E-LOGOS – Electronic Journal for Philosophy 2016, Vol. 23(2) 37–41
ISSN 1211-0442 (DOI: 10.18267/j.e-logos.435),Peer-reviewed article
Journal homepage: e-logos.vse.cz

Rawls versus utilitarianism: the subset objection

Terence Rajivan Edward[1]

Abstract: This paper presents an objection to John Rawls's use of the original position method to argue against implementing utilitarian rules. The use of this method is pointless because a small subset of the premises Rawls relies on can be used to infer the same conclusion.

Keywords: original position, utilitarianism, separateness of persons, subset objection.

[1] *School of Social Sciences, Arthur Lewis Building, The University of Manchester, Manchester. M13 9PL, email: t.r.edward@manchester.ac.uk .*

In *A Theory of Justice*, John Rawls asserts that justice is the first virtue of social institutions (1999: 3). Institutions must be improved or else abolished if they are unjust. Here we can interpret him as using the word 'just' to mean that they distribute rights, duties and the benefits of social cooperation in a fair way. But when do they do this? Different people may put forward different views about this. One person may say that institutions are just when they abide by rule A, another person may say that institutions are just when they abide by rule B, a third person may say that institutions are just when they abide by rules C and D, and so on. One of Rawls's contributions to political philosophy is to propose a method for resolving these disagreements, namely the original position. In this paper, I introduce an objection to the use of this method to resolve the disagreement between Rawls and utilitarians over when institutions are just.

Before presenting this objection, it is necessary to explain what the original position is. The original position is a thought experiment. Rawls asks us to imagine a set of people deciding to live together as a society and forming an agreement on rules (1999: 10). These rules concern how social institutions should distribute rights, duties and the benefits of social cooperation. Each individual in the original position is self-interested and is interested in gaining as many goods of a certain kind as possible. Rawls refers to the goods in question as primary goods. For the purposes of his thought experiment, Rawls assumes that the primary goods that a society distributes are rights, liberties and opportunities, income and wealth, and the social bases of self-respect (1999: 54).

Now if an individual in the original position knows certain features of themselves which are not shared by all others, they might well seek to achieve an agreement on the societal rules that is biased towards those features. For example, if an individual knows that they have a talent for growing food, they might seek an agreement which requires that institutions within society accord a number of privileges to people with this talent. Rawls thinks that a fair agreement would be achieved if the parties are self-interested but are unable to access information about themselves which would lead them to prefer an agreement biased towards the particularities of their own case (1999: 11). So he asks us to imagine that they do not have access to this information. They do not know their occupation, gender, class position, natural endowments, or conception of what a good life would be. Rawls describes individuals in the original position as behind a veil of ignorance, owing to this lack of knowledge (1999: 118). The method of the original position is to present such individuals with a menu of options, each option specifying a set of rules, and then consider which option they would select, given that they are all self-interested, that what they all want is primary goods and that they are behind a veil of ignorance. If individuals in the original position would prefer one option over another, then the rules of that option are more just than the rules of the second option.

This method provides a way of resolving disagreements over which rules institutions must abide by in order to be just. If one person thinks that they must abide by rule A and another person thinks that they must abide by rule B, we can consider which rule people in the original position would prefer. If they would prefer A over B, then rule A is more just than rule B. Rawls uses the original position method in this way to argue against utilitarian rules. These rules are called utilitarian because the tradition of utilitarianism suggests these rules. One utilitarian rule that Rawls considers is often referred to as 'the total utilitarian rule.'[2] The rule is that a society's institutions should produce the greatest amount of happiness within that society. I focus on this rule below, but what I have to say applies to the other utilitarian rule he considers as well.

[2] Rawls himself does not use this label. He refers to this rule as classical utilitarianism (1999: 160).

Rawls argues that individuals in the original position, if presented with certain rival options, would prefer those options to the total utilitarian rule (1999: 160). Rawls's rival option consists of two principles. Here we need only present the first of these: each person is to have an equal right to the most extensive set of basic liberties compatible with a similar scheme of liberties for all (1999: 53). The basic liberties are: political liberty, which is the right to vote and the right to be eligible for public office; freedom of speech and of assembly; liberty of conscience and freedom of thought; freedom of the person, along with the right to hold personal property; and freedom from arbitrary arrest and seizure (Hart 1973: 539). The total utilitarian rule allows that an individual can be forcibly enslaved or killed if this maximizes overall happiness, but Rawls's alternative option denies this because his first principle (his first rule) protects each individual's right to basic liberties. Individuals in the original position are self-interested and so prefer this alternative, because in the worst-case scenario they will not be sacrificed for the happiness of others. Hence Rawls's alternative is more just, using the original position method. We can describe this as an original position argument against the total utilitarian rule.

The material above can be elaborated upon and rendered more rigorous. I think it will be excusable if I mostly pass over these tasks here, in order to move more quickly to an objection. But I do need to engage in some elaboration first. The elaboration concerns why Rawls thinks that his original position thought experiment is an adequate method for working out when societal rules are just.

Rawls thinks of the original position as an adequate method partly because it takes into account the separateness of persons (Richardson 2005). As I understand Rawls, what he means by taking into account the separateness of persons is being consistent with the view that each person is a distinct being with value in themselves and with their own life to lead, which ought to be respected. This statement could do with more clarification, and I think Rawls could have clarified it more. But he does imply that the following example is an illustration what it is to not take the separateness of persons into account (1999: 24), which is all the clarification that we need here. Let us imagine, in a somewhat artificial way, that there is only one society and it consists of just three people and that there are no other beings alive that are capable of happiness. Person 1 has only one unit of happiness, person 2 has nine units of happiness and person 3 has ten units of happiness. But if person 1 is forcibly enslaved, resulting in no happiness for them, person 2 would have twenty units of happiness and person 3 would have thirty units of happiness, such are their preferences. So enslaving person 1 produces a greater overall amount of happiness within the society, since fifty units is more than twenty units. In a way that is analogous to how a single person might reason that a part of themselves must be sacrificed in order to achieve a higher level of personal happiness (e.g. a tooth must be removed), we can treat all the members of a society as if they were parts of a single organism and do things to a part of the social organism, such as person 1, for the greater overall happiness of this organism (Nozick 1974: 32; Rawls 1999: 163). But this approach does not take into account that each person is a distinct being with value in themselves and a life of their own to lead, which means that some ways of producing greater happiness are wrong, such as forcibly enslaving person 1.

A full statement of one of Rawls's original position arguments[3] against utilitarianism must include the reasons for thinking that the original position is an adequate method and so must

[3] Rawls has more than one argument against utilitarianism using the original position method. What I presented earlier was only one of these, but the lesson I extract applies to all of them.

include that the original position takes into account the separateness of persons. Here is a list of components that must be included in a full statement of such an argument:

(i) The claim that the first virtue of institutions is to be just.
(ii) The claim that the original position is an adequate method for deciding between proposed rules of justice.
(iii) The reason for accepting that the original position is an adequate method for deciding between proposed rules, part of the reason being that *an adequate method must take into account the separateness of persons and this method does so*.
(iv) The application of this method to show that a utilitarian rule is unjust.

The problem is that once you have introduced (iii), have you not already implied that the utilitarian rule is unjust before even getting to (iv)? Component (iii) involves explaining what it is to take into account the separateness of persons and what it is to not do so. This component also involves asserting that the separateness must be taken into account. In the process of doing all this, one implies[4] that the utilitarian rule is unjust. It is unjust because it treats persons as if they were mere parts of an organism, to be sacrificed if doing so increases the happiness of the whole.[5] Given this implication, it is pointless to go through the detailed application of the original position method to show that the utilitarian rule is unjust. Component (iv) can be dropped.

Sometimes when Rawls's attack on utilitarianism is summarized, he is portrayed as having arguments against it using the original position method and also other arguments against it which do not require using this method (Goldman 1980: 351; Scheffler 2001: 152). One of these other arguments says that utilitarianism fails to take into account the separateness of persons (Goldman 1980: 363). But why would an original position argument and the separateness of persons argument both be of interest for establishing that utilitarianism is unjust? Multiple arguments against something are not always of interest. Two arguments against something are of interest when it is consistent to endorse either argument without endorsing the other argument. They provide independent lines of attack. Two arguments against something are not both of interest, for the purpose of establishing their conclusion, when the premises of one argument are a smaller subset of the premises of another argument yet the reasoning of the briefer argument is valid. For example, if you make an argument for a certain conclusion and this argument begins with ten premises, but the very same conclusion can be deduced from just three of the ten premises, then the ten premise argument is a pointless exercise. *The worry I have about Rawls's original position arguments against utilitarianism is that a small subset of the premises involved will allow one to arrive at the same conclusion, hence original position arguments are pointless.* An original position argument against a particular utilitarian rule must also include material to justify the original position method. That material itself entails the unjustness of the utilitarian rule under consideration. So there is no need to go on and apply the original position method to discover this unjustness. Note that there is another utilitarian rule that Rawls considers, namely the average utilitarian rule (1999: 130).

---

[4] Strictly speaking, to validly deduce this conclusion, one will need a definition of the utilitarian rule under consideration, so 'implies' may be too strong. But one can have that definition without having the entire component (iv).

[5] I have doubts about the separateness of persons requirement, but the point here is that if this requirement is granted, then one can infer the unjustness of the utilitarian rules that Rawls considers (see also Rawls 1999: 24 and 163).

There is no reason why the subset objection[6] cannot be applied to original position arguments against this rule as well.

References

Goldman, H.S. 1980. Rawls and Utilitarianism. In G. Blocker and E. Smith (eds.), *John Rawls' Theory of Social Justice*. Athens, Ohio: Ohio University Press.

Hart, H.L.A. 1973. Rawls on Liberty and its Priority. *The University of Chicago Law Review* 40: 534-555.

Nozick, R. 1974. *Anarchy, State and Utopia.* New York: Basic Books.

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

Richardson, H. 2005. John Rawls (1921-2002). *Internet Encyclopedia of Philosophy.* Accessed on 2nd July 2014 from: http://www.iep.utm.edu/rawls/

Scheffler, S. 2001. *Boundaries and Allegiances: Problems of Justice and Responsibility in Liberal Thought*. Oxford: Oxford University Press.

[6] This is not an ideal name for the objection because a set which is a subset of another set is not necessarily smaller. It should be kept in mind that, in this context, the subset of premises has to be smaller.