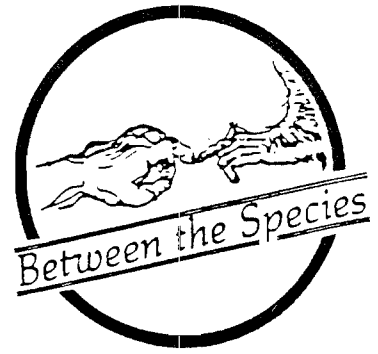
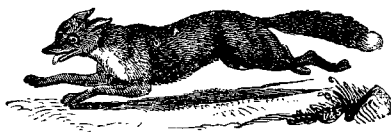


Animals in the Original Position

Lilly-Marlene Russow
Purdue University



Editors' Note: The following paper by Professor Russow, commentary by Professor Sapontzis, and reply by Russow were presented at the Pacific Division of the Society for the Study of Ethics and Animals, held in Portland, Oregon, March, 1992.



In *A Theory of Justice*, John Rawls¹ defends two principles of justice by means of an appeal to a hypothetical “original position.” However, these principles are strictly principles of justice by and for human beings, or at least by and for *persons*. Several philosophers have recently argued that this argument is flawed—is “speciesist”—in that there is no adequate justification for excluding animals from consideration as beings to whom the principles of justice ought to apply.² If those in the original position had to consider the possibility that they might be reborn as a calf rather than a human, they might well choose different principles. If they did not consider that possibility, the original position, and the “veil of ignorance” which helps define it, would fail to protect against prejudice or guarantee fairness and impartiality.

One must be careful to distinguish at the outset different bases for objecting to the exclusion of nonhuman animals from Rawls’s theory. One might argue that Rawls simply cannot exclude animals from his theory, that doing so is internally inconsistent. Alternatively, one might argue that while animals can consistently be excluded, there are good reasons for including them. This line of argument might either

attempt to find these good reasons within Rawls’s own theoretical proclivities (the second possible strategy) or—a third type of project—introduce new themes which are independent of, even contrary to, some of Rawls’s basic commitments. On this last strategy, one would be arguing that the notion of an original position can be used in contexts other than the Rawlsian one to generate a concept of justice that would be, in VanDeVeer’s terms, “interspecific.”³ Of course, all of these attacks presuppose the more basic claim that it is logically possible to include animals in the original position.

In this paper, I shall only briefly comment on the argument that inclusion does not lead to logical absurdity; Rawls’s critics, most notably VanDeVeer, are correct in noting that the original position already has in place the mechanisms by which the interests of animals could be considered in the same way as those of humans. I shall argue that any attempt to *force* an expansion of the original position fails, and that Rawls is quite justified in restricting the argument so that it does not apply to most animals. While the critics apparently take themselves to be engaged in arguments of the first or second sort, they must, at best, content themselves with something like the third strategy. In order to see why, we must first examine the concept of the original position and then step back for a sense of the broader context in which talk of the original position is embedded. The description of Rawls will necessarily



PHILOSOPHY

be oversimplified, but the relevant portions of the argument can be identified.

Rawls argues for his two principles of justice by appealing to “the original position,” a hypothetical situation in which “free and equal” individuals choose the principles of justice that seem to them most rational and advantageous—principles to which they would voluntarily accede (TOJ, pp.12-13). To insure impartiality, we must imagine that these deliberations take place behind a “veil of ignorance”; the participants have no knowledge of social or economic status, gender, particular talents or abilities, or “particular inclinations and aspirations” (TOJ, p. 18). Rawls claims that individuals in this position would choose the following two principles:

First: each person is to have an equal right to the most extensive basic liberty compatible with a similar liberty 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. (TOJ, p. 60)

Both principles tell us only about the proper treatment of persons. If ‘person’ does not include most nonhuman animals, adoption of these principles alone would still allow for the most egregious mistreatment of animals, and this gives rise to the criticism under consideration.⁴ The challenge is most forcefully posed by Donald VanDeVeer, who argues that we can redescribe the original position without lapsing into the conceptual incoherence of trying to imagine that we have a person, a dog and a pig deciding which principles to choose. Moreover, he argues that we must redescribe it in order to guarantee fairness, or else it will fail to function as Rawls intended it to.

People in the original position are envisaged as a fairly sophisticated bunch. They are rational, can recognize a valid argument when one occurs to them, and can grasp complicated economic calculations such as maximin. However, if they knew that they would be that clever in “real life,” they would tend to choose principles that would give favored treatment to clever people. Thus, the veil of ignorance is constructed to preclude that sort of inequality: those in the original position must find principles that they could accept even if they turn out to be ill equipped to do moral philosophy.

This is the point at which VanDeVeer raises his challenge. If people in the original position are not allowed to act on the knowledge that they will be “sophisticated reasoners,” it seems equally legitimate to insist that they should not be allowed the knowledge that they are moral persons at all.⁵ Doing so would encourage those in the original position to adopt principles that would protect nonhuman animals, without the absurd requirement that they must formulate these principles *using* only the limited capacities of such animals. It is as if Rawls recognizes that sophisticated intellectual ability should not be a morally relevant characteristic in determining just treatment, and so sets up the original position to make sure that it will not be favored by the chosen principles. But he simply assumes without argument that being a rational (though not necessarily moral) agent or moral person is morally relevant, and hence allows that information to be accessible, even under the veil of ignorance. In VanDeVeer’s words:

Should not the veil of ignorance exclude the knowledge that members of the society [i.e., all individuals whose treatment is dictated by the principles chosen in the original position] will have a concept of their good or of justice?...The rationale of a veil of ignorance, on Rawls’s own view, is to guarantee impartial consideration of principles.⁶

If VanDeVeer is correct, Rawls has no adequate justification for assuming that the original position should allow those choosing the principles to proceed on the assumption that they will be moral persons; indeed, such an assumption would lead to just the sort of self-interested prejudice that the veil of ignorance was meant to prevent. If they are *not* entitled to that assumption, VanDeVeer suggests that they would choose different principles:

they might choose principles which preclude treatment of any sentient creature (not posing a serious threat to others) which would render *no life at all* for that creature, on balance, ...preferable to its living. I call this the Life-Preferability Principle.⁷

In short, the problem to be addressed is whether it is legitimate to justify one’s choice of principles by

appeal to an original position in which the participants are allowed to assume that they are/will be moral persons, as that concept is defined by Rawls. Going back to our second and third strategies, we can first ask whether the Rawlsian framework provides any reason for doing so and, second, consider whether abandoning Rawls at this point yields a more satisfactory theory. VanDeVeer does not distinguish these two tactics, but since I shall argue that the first will not work, he can at best hope for the second.

To determine how a Rawlsian might respond to VanDeVeer's challenge, we must go back to an essential aspect of Rawls's starting point: his rejection of utilitarianism. I do not wish to re-examine his reasons for that rejection here but merely to note that it is essential to his theory: any attempt to extend or modify the description of the original position *in a Rawlsian spirit* must be consistent with that rejection of utilitarianism. More specifically, the entire idea of an original position and the veil of ignorance is presented as an alternative to utilitarianism; therefore, any criticism of Rawls's strategy must be consistent with the rejection of utilitarianism that motivates the strategy, or else must reach further and attack the anti-utilitarian arguments on which the original position is based.

We can frame the same point in a more positive light. Describing the original position is the second step in an argument which begins by arguing that a theory of justice must, above all, be structured so as to avoid what Rawls has already argued is the main flaw of utilitarianism: the failure to respect individuality. Rawls seems to assume what others have argued for: to respect individuality one must reject the idea that it is rational (rationally obligatory?) to have no or less interest in the fulfillment of one's own desires as in the fulfillment of another being's equally strong interest.

If one rejects utilitarianism in favor of respect for individuality, one moves inevitably, as Rawls does, toward a system which protects the individual's ability to choose and carry out his/her own plans, to further her/his own purposes. In Rawls, this is carried one step further: the sense of individuality which carries moral weight involves having and caring about a life-plan. Thus, he defines moral persons as having "a conception of their good (*as expressed by a rational plan of life*)" (TOJ, p. 505, my emphasis). Whether all of these conclusions follow from a rejection of utilitarianism remains to be seen, but that is how, in Rawls, the notion of moral personhood is developed and how it comes to

be built into the description of the original position. The choice and ordering of the two principles of justice is then no surprise, since the original position is intended to reflect this prior commitment to the moral significance of an ability to act as an individual, more specifically an ability to choose and act on a life-plan of one's own. This emphasis on individuality is, again, something that classical utilitarianism is supposed to be unable to accommodate.⁸

All of this suggests three possible criticisms of Rawls:

1. He has not adequately justified his first step, the rejection of utilitarianism in favor of a theory that puts primary weight on freedom and moral personhood.
2. The description of the original position does not follow from the first step, and cannot be justified by appeal to it.
3. There is an inconsistency in the overall argument, and the burden of proof is on Rawls, not the critic, to diagnose the source of the problem.

The first criticism might lead us to abandon some basic elements of the Rawlsian view and perhaps follow the third strategy described at the beginning of this paper. The second criticism leads naturally to the second strategy, and the third criticism follows the first strategy. Let us consider each of these in turn.

The first—a frontal attack on Rawls's rejection of utilitarianism, and his consequent emphasis on respect for the individual agent or subject—is certainly open to fierce debate, as demonstrated by numerous commentators on Rawls. Moreover, it is not limited to concerns about animals: the controversy will not center on the exclusion of animals but, rather, on general considerations for or against utilitarianism.

Despite its apparent attractiveness, I propose not to evaluate this tactic. My primary justification for doing so is that the criticisms leveled against Rawls's exclusion of animals have not objected specifically to this aspect of the argument. Two other reasons reinforce this inclination. First, any dispute over this starting point is not peculiar to concerns about animals: it can be, and has been, raised purely within the context of duties, obligations, and justice toward anyone or anything who is not a moral person in the relevant sense—and this

would include some human beings. Therefore, such an objection would not support the specific charge of speciesism. Second, several debates within the "ethics and animals" literature serve to reinforce the Rawlsian line that utilitarianism (at least in its classical guise) pays insufficient attention to the status of the individual. Regan's emphasis on the moral status of being a "subject of a life"⁹ and even Peter Singer's qualified position on replaceability¹⁰ move us in a direction congenial to Rawls's first premise. The sense of "individuality" (as opposed to merely being a receptacle of pleasurable experiences) which is morally relevant is sure to be hotly debated, but until one argues (as someone like Regan might want to¹¹ but has not explicitly done) that Rawls has set the standard for "subject" too high, this first method of attack on Rawls must be discounted.

The second strategy would be to argue that Rawls's rejection of utilitarianism does not justify the most common interpretation of the description of the original position: that it does not warrant allowing those in the original position to assume that they will not be 'incarnated' as nonhuman animals. I think this complaint is correct but achieves only a very minor victory for two reasons. Rawls himself emphasizes moral personhood, not biological facts about membership in the species *Homo sapiens*, as a crucial factor.¹² Thus, a few nonhuman animals, perhaps dolphins or gorillas, might be capable of formulating a rational life plan and valuing it because it is their own life plan; they would thereby qualify as the sort of beings that participants in the original position ought to include in their considerations. Nonetheless, the best scientific evidence we have to date would indicate that this is a plausible hypothesis only with a very small number of species: it will not, for example, save the veal calf or the laboratory rat. Elliot does not think this is an obstacle to including animals, because we can still "think ourselves into their position" and make reasonable inferences about what sort of life would be preferable to and for them.¹³ This, however, misses the point: the issue is not whether nonhuman animals prefer one sort of life to another but whether the very choosing of a life-plan of one's own is important to them.

According to Rawls's argument, those in the original position, even though they do not know what their life-plan will be, are supposed to know that they will *have* a rational life plan, i.e., to know that following plan A rather than plan B matters simply because one has *chosen* A rather than B.¹⁴ This, as noted earlier, is

presented as a consequence of the rejection of utilitarianism. Once one grants that much, it follows that those in the original position are justified in assuming that they will not be reincarnated as anything other than a human being or perhaps a member of one of a very select and limited group of species. As hinted at above, there may be a sense of "individuality" which at the same time preserves the force of the objection to utilitarianism while at the same time demanding less than Rawls demands of moral persons, but such a middle ground has not been successfully delineated.¹⁵

The third strategy in attacking Rawls is simply to point out an inconsistency and leave it to the Rawlsian to decide how to fix it. Those who have pursued this course have identified such an inconsistency, the discussion of so-called "marginal cases".¹⁶ Although the inconsistency is real, the modification of Rawls's position on this point that seems most consistent with his overall theory is not one that results in animals being considered in the original position.

The term "marginal cases" refers to human beings who are so severely retarded or brain damaged that they lack the most rudimentary ability to make conscious, deliberate, rational choices. We can restrict the field even further by specifying that they are individuals who never had such abilities, and have no realistic chance of ever acquiring them, since individuals in the original position may want to take into account the fact that they will start life as helpless infants and may be the subject of an accident or disease that could destroy their ability to function as a moral person (TOJ, pp. 248-250). Such "marginal cases" are clearly less able to construct or follow a life plan than a pig or a hamster,¹⁷ yet Rawls still accords them protection that he does not extend to nonhuman animals. It should be noted immediately that Rawls's discussion of this topic is both tentative and equivocal (TOJ pp. 248-250, 506-510). *If* we accept his critics' assumption that those in the original position would extend the principles of justice to all human beings, even marginal cases, then Rawls is guilty of inconsistency. But the inconsistency is easily remedied, if indeed it is supported by the text at all.

Given Rawls's commitment to liberty, moral personhood, and the importance of freedom to choose, the only solution consistent with a Rawlsian framework would be to accept that marginal cases are not included in considerations of justice. This, however, should not horrify us, because it does not leave marginal cases without any moral protection. Rawls takes pains to

admit that the theory of justice is not meant as a complete moral theory (TOJ, p. 512) and we might well have many good reasons for extending special protection to marginal cases.¹⁸ The possibility for inconsistency exists, in that the full but as yet unarticulated moral theory *could* turn out to accord protection to marginal cases that it denies to animals without sufficient reason, but that hypothetical possibility does not pose a serious threat to the theory of justice as presented by Rawls.

To sum up: we can describe a theory of justice that is consistent in both letter and spirit with Rawls's that has the following characteristics:

1. Rawls is justified in specifying that all those in the original position may assume that they will be moral persons in the society to be governed by the principles which they choose. Thus, the principles of justice which they choose might easily turn out not to cover many who are not moral persons.
2. Nonetheless, some individuals who are not included as moral persons, or (not necessarily equivalently) those to whom the principles of justice apply, may be moral subjects: our complete moral theory of duties and obligations may show that we have direct moral duties toward some who are not moral persons, e.g. marginal cases and nonhuman animals.
3. It is not a foregone conclusion that marginal cases and nonhuman animals cannot be accorded differential treatment without lapsing into speciesism, but any attempt to justify differential treatment must still be produced.

The implications of this issue extend beyond an abstract interest in the internal consistency of *A Theory of Justice*; they provide important clues about what an adequate account of our duties toward animals might look like. There are at least two lessons to learn. First, the notion that being a moral person or subject is morally relevant has not yet been impeached; it deserves further consideration and elaboration as an important difference between most humans and most nonhuman animals. Second, any attempt to build a theory on such a foundation must take care to avoid inconsistency in its treatment of marginal cases. Although these two points

are only a start, they do point in what I believe is the right direction.

Notes

¹ John Rawls, *A Theory of Justice* (Cambridge: Harvard University Press, 1971). Specific references to this work will be included in the body of the paper, identified as 'TOJ.'

² The following represent the most important versions of this criticism: Robert Elliot, "Rawlsian Justice and Non-Human Animals," *Journal of Applied Philosophy*, v. 1 (1984), pp. 95-106; Edward Johnson, *Species and Morality* (Ann Arbor: University Microfilms, 1976); Donald VanDeVeer, "Of Beasts, Persons, and the Original Position," *Monist* (1979) pp. 368-377; and Tom Regan, *The Case for Animal Rights* (Berkeley: University of California Press, 1983).

³ Donald VanDeVeer, "Interspecific Justice," *Inquiry*, (1979), pp. 55-79.

⁴ In this paper, I propose to set aside Regan's charge in "Duties to Animals: Rawls' Dilemma," *Ethics and Animals*, vol. 2 (1981), pp. 76-82 (see also *The Case for Animal Rights*, pp. 167-170) that Rawls is guilty of a simple inconsistency in his treatment of natural duties. Alan Fuchs has, I believe, successfully rebutted this charge in "Duties to Animals: Rawls' Alleged Dilemma," *Ethics and Animals*, vol. 2 (1981), pp. 83-88. Fuchs argues that Rawls's account of natural duties does not include the duty not to be cruel to animals. Although Rawls does acknowledge that duty elsewhere (TOJ, p. 512), it is not a natural duty in the sense Regan is worried about. As Fuchs points out, Rawls states several times that his theory of justice was never intended to be a complete moral theory and strongly suggests that our treatment of animals might well be restricted by moral principles which, while valid, fall outside the scope of a theory of justice.

⁵ Rawls takes the necessary and sufficient conditions for moral personhood to be something like: (a) having one's own ends, a conception of one's own good, or a life plan and an interest in following it, and (b) having the *capacity* for a sense of justice (p. 505; cf pp. 12, 19). We shall see presently that this definition of "moral person" is fraught with difficulties. It should be emphasized at the outset, however, that moral personhood is not restricted by definition to any one biological species; nor is it equivalent to "moral subject." That is, Rawls's argument allows for the possibility that beings who are not *Homo sapiens* might be moral persons and explicitly agrees that we have moral obligations that are owed to beings who are not moral persons in his sense.

⁶ VanDeVeer, pp. 371-372. It should be noted that, at least in principle, VanDeVeer's notion of "members of society" is broader than Rawls's "moral person". It is certainly logically possible that participants in the original position might decide

to extend the coverage of whatever principles they adopt so as to cover more than the class of moral persons—e.g. the children they might be born as.

⁷ VanDeVeer, p. 375.

⁸ This orientation helps justify Rawls's demand for a thin theory of the good, and a special emphasis on the primary goods which enable a person to pursue her chosen life plans. Elliot correctly notes that controversy rages about whether Rawls is successful in avoiding "want-regarding principles"; however, he does not think that animals pose any insurmountable difficulties in this regard (Elliot, pp. 103-104).

⁹ Regan, *The Case for Animal Rights*, pp. 243-248.

¹⁰ Peter Singer, "Killing Humans and Killing Animals," *Ethics*, vol. 22 (1979), pp. 145-156.

¹¹ It would seem that Regan's analysis of a "subject of a life" is also meant to capture a sense of individuality, but one which requires substantially less intellectual capacity than Rawls's account of a moral person.

¹² Rawls, pp. 505-510, *passim*.

¹³ Elliot, p. 104.

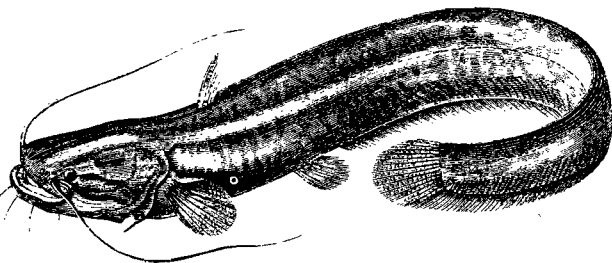
¹⁴ Cf. Rawls, p. 142.

¹⁵ Again, this is predicated on serious doubts about Regan's analysis of a "subject-of-a-life". Cf. fn. 10.

¹⁶ VanDeVeer and Regan both discuss this point, although it does not seem to be the major focus in either case.

¹⁷ This comparison tacitly recognizes an additional complicating factor that is not discussed in this paper: the fact that being a moral person may well be a matter of degree. I do not think this simplification affects the analysis of the arguments under analysis here, since none of them mentions this dimension, but it is dangerous to ignore it entirely.

¹⁸ For more on this topic, see Michael Wreen, "In Defense of Speciesism," *Ethics and Animals*, vol 3 (1982), pp. 47-60, and James Lindemann Nelson, "Animals, Handicapped Children, and the Tragedy of Handicapped Cases," *Journal of Medical Ethics*, vol 14 (1988), pp. 191-193.



Commentary: On the Utility of Contracts

Steve F. Sapontzis

California State University, Hayward

In many discussions that touch on animal rights, the participants clearly feel they know the truth and proceed to shape arguments to fit that truth. We owe Professor Russow our thanks for a careful, thoughtful discussion which has no axe to grind.

If I understand her argument, it runs something like this.

- A. To understand whether participants in the original position could be incarnated as nonhuman animals, we need to understand what the original position is supposed to accomplish.
- B. The original position was set up in response to Rawls's dissatisfactions with utilitarianism, principally to overcome (what he believes to be) utilitarianism's failure to respect individuality.
- C. Individuality, in the morally significant sense, involves having and caring about a life-plan.
- D. Consequently, the participants in the original position can be assured of being incarnated only as beings capable of having and caring about life-plans, so-called "moral persons," because only such beings are the object of the original position exercise.
- E. Very few, if any, nonhuman animals are capable of having and caring about life-plans.
- F. Consequently, the participants in the original position can be assured that with, at most, very few exceptions, they will be incarnated as human beings.



DISCUSSION