

## 1 Introduction

This Element has two aims. It first intends to bring its reader up to speed with the recent developments in contractualist ethical theorising. The idea of understanding morality in terms of a contract is old. It was described in Plato's masterpiece the *Republic* about 2,400 years ago (Plato, *Republic*: 358e–359b). Different social contract theories were also popular during the early modern period (1500–1800) when Thomas Hobbes (1651/1996), John Locke (1689/2002), Jean-Jacques Rousseau (1762/1997) and many others developed these theories further.

This Element, however, focusses only on the most recent discussions of contractualism. They tend to begin with a version of contractualism, which T. M. Scanlon introduced in 'Contractualism and Utilitarianism' (1982) and then developed further in *What We Owe to Each Other* (1998). Section 2 outlines Scanlon's theory, which is based on the moral principles no one could reasonably reject. Section 3 then explains the main objections to Scanlon's contractualism. It has been argued that his theory has implausible consequences with respect to how different-sized groups, non-human animals and cognitively limited human beings are to be treated. It has also been claimed that Scanlon's contractualism cannot vindicate the so-called deontic distinctions and that it is merely a redundant theoretical device that is not needed for understanding right and wrong.

Scanlon's contractualism is, of course, not the only contemporary version of contractualism. Thus, in Section 5.2, I outline Derek Parfit's (2011) Kantian contractualism, Nicholas Southwood's (2010) deliberative contractualism and David Gauthier's (1986) contractarianism. I explain both what is common to these contractualist views and how they differ from one another. Thus, by reading Sections 2–3 and 5.2 (as well as Section 5.4, which discusses contractualist responses to the main objections to contractualism), the reader should be able to understand the key issues in the recent debates about contractualism in normative ethics.

This Element, however, also makes an original contribution. It begins from the observation that ethicists have recently come to understand contractualism's main competitor, consequentialism, in a new way. Instead of taking consequentialism to be a distinct view of which actions are right and wrong, many have started to think of it as a flexible framework in which different ethical views can be formulated as versions of consequentialism. The so-called consequentialisers have tried to show that, for every plausible first-order ethical view, there is a version of consequentialism that

is co-extensive with it (Dreier, 2011). This so-called consequentialising project promises that the previous debates concerning which ethical theory is true can now be translated into disagreements over which version of consequentialism is correct. Section 4 summarises these recent radical developments.

Section 5 then suggests that we should revise our understanding of contractualism in the same way. It argues that we should draw a distinction between different versions of contractualism (including Scanlon's contractualism) and the general framework of contractualism in which different versions of contractualism can be formulated. This allows us both to explain more precisely how contractualist ethical theories differ from consequentialist theories and to state the differences between different versions of contractualism more accurately.

Section 5.3 furthermore argues that the contractualist framework is just as flexible as the consequentialist one. It turns out that, for every plausible ethical view, there is likewise a version of contractualism that is co-extensive with it. Section 5.4 goes on to explain how this result has radical consequences for our understanding of the traditional objections to contractualism. The previous disagreements over these objections can now be understood as internal questions concerning which version of contractualism is most plausible.

Finally, Section 6 concludes by first summarising the main discussions of this Element. It also considers what consequences these new developments in our understanding of the traditional ethical theories will have for future debates in normative ethics. I suggest that we should no longer argue over which ethical theory is in some general sense true or correct. This is because it is likely that there is a version of every major ethical theory that corresponds to the correct view of which actions are right and wrong. Rather, we should take seriously the idea that, when we try to solve difficult new ethical problems, these theories as general frameworks direct our attention to different kinds of considerations. The consequentialist framework makes us investigate which theories of value are plausible, whereas the contractualist framework guides us to consider what kinds of views of reasons or rationality we should accept. My pragmatic suggestion, then, is that we should accept ethical theories on the basis of how helpful the ways of thinking about new ethical problems they provide are. I also recommend accepting a form of pluralism. In advance, we have no reason to think that one framework of ethical theorising will be the most fruitful one in every context. Rather, different ethical views may well enable us to investigate different problem cases in more illuminating ways.

## 2 Scanlon's Contractualism

This section begins by reconstructing Scanlon's contractualism on the basis of how it was presented in *What We Owe to Each Other*.<sup>1</sup> Scanlon's starting point is that any plausible ethical theory ought to answer two questions at the same time (147–9). Such a theory should first be able to provide a plausible view of which actions are right and wrong. After all, any plausible ethical theory should entail, for example, that we ought to keep our promises. At the same time, a plausible ethical theory should also be able to explain *why* we ought to do what is right and avoid acting wrongly. Scanlon does not understand this second question as a psychological question that is about which mental states move us to do the right thing but rather as a question concerning normative reasons (148). What are the '(almost?) always' overriding reasons we have for acting morally?

Providing a satisfactory answer to this second question is difficult due to 'Prichard's dilemma' (149–53; Prichard, 1912). We cannot merely claim that we should do right actions because they are right. This answer would merely take the reason-providing force of morality for granted. We cannot, however, offer merely non-moral reasons for why we ought to do the right thing either. Purely self-interested reasons based, for example, on the costs of being ostracised by others would simply not be the right kinds of reasons. They would not be the kinds of reasons we intuitively take ourselves to have for acting morally. The challenge is thus to provide an account of our reasons for doing the right thing that is at the same time both informative and recognisably moral in nature.

Scanlon's contractualism thus attempts to describe the nature of rightness and wrongness in a way that both (1) makes sense of which actions are right and wrong and (2) explains why we ought to do right actions understood in that way.<sup>2</sup> He believes that we should begin from what we can justify to others on

<sup>1</sup> All unattributed references are to Scanlon (1998). In addition to contractualism, Scanlon's book addresses a number of other topics including the nature of value (of which he gives a buck-passing account), the nature of well-being and the role of teleology in practical reasoning. It is true that these views also to a large extent underwrite Scanlon's own version of contractualism, and so Scanlon's discussions of these topics are essential reading for anyone interested in Scanlon's contractualism.

<sup>2</sup> In *What We Owe to Each Other*, Scanlon explicitly presented his view as a metaethical or metaphysical theory of the nature of the properties of rightness and wrongness (12) that constitutes the subject matter of our judgements of right and wrong (2). He then considers being able to make sense of which actions are right and wrong and being able to explain the reasons for acting rightly to be the standard by which to evaluate the previous kind of metaphysical/metaethical theories of the nature of right and wrong. Later on, he acknowledged that he 'should have avoided describing contractualism as an account of the property of moral wrongness' in the previous way (Scanlon, 2004: 137). His more recent proposal is to understand contractualism as a view of which standards we should use to evaluate our actions and what reasons support these

grounds they could not reasonably reject (153–5).<sup>3</sup> He then argues that focussing on this type of idealised justification will enable us to explain both which actions are right and what reasons we have for doing those actions. Next, in Section 2.1, I focus on how Scanlon derives an account of right and wrong from the previous type of justification. Section 2.2 then outlines how that account can be used to answer the question of why we should be moral.

## 2.1 Which Actions Are Right?

How could we then get from justification to an account of which actions are right and wrong? Let us use a simplistic version of contractualism as a starting point. When our actions affect others, they often question what we have done to them. In such cases, it clearly helps if you can explain *why* you did what you did in a way that the other person accepts. If others accept your reasons, they will no longer disapprove of what you did, which leads to pleasant harmony (154).

As a consequence, you might think that we could explain what is right and wrong by relying on which justifications others actually accept. This starting point would allow us to explain what reasons we have for doing the right thing – right actions will help us to avoid conflicts and they also lead to harmonious social interaction. Starting from actual justification would furthermore lead to a view of which actions are right and wrong. Right actions would be precisely the ones we can justify to others on grounds they accept.

This proposal admittedly has advantages. Firstly, whether an action is right becomes a tractable empirical question concerning which justifications others accept on the basis of what they ultimately care about. Secondly, as we have just seen, the view offers at least some kind of an explanation of why we ought to do the actions that are right in the relevant sense. The reason for doing those actions is that acting in those ways makes our interactions with other people run more smoothly, which often is in our selfish interests.

According to Scanlon, the previous proposal has one fundamental flaw (155). The problem is that it leads to an implausible account of which actions are right and wrong. Some individuals, as a matter of fact, do not think that their own interests are as important as the interests of others. These individuals are so modest that they are often happy to receive less than others, and sometimes they

---

standards (ibid.). This fits with the description of his project given in Sections 2.1 and 2.2 in this Element.

<sup>3</sup> The idea of justifiability to others thus plays a crucial role in Scanlon's own version of contractualism. In Section 5.1, I have suggested that, despite this, this notion should not be thought to be what makes a theory a version of contractualism. This is because there are versions of consequentialism that equally rely on the idea (see Hooker, 2000: 99–102) and versions of contractualism that do not (Southwood, 2010).

do not even object to being treated badly. The previous simple view would entail that it would be morally right to give these modest, self-sacrificing individuals less than others and sometimes it would not even be wrong to mistreat these individuals as long as they accept such mistreatment. For this simple reason, the previous simple view fails as a view of which actions are right.

Right and wrong cannot, therefore, be a matter of which actions can be justified to others on grounds they actually accept. We can then think of Scanlon's own theory as an attempt to formulate a more sophisticated view that could avoid the problems of the previous simple view. Scanlon believes that we can still make sense of right and wrong in terms of what can be justified to others. However, instead of actual justification, according to him, we will have to focus on which actions we can justify to others on grounds *they could not reasonably reject*. Even if the self-sacrificing individuals accepted our attempts to justify the ways in which we mistreat them, they still *could* reasonably reject those justifications. This is why the reformulated view does not entail that treating self-sacrificing individuals badly is right.

According to Scanlon's contractualism, therefore, actions are right insofar as they are authorised by the set of moral principles no one could reasonably reject (and wrong insofar as they are forbidden by those principles) (4–5, 153). We, of course, want to know next what would make a set of moral principles one that could not be reasonably rejected. This is easiest to explain with a simple case (195).<sup>4</sup> Imagine that you and I are considering whether acting in a certain way ( $\varphi$ ing) would be wrong in the circumstances (C) we are in. According to Scanlon, we then first need to compare different principles that could be adopted to govern  $\varphi$ ing in different C-like circumstances.

The relevant principles we could adopt should not be understood merely as simple, absolute and exceptionless principles of the form 'Always  $\varphi$ !' or 'Never  $\varphi$ '. Rather, it is better to understand them as complex elements that could be added to our overall moral sensitivities, where these elements determine in part whether  $\varphi$ ing would be seen as the thing to do in different situations. These elements of our moral sensitivities determine when the relevant rule applies, what exceptions it allows and what strength it has in different cases of conflict (197–202; see also Hooker, 2000: 90–1).

Let us then consider two alternative principles – one that would authorise us to  $\varphi$  (hereafter 'the permitting rule') and one that would forbid  $\varphi$ ing ('the forbidding rule'). According to Scanlon, the next step is to examine what consequences the general adoption of these principles would have for different

<sup>4</sup> Here, for the sake of simplicity and for illustrative purposes, I focus on comparing individual principles rather than whole sets of principles (as properly speaking I should).

individuals (195, 202–6). After all, the general adoption of the permitting rule and the general adoption of the forbidding rule would mean that people would act in different ways. As a result, which principle we adopted would eventually make a difference to what kinds of lives we, as individuals, would come to live.

For example, imagine that  $\varphi$ ing is something that is mildly amusing to do to others, trivially harmful for the victims in individual cases and seriously harmful for people who are repeatedly treated in that way. In this case, it is easy to see that people come to live very different kinds of lives depending on whether the permitting rule or the forbidding rule is adopted. In the scenario in which the permitting rule is adopted, some people benefit a little because they are able to  $\varphi$ , whereas others suffer the trivial harms that together constitute a serious burden to them over time. In contrast, when the forbidding rule is adopted both the trivial benefits and the resulting serious harms go away.

Scanlon calls the lives which individuals come to live under the alternative principles their ‘standpoints’ (202). Furthermore, we can call ‘burdens’ the features of these standpoints that make them less choice-worthy as lives (195). These burdensome features of the lives of different individuals then provide those individuals with reasons to make objections to the principles that are causally responsible for the burdens in question. For instance, in the previous example, the individuals who repeatedly suffer a minor loss because others  $\varphi$  can use the burden of having to suffer all those losses as an objection to the adoption of the permitting rule.

Scanlon’s theory of reasonable rejectability still needs two further elements. Firstly, we need an account of which features of the lives of different individuals ground legitimate objections to different moral principles (213–18). Here, Scanlon seems to rule out two kinds of considerations and explicitly include at least two others. Firstly, on the pain of vicious circularity and redundancy (see Section 3.4.1), the fact that others would wrong you by  $\varphi$ ing cannot be a feature of your standpoint that enables you to object to the permitting rule. After all, when you give an account of what is right and wrong, your account cannot rely on rightness and wrongness of different actions. We can call this the ‘deontic restriction’ on the reasons to reject principles (216; see also Parfit, 2003: 370). Secondly, your reasons for making objections to a given moral principle must be features of your own life. Your reason to object to a principle cannot be based on the combined, aggregate burden that both you and some other people are burdened in a certain way by the adoption of that principle.<sup>5</sup> Call this the

<sup>5</sup> This condition does not preclude counting the empathic pain I would feel because of someone else’s suffering as a burden which can ground an objection to a moral principle. However, usually such an objection is considered to be weaker than the one based on the original suffering that causes me to feel the empathic pain in the first place.

‘individualist restriction’ on the reasons to reject principles (229; see also Ridge, 2001).

What would then count as a legitimate reason? Firstly, Scanlon accepts that, if a principle lowers your level of well-being, you can make an objection to that principle on that basis (216). However, he denies that well-being would be the only ground for an objection. He also rejects the idea that well-being would provide a master currency that could be used to weigh different objections. Scanlon instead claims that we should consider more concrete qualities of individual lives as the relevant burdens that count as ‘generic’ reasons to make objections to different principles. These more concrete personal considerations include, for example, bodily harms, not being able to rely on the assurances of others and not having control over what happens to one’s own body (204). These considerations count as generic reasons because their status as a reason does not depend on particular aims or preferences of individuals.

Scanlon also accepts that, when we compare moral principles, individuals can also make objections on the basis of other moral rules that are not at issue in the comparison in question (214). When we evaluate one principle, say, governing our obligation to help those in need, we need to take into account how the adoption of that principle would affect what people are entitled to on the basis of other moral principles. To assess one principle, we thus need to hold at least some other principles fixed. Of course, those other principles can also be evaluated in the contractualist comparisons of different principles but, when we do so, we again need to hold at least some other principles fixed.

We then have a sense of what kinds of objections individuals can make to different principles. Which principles cannot then be reasonably rejected? Here, we must consider the strongest personal objections which individuals can make to the different alternatives. We must first identify the individuals who would be personally burdened the most by the general adoption of every principle that could potentially govern a certain kind of situation. This is because these individuals could make the strongest objections to those principles. We then compare pairwise the strength of the personal reasons which these individuals have for rejecting the relevant principles (195). The individuals who have the stronger reasons to make objections in these comparisons can reasonably reject the principles that are making their lives burdensome. In contrast, the one person whose objection is grounded on the weakest personal reason cannot do so, because some individuals would always have stronger personal objections to all other moral principles that could potentially govern the relevant type of situations (229).

To see how this works, recall that in the earlier example the general adoption of the permitting principle would provide minor benefits to some even if it



would also make the lives of many people more burdensome over time. In this case, those individuals could make a serious objection to the permitting principle, whereas no one could make an equally strong objection to the forbidding principle (no one would lose more than a minor benefit if that principle were adopted). In Scanlon's view, the forbidding principle could thus not be reasonably rejected because there would be stronger personal objections to the alternative. Moral principles that no one could reasonably reject are therefore more generally such that there are stronger personal reasons to reject all the principles that could replace them.

If there are many different principles that could govern certain kinds of situations to which the most serious personal objections are equally the least serious, we need to consider to which of these principles the fewest individuals could make those objections (230–4). If there still are many principles to which equally many individuals could make the most serious personal objection, we need to compare the second most serious objections in the same way. If these objections and their number are also tied, we keep moving to the less serious objections until we find a unique principle that could not be reasonably rejected because all other principles would produce more serious unnecessary burdens to some individuals.

Scanlon's claim, then, is that understanding right and wrong in terms of the previous non-rejectable principles gives us an extensionally adequate account of right and wrong. The view would fit our moral intuitions as the following illustrations are supposed to show.

Firstly, the previous theory promises to make sense of the wrongness of promise-breaking (295–327). When you make a promise, you knowingly create expectations in the other person to whom you make the promise. Breaking the promise will mean that those expectations will not be met. Consider, then, two potential principles that could govern promise-breaking: a forbidding rule and a permitting one. If we all adopted the latter rule and thus believed that there is nothing wrong with breaking a promise, many more promises would not be kept. This would burden many individuals because their expectations, which the relevant promises had created, would not be met. In this scenario, we would also lose the ability to rely on the assurances of others. We would just not believe them when they would try to make a promise. This presumably would make our lives more difficult in a number of ways.

In contrast, there would not be equally strong objections to the forbidding rule. It is difficult to see how the burden of having to keep a promise could be claimed to be unreasonable. This is because there is always an easy way to avoid that burden – all you need to do is not to make the promise in the first place. We must, of course, recognise that in some situations it is permissible to break



a promise. For example, you are not required to keep your promise to meet your friend at a café if keeping that promise would require not saving the easily saveable drowning child on your way. One promise of contractualism is that it will help us to understand this type of exception to the promise-keeping principle. After all, as the previous example shows, the most stringent principles could be reasonably rejected because they would cause certain serious personal burdens to some individuals, which no one would need to experience under the less stringent principles that contain some exceptions to the general requirement.

Scanlon also argues that his contractualism fits our intuitions about the so-called *anti-utilitarian protections*, according to which morality does not allow us to sacrifice individuals for the sake of the general good. Consider Scanlon's own example in which millions of people are enjoying watching the World Cup final live on television (235). Unfortunately, a piece of equipment falls on Jones in the transmitter room. As a result, Jones is receiving extremely painful electric shocks. The only way to help him is to cut off the transmission for 15 minutes. This would mean denying millions of people the pleasure of watching the final.

Many people agree that the right thing to do is to save Jones. Scanlon's contractualism promises to tell us just why we ought to do so. We must again begin by comparing two principles that could govern the situation. The strongest objection any person could make to the principle that requires saving Jones is that this principle would require them to give up the trivial benefit of being able to follow the match. Yet Jones clearly has a much stronger objection to the principle that would forbid saving him based on the painful electric shocks. Given that the 'Save Jones!' principle would not cause equally serious burdens to anyone, that principle could not be reasonably rejected. Contractualism, due to the individualist restriction, thus does not allow us to sacrifice individuals like Jones for the sake of the general good.<sup>6</sup>

These are, of course, just two cases in which Scanlon's contractualism seems to have appealing consequences. Scanlon's hope, however, is that, whatever intuitively wrong kind of an action we take, we can always identify some serious personal burdens on the basis of which some individuals could reasonably reject the principles that would permit that kind of action. If we are always able to identify the previous kind of burden, then considering which principles could not be reasonably rejected seems to offer us a compelling account of right and wrong.

<sup>6</sup> In Scanlon's view, objections are always compared one-to-one. This entails that the only way that the fact that an option would burden many people could be taken into account would be if one person could include the other people's burdens into their objection. The individualist restriction rules this out.

## 2.2 Why Be Moral?

As discussed, Scanlon believes that a plausible ethical theory should be able to explain why we ought to do the right thing when we, at the same time, use that same theory to understand which actions are right.<sup>7</sup> Let us return to the simplistic version of contractualism introduced in Section 2.1. According to it, which actions are right is a function of which justifications others are actually willing to accept. Even if this view fails as an account of which actions are right, it admittedly provides a compelling explanation of why we ought to do right actions where rightness is understood in terms of actual justifiability. Such actions tend to lead to harmony, co-operation and fewer conflicts. Insofar as we then have reasons to live in such a pleasant way with others, we have reasons to do actions that we can justify to others on grounds they accept.

This account of why we ought to be moral is not available for Scanlon, because, in his view, what we can justify to others on grounds they actually accept is not relevant at all. Instead, we are now considering actions that are authorised by the principles no one could reasonably reject. However, even if an action were authorised by those principles, the person affected by the action might not actually accept the justification you offer to them on grounds they could not reasonably reject (and so they would be acting unreasonably, as we often do). This is why following the principles no one could reasonably reject need not lead to actual harmony. The ability to live in harmony with others cannot therefore be our only reason to do the actions that are authorised by the non-rejectable principles.

What, then, are the reasons to follow those principles? Here, three types of reasons are relevant. Firstly, there are almost always concrete first-order reasons not to do actions that are forbidden by the non-rejectable principles (156; see also Stratton-Lake, 2003: 74; Wallace, 2002: 455). The principles that would allow doing those actions could always be reasonably rejected because of some concrete burdens that some individuals would have to bear as a consequence of those principles. Principles that authorised physical attacks would, for example, make some people's lives more burdensome because they would have to suffer the consequences of the attacks which these principles would permit. This suggests that there are always concrete, ordinary first-order reasons not to do

<sup>7</sup> So, for example, Scanlon recognises that utilitarianism can, in many cases, provide a plausible account of which actions are wrong. For example, utilitarianism correctly entails that we should help people who live in extreme poverty (152). However, Scanlon objects that this is not enough to make utilitarianism a plausible account of what it is for actions to be right and wrong, because utilitarianism fails as an account of the reason-giving force of the idea of moral wrongness (152–3); that is, a plausible account of right and wrong also has to be able to provide a compelling account of just why we ought to avoid acting wrongly.