Contractualism

Jussi Suikkanen

Final author copy; To be published in Michael Hemmingsen (ed.): *Ethical Theory in Global Perspective* (SUNY Press).

Introduction

There is a long historical tradition of trying to understand morality in terms of a contract. The core idea in this tradition is that what is right and wrong is in some way grounded in either what we have agreed to do or in what we could be expected to agree to in some hypothetical circumstances. This contractualist way of thinking goes back to at least Ancient Greece (Plato, *The Republic,* 358e–359b), but it really became the prominent way of thinking especially about our political obligations during the Early Modern period through the social contract theories of Thomas Hobbes (1996 [1651]), John Locke (2002 [1689]), and Jean-Jacques Rousseau (1997 [1762]).

Contractualism is not, however, merely a historical tradition, but rather it continues to be a popular approach. In political philosophy, many debates concerning justice still tend to take John Rawls's (1971) contractualism as their starting point. Similarly, in moral philosophy, different ways of developing the basic contractualist insights are at the centre of several key theoretical debates (Gauthier 1986, Scanlon 1998, Southwood 2010, and Parfit 2011). That so many people have approached morality through the idea of a contract for over two millennia suggests that the contractualist framework must be getting something right. Yet, as we will see below, the devil will be in the details.

For the sake of simplicity, this chapter focuses on just one contemporary formulation of contractualism – the version outlined by T.M. Scanlon in his 1998 book *What We Owe to Each Other*. This view is often considered to be the paradigmatic version of contractualism – a pinnacle in the long tradition of contractualist thinking. It's also one of the clearest, most appealing, and most debated formulations of contractualism.

The next section provides an outline of Scanlon's theory. The following section then explains how contractualism differs from classical utilitarianism especially in its treatment of cases that involve different sized groups. After this, I will outline a key internal debate in contractualist ethics today to do with how the view should be formulated with respect to time. Finally, the last section focuses on two traditional objections to contractualism – the redundancy objection and the question of whether the view can explain our obligations towards non-human animals and cognitively impaired human beings.

Outline of Scanlon's Contractualism

According to Scanlon (1998, 147–153), ethical theories must be able to answer the following two questions:

1. Which actions are right and wrong?
2. What good reasons do we have for not doing the actions that are wrong?

When it comes to the first question, we are looking for a theory of right and wrong that would fit our carefully considered moral convictions about individual cases (Scanlon 2003, 149). Thus, a theory that entails that we should keep our promises and not be rude to our friends will be much more plausible in this respect than a theory that entails that it would be morally permissible to harm innocent babies for fun.

In the case of the second question, Scanlon emphasizes that a plausible ethical theory should be able to answer that question in an informative way (Scanlon 1998, §4,3; Prichard 1912). For this reason, a theory that claims that we should do right actions merely because they are right will be less plausible. In addition, if a theory claims that we should avoid doing wrong actions because acting in those ways will make us worse off (due to, for example, the disapproval of others), that theory too fails to provide a satisfactory answer to the second question. The reason offered here is a wrong kind of a reason – not a moral reason but rather one based on selfishness. We expect, after all, that good moral agents do what is right for some other reason than merely their own selfish interests.

Let us then consider how Scanlon's contractualism tries to answer the previous two questions. At the heart of his view is the following principle (Scanlon 1998, 153):

> An action is wrong if and only if it is forbidden by the set of principles no one could reasonably reject (and right if it is authorized by those principles).

Notice that the view is not formulated in terms of which principles individuals actually agree to accept or could accept. There are two reasons for this. Firstly, it is questionable whether there ever is or could be principles that everyone would accept at the same time – there just seems to exist too much moral disagreement for that to be possible. Secondly, Scanlon (1998, 155) worries that some self-sacrificing individuals will agree to principles that would be very bad for them, because those principles benefit others. If we understood right and wrong in terms of actual contracts, then it would be right to treat those individuals badly, which still seems objectionable.

So, instead of an actual agreement, Scanlon formulates his view in terms of principles no one could reasonably reject. What are they then? Scanlon (1998, 195) thinks that we should first consider what would be the consequences of adopting together different sets of moral principles. Different principles would, of course, make a difference to what kind of lives different individuals would come to live. We can then call the lives that individuals would come to live under the different principles their "standpoints" (Scanlon 1998, §5.4). Here some elements of those standpoints will be good and make the lives choice-worthy, whereas other elements will be different burdens the individuals occupying those standpoints will have to bear as a result of the moral principles they live under.

Scanlon (1998, §5.2) then claims that individuals can make objections to the principles we could adopt together on the basis of the personal burdens they would have to bear under them. What are these burdens? Scanlon (1998, 204) suggests that we cannot understand them on the basis of people's personal tastes or specific interests but rather we need to focus on what "generic reasons" individuals would have for objecting to the principles on the basis of their own personal standpoints. For example, we can ask whether the principles entail that individuals would be physically harmed, unable to trust other people or to form personal relationships with them, suffer from poverty or ill health, unable to be autonomous or to make free choices, have limited opportunities to engage with education, art, sciences, and so on.

Scanlon's (1998, 195) contractualism then focuses on the individuals who can make the strongest personal objections to the principles they live under based on the burdensome elements of their personal standpoints. Those individuals can reasonably reject a set of moral principles they live under if there is an alternative set to which no one has an equally serious personal objection. The set of principles no one can reasonably reject is then the one where some individuals have stronger personal objections to all other sets (ibid.). According to this view, the actions that are authorized by that non-rejectable set are then the morally right things to do for us whereas the actions that are forbidden by it are wrong. This is Scanlon's answer to the first question above. I'll use a concrete example in the next section to illustrate how the previous test works in practise, and the later sections will also consider how well the view fits our moral intuitions. Before that, let me introduce Scanlon's answer to the second question.

Why should we then follow the non-rejectable principles? The key claim here is that by doing so you can justify your actions to others on defensible grounds (Scanlon 1998, §4,3). In contrast, if you do not follow these principles, you are telling other people that you are willing to overlook their serious objections to your way of treating them. This is because, by violating the non-rejectable principles, you are expressing to others that you would rather follow some other principles that cause wholly unnecessary serious burdens to some individuals, burdens which no one would need to bear if you followed the non-rejectable principles. Furthermore, those who will be the victims of your unjustifiable actions will take the fact that you don't care about their objections to indicate that you are not willing to grant them the same equal moral status as for others and this they will see as a serious moral harm done to them. In Philip Pettit's words, for this reason we tend to "shrink from the gaze of another when we realise that it is impossible for us to justify our behaviour to someone else" (Pettit, 2000, 231).

On the positive side, if you follow the non-rejectable principles and thus show that you care about whether you can justify your actions to others on grounds no one could reasonably reject, this will allow you to form valuable moral relationships with others – the kind of relationships we have good reasons to be in for its own sake. Scanlon (1998, 162) calls these relationships ones characterized by "mutual recognition". In them, we can wholeheartedly stand by our actions knowing that no one could reasonably reject the principles we act on. Generally, this will also lead to an atmosphere of trust and mutual co-operation, which will also be in our interests in other ways too. Thus, to summarize, in this way Scanlon believes that there is a close connection between (i) his account of right and wrong in terms of the principles no one could reasonably reject and (ii) his story of why we ought to avoid acting wrongly based on the valuable moral relationships avoiding those actions will enable us to have.

Contractualism, Utilitarianism, and Aggregation

To see how the previous view works, let us consider an example Scanlon (1998, 235) uses both to illustrate the view and to compare it to classical utilitarianism. According to classical utilitarianism, an action is right if and only if it would lead to a higher total amount of general happiness than any other alternative available to you in the situation you are in. Consider then the following case.

It's the World Cup final – a match watched by millions of people. Unfortunately, there is an accident at the transmitter room and some crucial electronic equipment has fallen on Jones's arm. This is giving him extremely painful electric shocks but the only way to save Jones would be to cut off the transmission for fifteen minutes. The problem is that doing so would deny millions of people the enjoyment of watching the match for that time. What is the right thing to do?

Classical utilitarianism's answer to this question is that we should continue the transmission despite the suffering this would cause to Jones. After all, if we cut the transmission, we will lose millions of fifteen-minute-long periods of enjoyment added together, and this would be a worse loss in terms of the total amount of happiness than if Jones has to suffer for the rest of the match.

Scanlon's (1998, 235) contractualism, in contrast, entails that we should stop the transmission and save Jones. It makes us focus on the most serious objections that individuals would have to different solutions to the dilemma. If we continue the transmission, Jones will have a very strong objection to this based on the extremely painful electric shocks. Yet, if we stop the transmission, the strongest objection any one individual will have to this will be the loss of fifteen minutes of enjoyment. If we compare these two personal objections pairwise, it is evident that Jones's objection is stronger. This is why, on Scanlon's view, Jones can reasonably reject the principles that would authorize continuing the transmission whereas no one can reasonably reject the principles that require us to save him.

Most people here share the intuition that saving Jones would be the right thing to do. Consequentialist views, such as classical utilitarianism, that allow us to aggregate the minor benefits of many people together appear to let us sacrifice few individuals for the sake of the common good. They allow using minorities as a means to the happiness of the majority in a way that we tend to find objectionable. One motivation for Scanlon's (1998, 234–235) contractualism thus is that it offers individuals "anti-utilitarian" moral protections due to the "individualist restriction" built into it. This is the core idea that we can compare only the personal burdens and objections to different principles and never aggregate them together.

This important motivation for contractualism, unfortunately, also seems to lead to problems in the cases in which aggregating the objections seems quite appealing. For example, consider a case where you can save either one person from a certain death or a large group of people from something almost as bad, for example from permanent blindness or paralysis (Scanlon 1998, §5.9). In this case, no matter how large the group (say, millions), Scanlon's view seems to entail that we still ought to save the one individual from the slightly more serious harm. This is because her personal objection (death) to not saving her will be stronger than the personal objections the individual members of the large group (blindness or paralysis) have to saving the one individual and we are not allowed to add those objections together. As a result, Scanlon's contractualism appears to require us to save always the one individual from a serious harm rather than a very large group from a slightly less serious harm.

Scanlon (1998, §5.9) tried to address this concern already in *What We Owe to Each Other*. Firstly, he suggested that different harms belong to "broad categories of moral seriousness" (Scanlon 1998, 238). This allows us to think that harms that are almost as serious really belong to the same category and so they can be treated as comparable to one another. Thus, in the previous example, blindness and paralysis would belong to the same category as death. Scanlon (1998, 232) also suggested that, even if we are not allowed to aggregate the objections of different individuals together, we should still always save the larger group in the cases where the harms on both sides belong to the same category of seriousness. This is because, if the larger group is not saved in those cases, each member of that group can object not only to the harm they would suffer but also for the unfairness of the fact that their additional presence in the larger group is not making any difference to what we are to do in the case. Scanlon then suggests that this additional consideration counts as a tiebreaker when we must choose between saving a smaller group and saving a larger group where the personal objections on both sides are roughly equally strong. Whether these responses work continues to be debated intensively.

Ex-Post **versus** ***Ex-Ante*** **Contractualism**

This section introduces a more recent lively debate about how Scanlon's contractualism should be formulated with respect to from which temporal perspective we should consider the personal objections to different moral principles. The way I formulated Scanlon's contractualism above is today known as *ex-post* contractualism ("*ex-post*" is Latin and means "after the fact"). According to this type of contractualist views, when we compare different moral principles, we focus on what we can predict would actually happen to specific individuals as a result of the adoption of those principles. Here we are imagining a process where the moral principles are first adopted by everyone and then, because of this, different things happen to different individuals. After all of this has happened in the imagined process, we then retrospectively collect information about what happened to different individuals and how those individuals can now object to the principles they have lived under. Because of this temporal perspective of looking backwards, one thing that we will not take into account is how likely it was that any one individual had to bear the burdens he or she came to bear.

The other alternative is *ex-ante* contractualism ("*ex-ante*" means "before the event"). According to such theories, we are to consider objections which individuals would have to the adoption of different principles from a temporally antecedent perspective, from before those principles are adopted. From that perspective, individuals cannot object to the principles based on what actually happens to them as a result of their adoption. Such consequences, after all, haven't been produced yet by the principles, and it might be uncertain and unpredictable how different individuals will be affected. Because of this, from the temporally antecedent perspective individuals can object to the principles only based on what kind of *prospects* the principles give to them. This means that individuals can object to the principles based on different potential burdens the principles might produce, but importantly these objections must be discounted by how improbable the relevant burdens are.

In this framework, everyone can thus object to the principles based on the risks the principles impose on them. These risks can be understood more technically in terms of what is called an individual's "burdensomeness expectation" for a given set of principles. To get that expectation, an individual can multiply each personal burden a given principle could produce with its probability and then these products are summed up. The thought then is that the non-rejectable set of principles is such that all other principles create higher burdensomeness expectations for some individuals.

To see what these two alternatives mean in practice, let us consider two examples – one which supports the *ex-post* views and another which supports the *ex-ante* views. The first example is Scanlon's (1998, 208–209) own original reason for formulating his view in the *ex-post* way. Consider a health-care policy that would require "us to impose very severe hardship on a tiny minority of people, chosen at random (by making them involuntary subjects of painful and dangerous medical experiments, for example), in order to benefit a much larger majority" (Scanlon 1998, 208). According to *ex-post* contractualism, this policy is clearly reasonably rejectable. We know that, when we adopt the principle, there will be some individuals who will be coercively subjected to painful medical experiments. After the principle has been adopted, these individuals can reasonably reject the principle given that there are alternative principles under which no one has to bear equally serious burdens. This is why the *ex-post* views get this case intuitively right.

In contrast, the *ex-ante* views seem objectionable here because the policy's burdensomeness expectation for each individual will actually be quite good from the temporally antecedent perspective. For any individual, the policy has two possible outcomes: the very likely outcome that you will benefit from the policy and the very unlikely outcome that you will have to suffer due to the painful medical experiments. If we assign the positive outcome for you the value of 100 and a probability of 0.999 and the bad outcome for you the value of -1000 and the probability of 0.001, then the policy's burdensomeness expectation for you would be (100 x 0.999) + (-1000 x 0.001) = 99.9 + -1 = 98.9. This could well be a better expectation for you than if the policy were not adopted. For this reason, the *ex-ante* versions seem to support objectionable policies that will in the future sacrifice some random individuals for the sake of minor benefits for everyone else.

There are, however, also cases where the *ex-ante* views work better. Consider the case of social risk imposition first described by Johann Frick (2015, 181). In it, a terrible virus will kill one million young children if we do nothing. Fortunately, there are two vaccines we could produce (even if we cannot produce both). Vaccine 1 will save every child from death, but it does not offer complete protection. Every child is still certain to get one of their legs paralyzed. In contrast, the Vaccine 2 gives every child a 0.999 chance of surviving the virus completely unharmed. This means that every child has a 0.001 chance that the Vaccine 2 they take does not work for them in which case the child will sadly die. Furthermore, whichever of these two vaccines we produce, that vaccine will then be given to all the one million children. Which of the vaccines should we then produce?

Here the *ex-post* views have implausible consequences. According to them, we should produce the Vaccine 1 even if we know that the consequence of this policy will be that every one of the million young children will have a paralyzed leg. This is because the *ex-post* views direct us to compare the most serious personal objections the children will eventually have because of the adoption of the two different policies. In the scenario where we produce the Vaccine 1, the strongest objection any one child has to this policy will be that one of their legs gets paralyzed for the rest of their life. In contrast, in the scenario in which we produce the Vaccine 2, the strongest personal objections few children will have to this policy will be that because of the policy they die very young. Because these latter personal objections to the production of the Vaccine 2 are stronger, the principle that we ought to produce the Vaccine 1 cannot be reasonably rejected. This is why the *ex-post* views entail that producing the Vaccine 1 is the right thing to do, but this is intuitively wrong.

In contrast, from the temporally antecedent perspective, it seems like for every individual child the policy that requires producing the Vaccine 2 will offer them better prospects. From this perspective, each child will compare the certainty of having one of their legs paralyzed to getting a 0.999 chance of surviving unharmed and a 0.001 chance of dying. Here most of us would rather take the second option as it seems to us that it is worthwhile to take the small chance of dying in order to escape the certainty of getting one of your legs paralyzed. This is why according to the *ex-ante* views the policy that requires producing the Vaccine 2 cannot be reasonably rejected, which intuitively seems to get things right.

Interestingly, we then have one case that supports the *ex-post* views and one that supports the *ex-ante* versions even if both views cannot be true. As a result of this stand-off, the defenders of the *ex-ante* views have tried to develop their views further so as to find a version that would not entail that we should adopt the policy that permits subjecting some individuals to painful medical experiments in the first case. Likewise, the defenders of the *ex-post* views have tried to develop their views further so as to avoid the objectionable consequence that we should produce the Vaccine 1 in the second case. At this point, it still seems an open question which of these versions of contractualism will turn out to best fit all our carefully considered moral convictions about different cases.

Objections to Contractualism

This final section will finally consider two common objections to contractualism. According to the first and the most frequent one, contractualism is theoretically redundant – a spare wheel in the machine of ethical theorizing that does no genuine work but rather merely whizzes around. According to the second objection, contractualism fails because it cannot ground our obligations to treat non-human animals and cognitively impaired human beings well.

The basic assumption of the first traditional objection is that, as an ethical theory, contractualism is an attempt to explain something important about right and wrong actions – be this (i) which actions are right and wrong, (ii) what makes those actions right and wrong, or (iii) what it is for them to be right and wrong. At least three reasons have then been given for why the contractualist explanations for these things would be bad explanations (Southwood 2009, §3.5).

Firstly, it has been argued that there are always better explanations available (McGinn 1999, 35–36). For example, what explains why harming an innocent baby for fun is wrong must be how much the baby would suffer and how monstrous the action would be. This seems like a much better explanation than the one based on the idea that the potential baby-torturer could not justify his or her actions to others on grounds no one could reasonably reject.

Secondly, it has also been argued that the contractualist explanation would be circular – it is presupposing things that it is supposed to try to explain (Hooker 2000). For example, it could be claimed that, when we consider the different objections individuals can make to different principles, at this point we smuggle in prior features of morality (that is, objections already based on what is right and wrong) that are not validated by the contractualist theory.

Finally, the most common version of the objection is that the contractualist explanations are redundant (Blackburn 1999). Whenever someone can reasonably reject a principle allowing some action, for example, because of the harm that those kinds of actions cause to them, that relevant harm itself seems enough to explain the wrongness of the actions in question. Going in the explanation first from the harms to reasonable rejection and then from reasonable rejection to wrongness is just an unnecessary detour where the middle stage can be cut off without a loss.

Contractualists have, of course, tried to address these concerns. The main line of response is based on the thought that contractualism is needed to do genuine explanatory work to get us from other people's potential personal objections to your actions to the reasons you personally have for acting in certain ways (Ridge 2001). For example, in the social risk imposition case described above, the theory is needed to explain how we get from what different objections each child would have for producing either the Vaccine 1 or the Vaccine 2 to the conclusion that you have most reason to produce the Vaccine 2, as that is not obvious before we compare the different objections the children have. Whether this line of response works is something that continues to divide opinions.

The second common objection concerns our duties towards non-human animals and cognitively limited human beings such as infants or severely disabled individuals (Phillips 1998; Hooker 2000, §2.9; Nussbaum 2006). The problem is that Scanlon's contractualism understands right and wrong in terms of the reasons individuals have for objecting to different principles. Yet, having a reason to make an objection seems to require (i) being able to evaluate how strong reasons different considerations provide for you to make objections, and (ii) being able to make the objection in question. These things, however, are something that non-human animals and cognitive limited human beings cannot do. For this reason, both groups do not seem to have any reasons to object to the principles that would permit mistreating them, and so contractualism seems unable to explain our obligations towards non-human animals and cognitively limited human beings.

Several contractualist responses to this objection have been offered, but these responses remain controversial too. Firstly, it could be suggested that many of us who can evaluate and make objections deeply care about non-human animals and cognitively impaired human beings, and so at least we have reasons to reject the principles that permit mistreating them. The problem is that this response only gives non-human animals and cognitively limited human beings an objectionably indirect moral status – they would only matter morally because they happen to matter to us (Hooker 2000, 68). It has also been suggested that we can imagine that non-human animals and cognitively limited human beings would have trustees, who could then reasonably reject principles on their behalf (Scanlon 1998, §4.8). It has likewise been thought that contractualism is only attempting to capture a core part of interpersonal morality – what we owe to each other – and so it is not surprise that some other theory is required to explain the other parts of the moral reality (ibid.).

Conclusion

This chapter has then tried to introduce contractualism in a positive light. Whilst doing so, it has also tried to be open about the main problems of the theory. The first section explained how, according to contractualism, what is right and wrong is determined by the principles no one could reasonably reject, where these principles are a function of the personal objections individuals can make to different alternatives. This section also introduced how, according to Scanlon, we ought to follow the non-rejectable principles because of the valuable personal relationships this enables us to form with others.

The second section then used the television transmitter case to illustrate how contractualism differs from classical utilitarianism. After this, I explained how there are two different versions of contractualism depending on from which temporal perspective we consider the relevant objections. Finally, the last section introduced the redundancy objection and the challenges contractualists face when they try to explain our obligations towards non-human animals and cognitively impaired human beings. I can only hope that this brief introduction inspires a new generation of ethicists to explore the contractualist tradition further.

Further Reading

Ashford, Elizabeth. 2003. "The Demandingness of Scanlon's Contractualism." *Ethics* 113, no. 2: 273–302.

Blackburn, Simon. 1999. "Am I Right?" *New York Times*, 21 February: 24.

Frick, Johann. 2015. "Contractualism and Social Risk." *Philosophy & Public Affairs* 43, no. 3: 175–223.

Gauthier, David. 1986. *Morals by Agreement*. Oxford: Oxford University Press.

Hobbes, Thomas. 1996 [1651]. *The Leviathan*. Edited by Richard Tuck. Cambridge: Cambridge University Press.

Hooker, Brad 2000. *Ideal Code, Real World*. Oxford: Oxford University Press.

Hooker, Brad. 2003. "Contractualism, Spare wheel, Aggregation." In *Scanlon and Contractualism*. Edited by Matt Matravers. London: Frank Cass, 53–76.

Kumar, Rahul. 1999. "Defending the Moral Moderate: Contractualism and Common Sense." *Philosophy and Public Affairs* 28, no. 4: 275–309.

Locke, John. 2002 [1689]. *Two Treatise of Government*. Edited by Peter Laslett. Cambridge: Cambridge University Press.

McGinn, Colin. 1999. "Reasons and Unreasons." *The New Republic*, May 24: 34–38.

Nussbaum, Martha. 2006. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, MA: Harvard University Press.

Parfit, Derek. 2011. *On What Matters*, Vol. 1. Oxford: Oxford University Press.

Pettit, Philip. 2000. “A Consequentialist Perspective on Contractualism.” *Theoria* 66, no. 3: 228–245.

Phillips, David. 1998. “Contractualism and Moral Status.” *Social Theory and Practice* 24, no. 2: 183–204.

Plato. 2000 [circa 380BC]. *The Republic*. Edited by G.R.F Ferrari, translated by Tom Griffin. Cambridge: Cambridge University Press.

Prichard, H.A. 1912. “Does Moral Philosophy Rest on a Mistake?” *Mind* 21, no. 81: 21–37.

Rawls, John. 1972. *A Theory of Justice*. Cambridge, MA: Harvard University Press.

Ridge, Michael. 2001. “Saving Scanlon: Contractualism and Agent-Relativity. *Journal of Political Philosophy* 9, no. 4: 472–481.

Rousseau, Jean-Jacques. 1997 [1762]. *The Social Contract and Other Later Political Writings*. Edited by Victor Gourevitch. Cambridge: Cambridge University Press.

Scanlon, T.M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Scanlon, T.M. 2003. “Rawls on Justification.” In *The Cambridge Companion to Rawls*. Edited by Samuel Freeman. Cambridge: Cambridge University Press, 139–167.

Southwood, Nicholas. 2009. “Moral contractualism.” *Philosophy Compass* 4, no. 6: 926–936.

Southwood, Nicholas. 2010. *Contractualism & the Foundations of Morality*. Oxford: Oxford University Press.

Suikkanen, Jussi. 2019. "Ex Ante and Ex Post Contractualism: A Synthesis." *Journal of Ethics* 23, no. 1: 77–98.

Suikkanen, Jussi. 2020. *Contractualism*. Cambridge: Cambridge University Press.

Wallace, Jay. 2002. "Scanlon's Contractualism." *Ethics* 112, no. 3: 429–470.