Three assumptions of Rawlsian reflective equilibrium

Author: Terence Rajivan Edward

Abstract. John Rawls recommends a method for evaluating which principles institutions should abide by, known as reflective equilibrium. In this paper, I identify and challenge three assumptions that he makes.

Introduction. Imagine that a philosopher is interested in developing a theory about which principles institutions in a society should abide by. John Rawls recommends the following: the philosopher begins with their moral judgments about what such institutions should and should not do, and then tries to develop a theory which entails these judgments. The theory will consist of a few principles and, from the principles, they will hopefully be able to deduce these judgments. If a proposed theory is incompatible with many of the judgments, the philosopher should revise or abandon it. But if it entails most of them, they should consider whether it is a good idea to instead abandon an isolated judgment which does not fit. And if, after any appropriate revisions, the theory entails the judgments that are left, then the philosopher has reached what Rawls calls reflective equilibrium (1999: 18).

The term "reflective equilibrium" can actually be used in two senses. In one sense, it refers to this procedure of trying to achieve a theory which entails one's moral judgments, a procedure that allows for two-way receptivity. There is some possibility of judgments being revised to fit with a theory, as well as a theory being revised to fit with judgments. In another sense, it is the end state where a theory has been formulated that entails one's moral judgments.

The description of the reflective equilibrium procedure so far needs to be qualified in

certain ways to capture Rawls's thinking exactly. I shall add one qualification here. Until now I have written as if, when carrying out the procedure, a philosopher should pay attention to all their moral judgments about what institutions in a society should and should not do. But actually Rawls thinks they should only pay attention to "considered moral judgments." These, he tells us, are judgments made in conditions where our moral capacities are most likely to be displayed without distortion (1999: 42). He regards states of hesitation, of being upset and being frightened as examples of adverse conditions. Considered moral judgments can be more specific, such as that no government should set fire to Ned's house without reason; or more general, such as that no government should damage property without reason. Below I shall refer to considered moral judgments as "intuitions," though it is open to question whether this is an ideal term for them (Brun 2014).

Rawls hopes that by using the procedure of reflective equilibrium, philosophers can arrive at a consensus regarding which principles institutions in a society should abide by. An obvious worry is that it will not lead to this because different philosophers have different intuitions, different considered moral judgments. There is an assumption that Rawls makes and contesting the assumption reveals an unusual version of this worry. In the next section of this paper, I draw attention to the assumption. The later sections draw attention to and contest two other assumptions. Note that, although I have written of philosophers so far, Rawls recommends his procedure to all his readers, whether they can be called philosophers or not. And so I write simply of people below.

First assumption. In order to uncover the assumption I have in mind, it is useful to quote the passage where Rawls introduces his concept of a considered moral judgment:

> So far, though, I have not said anything about considered judgments. Now, as already suggested, they enter as those judgments in which moral capacities are most likely to be displayed without distortion. Thus in deciding which of our judgments to take into account we may reliably select some and exclude others. For example, we may discard those judgments made with hesitation, or in which we have little confidence. Similarly, those given when we are upset or frightened, or when we stand to gain one way or the other can be left aside. All these judgments are likely to be erroneous or to be influenced by an excessive attention to our own interests. (1999: 42)

In this passage, Rawls assumes that there will be agreement among readers about which states of mind are most likely to allow our moral capacities to be displayed without distortion. We can say he assumes this because he writes that "we may reliably select some" judgments, as if he and his readers will share beliefs about which states are reliable, and he shows no signs of anticipating disagreement about this issue.

As noted before, Rawls identifies some adverse conditions, such as when we are hesitant, upset or frightened. But even after this clarification, people will probably disagree about the conditions in which a person's moral capacities are most likely to be displayed without distortion. To illustrate this point, I shall quote from an essay by Bertrand Russell entitled "Mysticism and Logic." In the essay, Russell divides mystical experience into two parts, a negative part and a positive part. He writes:

> All who are capable of absorption in an inward passion must have experienced at times the strange feeling of unreality in common objects, the loss of contact with daily things, in which the solidity of the outer world is lost, and the soul seems, in

> utter loneliness, to bring forth, out of its own depths, the mad dance of fantastic phantoms which have hitherto appeared independently real and living. This is the negative side of the mystic's initiation; the doubt concerning common knowledge, preparing the way for the reception of what seems a higher wisdom. (1914: 784)

I have not experienced anything so extreme as this! Russell goes on to describe the positive part of mystical experience as consisting in an experience of certainty and revelation (1914: 785). This is followed by the formation of beliefs (1914: 785).

Now there are sure to be people who doubt that a mystical state is a state in which a person's moral capacities are most likely to be displayed without distortion. Instead they recommend relying on judgments made when one is in an ordinary sensible state of mind. So there will be disagreements about this issue, contrary to Rawls's assumption.

One response to the false assumption that Rawls makes is to just allow an individual to take intuitions from whichever state of mind they regard as most likely to let their moral capacities be displayed without distortion. But this may well lead to different people entering different intuitions into the reflective equilibrium procedure, taken from very different states of mind,[1] which in turn can lead to one person achieving reflective equilibrium with one theory and another person achieving reflective equilibrium with another theory.

Second assumption. Rawls instructs us to take moral judgments only from states of mind in which we are least likely to make errors. Let us grant, for the sake of argument, that readers share his non-mystical beliefs about which states these are (1999: 42). The judgments taken are supposed to be, for the moral theorist, what scientific data are for the scientific theorist. Rawls

[1] According to Russell, the mystic does not make judgments in the mystical state itself, rather afterwards, when reflecting on what was revealed in mystical experience (1914: 785). I do not think that this affects the main point of this section. The mystical experience is still treated as a source of knowledge.

therefore assumes that no information which is necessary for producing an adequate theory of societal principles is bound up with other, riskier states of mind. Could it not be that some important moral information is bound up with risky states: states of mind where there is a greater chance of erroneous judgments? By "bound up with a risky state," I am thinking of how a certain person might only make a certain judgment in a risky state of mind. If so, the distinctive information contained within that judgment is bound up with the risky state for that person.

Rightly or wrongly, people do sometimes treat important moral information as bound up with risky states of mind. For example, one might be given a strong argument for doing something, by someone very good at arguing, but one feels unease about the course of action and decides not to do this thing. I presume that there is nothing that abnormal about this. Indeed, I think that receptivity to unease is part of ordinary decency. But it is hard to know how much faith to place in such feelings. They can easily be wrong. Given that people do sometimes treat important moral information as bound up with risky states, Rawls's assumption requires defence.

I anticipate that Rawls would defend it by saying, "The assumption is defensible because I have produced an adequate theory of societal principles without relying on judgments taken from risky states." (See 1999: 35) I will not detail Rawls's theory here, but it strikes me as inadequate. It seems unsuitable for countries which have historical treasures, such as the pyramids and the monument to the Sphinx, since it treats government spending on these items as something that should not happen, even if the country becomes very wealthy (see Weinberg 2011).

Third assumption. If we look closely at the quotation from Rawls earlier, he tells us that we should rely on judgments made when our moral capacities are most likely to be displayed

without distortion. Only such judgments are to be entered into the reflective equilibrium procedure. He then says that we can discard those judgments made when we are hesitant, upset, frightened or stand to gain in some way, because these are likely to be erroneous or influenced by excessive attention to our own interests. Rawls is therefore making the following assumption: if a moral judgment is likely to be erroneous or influenced by excessive attention to our own interests, it is also a judgment which has not been made when our moral capacities are most likely to be displayed without distortion.

By registering this assumption, we can reconstruct Rawls's argument for discarding judgments made when we are hesitant, upset, frightened or stand to gain in some way:

(1) In the reflective equilibrium procedure, we should only rely on moral judgments made when our moral capacities are most likely to be displayed without distortion.

(2) If a moral judgment is likely to be erroneous or influenced by excessive attention to our own interests, it is also a judgment which has not been made when our moral capacities are most likely to be displayed without distortion. (Assumption)

(3) Moral judgments made when we are hesitant, upset, frightened or stand to gain in some way are likely to be erroneous or influenced by excessive attention to our own interests.

Therefore:

(4) In the reflective equilibrium procedure, we should not rely on moral judgments made when we are hesitant, upset, frightened or stand to gain in some way.

Rawls asserts (1) and (3) and concludes (4), but he needs premise (2) in order for the conclusion to follow from his premises. That is why we are justified in attributing it to him as an assumption.

I have a doubt about whether Rawls can always say that focusing mainly on one's self-

interest is a significant risk to displaying one's moral capacities without distortion. For Rawls, a way of working out what is fair is to think as a self-interested person in certain conditions would. An interpretation of Rawls is that, according to him, our capacity for judging without distortion what is fair is simply a capacity to think as such a person would (Putnam 2005: 116). But what then if a person finds themselves in precisely these conditions? It does not seem that Rawls can say that focusing on their self-interest increases the risk of their moral capacities displaying themselves with distortion, if these capacities are trying to work out what is fair. Consider this thought experiment, which is a minor variant on one that he recommends (1999: 74). Imagine that the hostess of a party is cutting a cake. The cake looks and smells delicious and she would like quite a lot; but in this culture the hostess is only supposed to take a piece after serving each guest. It is in her interests to cut the cake equally, because any other way of dividing it up would mean a chance of getting a smaller size. She tells guests that this way of dividing it up is fair, when she is mainly thinking about her interests. I do not see how Rawls can say that the hostess is at greater risk of her moral capacities displaying themselves with distortion.

References

Brun, G. 2014. Reflective Equilibrium Without Intuitions. *Ethical Theory and Moral Practice* 17: 237-252.

Putnam, H. 2005. John Rawls. *Proceedings of the American Philosophical Society* 149: 113-117.

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

Russell, B. 1914. Mysticism and Logic. *The Hibbert Journal* 12: 780-803.

Weinberg, J. 2011. Is government supererogation possible? *Pacific Philosophical Quarterly* 92: 263-281.