paper, I want to consider some suggestive material from Rawls. At one point in *A Theory of Justice*, Rawls imagines a person whose nature means that he does not regard communication or interaction as essential to a good life for him. Rawls imagines such a person, but without explicitly committing himself to this person being genuinely possible. He imagines someone who solves mathematical problems for a living, but whose sole pleasure is to count grass, grass which is in neatly-shaped areas (1999: 379). This person solves mathematical problems to fund the pursuit of his grass-counting pleasure. Rawls makes some suggestive remarks about this case:

Perhaps he is peculiarly neurotic and in early life acquired an aversion to human fellowship, and so he counts blades of grass to avoid having to deal with other people. But if we allow that his nature is to enjoy this activity and not to enjoy any other, and that there is no feasible way to alter his condition, then surely a rational plan for him will center around this activity. (1999: 379-380)

Some people might describe this person as mad. Rawls describes the person he imagines as ‘peculiarly neurotic,’ rather than using a term to indicate problems of a more severe kind. And he seems prepared to treat them much like any other citizen of society, if they are prepared to follow its rules and accept the importance of society being just.

There seems to be a gap between how the second objector and Rawls respond to the same kind of case. The gap involves the psychological theses they are attracted to: this person is suffering from an extreme mental disorder, versus this person is neurotic. It also involves

---

<sup>4</sup> But if Rawlsians are prepared to defer to expert opinion on man-made climate change, then why not do so on this issue as well?

the relevance for political philosophy: we do not have to take into account this person's views when working out principles for the major social institutions, versus we have to take into account this person as well. Assuming an acceptance of the original position, this gap will lead to other divergences: a perfectionist original position versus a non-perfectionist one; principles that aim to ensure that everyone gets communication or interaction versus principles that do not; perhaps a government that aims to reduce the levels of social isolation within some parts of the population, such as senior citizens, versus a government that does not. One might compare this to two roads, which begin close together, but gradually move further apart.

Even if Rawls did not include the grass-counter example and his suggestive remarks on it, I see no way for him to defend his version of the original position without saying the following: a human being who does not value communication or interaction as essential to a good life for them is sufficiently plausible that we must take them into account and cannot be biased against them, when designing the original position. A short way of capturing this commitment is to present it as a commitment to non-social human beings, but we must be careful to remember exactly what this means here. 'Non-social human being' means a human being who does not value communication or interaction as essential to a good life for them (and is not suffering from an extreme psychological disorder or mistaken about what kind of life they can enjoy). Strictly speaking, Rawls does not have to assert that there could be non-social human beings, but he has to regard the belief that there could be such people as plausible, plausible enough for him to design the original position so as to avoid bias against any such people. And if one is trying to avoid bias against them, this seems barely distinguishable from treating non-social human beings as possible.

## References

Edwin Baker, C. 1985. Sandel on Rawls. *University of Pennsylvania Law Review* 133: 895-928.

Rawls, J. 1999 (revised edition). *A Theory of Justice*. Cambridge, Massachusetts: Belknap Press.

Sandel, M. 1984. The Procedural Republic and the Unencumbered Self. *Political Theory* 12: 81-96.