

CONTRACTUALISM

Rahul Kumar

What is contractualism?

Contractualism is a non-consequentialist moral theory first advanced by T. M. Scanlon in his influential article, “Contractualism and Utilitarianism” (1982), and later developed in detail in his book, *What We Owe to Each Other* (1998). Its distinctive central claim is that one person morally wrongs another by treating him in a way whose permissibility is not justifiable to him on grounds he cannot reasonably reject.

The name, “contractualism,” suggests an account belonging to a tradition of thinking about morality associated with Hobbes. Its animating idea is that valid moral norms are those that can be thought of as the object of a hypothetical agreement, reached through a process of self-interested bargaining, among all those to whom they apply. In this respect, the name is misleading. Scanlonian contractualism’s roots lie not in Hobbes, but in the social contract tradition of Rousseau and Kant, one that treats what motivates the parties to the hypothetical agreement as the appeal of living in community with others on a basis of mutual respect for one another, along the lines of what Kant calls a the “kingdom of ends,” a “systematic union of different rational beings through common laws” (Kant 1997: 4:433).

This strand of the social contract tradition is one that John Rawls brilliantly develops in his *A Theory of Justice* (1971). Scanlon’s contractualism is importantly indebted to Rawls’s work, but is not helpfully understood as an extension of it. Rawls’s concern is with the nature of justice, and in particular, how a society’s basic institutions ought to be regulated so as to enable ongoing social cooperation on terms of fair reciprocity. Contractualism, on the other hand, focuses on a different, and more general, question: What makes it the case that a person is morally wronged by another’s treating her in a certain way?

In particular, contractualism has two principal theoretical aims. The first is that of providing a plausible alternative to consequentialist accounts of the subject matter of moral judgments. Consequentialists hold, roughly, that a person’s conduct is morally wrong if it is other than the way she ought to conduct herself if she is to best contribute to the promotion of aggregate well-being. Many find

some form of consequentialism intuitively compelling, despite the difficulty of defending, on consequentialist grounds, certain familiar moral convictions concerning how it is and is not morally permissible for individuals to treat one another. Contractualism offers itself as an account that is at least as intuitively appealing as any form of consequentialism, but one that provides a better basis for explicating the justification for common-sense moral convictions.

Its second aim has to do with the nature of moral motivation. It is prompted by the thought that knowing a certain way of conducting oneself to be morally wrong is not a kind of esoteric knowledge, to which a person could be blamelessly indifferent. Rather, those who take moral considerations to be important to take into account in their practical deliberations understand them as considerations that anyone capable of being guided by reasons should consider specially important. The question is, can this special importance be defended as appropriate? Part of the contractualist project is to defend a positive answer to this question, one that connects its understanding of the subject matter of moral judgments with the appropriateness of taking such judgments, so understood, to be of special importance.

Contractualism and the value of human life

Key to what contractualism says about both issues is a particular understanding of what respect for the value of human life requires. How this value figures in the account is easily misunderstood if it is simply assumed that the appropriate way to respond to something of value is to take it to be something one should aim to promote, or bring about more of. Reflection on different examples of things of value suggests, rather, that there is no one way of appropriately responding to something of value (Scanlon 1998: 87–103). A correct appreciation of the value of suffering plausibly requires that one strive to decrease the amount of suffering in the world. But it is a misunderstanding of the reasons one has to act, feel, and think in certain ways if one appropriately values friendship to take the appropriate appreciation of its value to require one to aim to bring it about that there are lots and lots of friendships in the world.

This way of thinking about the requirements for a proper appreciation, or respect for, something's value informs the contractualist rejection of the idea, associated with consequentialism, that respect for the value of human life requires taking oneself to have reason to contribute to bringing it about that there is more human life in the world. What it requires, rather, is a recognition, in one's practical thinking, of other human beings as creatures who have reasons for wanting to live and for having their lives go better, and who have a capacity to both assess reasons and justifications and select among the various ways there is reason to want a life to go, enabling them to live actively self-governed lives (Scanlon 1998: 103–7).

Respect for the value of others, then, requires that individuals be sensitive, in their practical thinking, to others' reasons for wanting, caring about, and pursuing certain things. But how exactly does respect for another's value require that another's reason for, say, caring about something figures in one's practical thinking? To paraphrase a point Scanlon makes in another context, a person's reasons for having a decent diet may well entitle him to claim the assistance of others in making that possible, while his equally strong reasons for building a monument to his God do not, even if he is quite willing to forgo the former in order to advance the latter (which is compatible with his still having a claim against others that they not interfere with his building it) (Scanlon 2003: 74).

The contractualist proposal is that how another's reasons, in a particular situation, ought to figure in one's practical thinking is fixed by a principle no one could reasonably reject for the regulation of that type of situation. That part of morality concerned with what we owe to each other as a matter of respect for the value of one another as persons is constituted, on this account, by an indefinite number of such principles, each for the regulation of a certain type of situation. For any situation, the relevant principle will spell out how other individuals are entitled, or may legitimately expect, to have their reasons figure in one's practical thinking, and how one is entitled to have one's own reasons figure in their practical thinking.

Grasping a relevant principle for the regulation of the type of situation one finds oneself in is not a matter of understanding a wholly stable rule that one knows how to apply. Rather, it is to exhibit a complex understanding of the reasons why in general it is important that there be constraints on how persons act in this type of situation. In particular, grasping a principle is a matter of understanding why there are good reasons for requiring that individuals, in determining what it is permissible to do in such situations, take certain reasons into account in their thinking, but not others. Judging a certain course of action permissible or impermissible under the circumstances is, on this account, then, a matter of assessing what there is or is not decisive reason to do, as informed by an understanding (which need not be readily articulable) of the moral rationale that supplies the relevant principle's evaluative point.

Scanlon gives several examples of contractualist principles, such as those regulating promissory obligations (Scanlon 1998: 295–327). Individuals have good reason to want to be able to sometimes assure one another of their future conduct by undertaking a voluntary obligation, like a promise, that binds one to do what one has promised to do unless released by the promisee. Displaying an understanding of the relevant principle in one's practical thinking is a matter of understanding when it is the case that one's normative situation has changed because one has successfully undertaken a promissory obligation; what one has reason to do and not to do because doing, or not doing, so would constitute violating the promise; when the circumstances justify the judgment that one has grounds to take the promise to have lapsed, that one is justified in breaking one's promise, etc.

What respect for the value of others as human beings requires on this account, then, is that individuals regulate their practical thinking by principles no one can reasonably reject. One *wrongs* another, or fails to comply with the requirements of respect for her value as a human being, when one fails to give her interests the kind of consideration in one's practical thinking that the relevant principle for the regulation of the situation requires.

Scanlon is clear in presenting contractualism as an account of the basis of those norms that constitute one important aspect of morality, those having to do with "what we owe to each other"; many moral norms, such as those having to do with the treatment of animals, fall outside this domain (Scanlon 1998: 171–7). This domain restriction is sometimes treated as evidence of contractualism's theoretical modesty. But it should now be clear that it is motivated by the recognition that what contractualism characterizes is just reasoning about what respect for one particularly important value – namely, that of human life – requires. There is every reason to think that the kinds of standards anyone takes to be constitutive of his morality will draw upon not just this value, but also a plurality of other, distinct, values.

Morality's authority and the value of mutual recognition

Why hold that what respect for the value of human life requires is that one be guided in one's practical thinking by principles no one could reasonably reject? A rationale for this connection starts to emerge from reflection on the full statement of the contractualist criterion of moral wrongness:

An act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement.

(Scanlon 1998: 153)

No one, that is, who is motivated by a concern to be guided in his practical thinking by principles no one could reasonably reject.

The characterization of valid principles here is as the objects of a *hypothetical*, not actual, agreement between all persons. That the agreement is hypothetical, not actual, does not, however, make the appeal to the idea of agreement spurious. What it draws attention to is an idea that lies at the heart of contractualism: that a person who takes complying with morality's requirements to be specially important is one who attaches importance to his conduct being *justifiable* to any other on grounds that that person cannot reasonably reject.

To care about the justifiability of one's conduct to another is not the same as the concern that others in fact accept one's conduct as justified. That they do so

is desirable, but as Scanlon notes, people often accept as justified that which they have good reason not to accept as justified (Scanlon 1998: 155). What principally matters to one who takes his conduct to be justifiable to any other is that others have good reason to accept his conduct as justified, even if they don't see that.

A common criticism frequently pressed against contractualism is that to care about the justifiability of one's conduct to others amounts to nothing more than a concern that one's conduct be morally justified. The crucial question, that of what the criterion is for evaluating moral justifiability, is one contractualism fails to address. But this criticism misses its mark. The central idea of contractualism is not that (a) conduct or a principle that is justifiable to others is so because it is justified, but the striking idea that (b) a principle or conduct that is morally justified is so because it is justifiable to others on grounds no one can reasonably reject. Being *justifiable to others* is itself a substantive view about what moral justification requires.

In particular, for a principle to be justifiable to another on grounds that a person cannot reasonably reject is for that principle to be one that he has reason to *authorize* or *license* others to be guided by in their practical deliberations. What it is for a principle to be one that a person has reason to authorize, or license, others to be guided by in their practical thinking is helpfully fleshed out by the metaphor of principles being the object of a hypothetical agreement. It invites us to think of a valid principle for the regulation of a certain type of situation as one that is worked out through a process in which all those within the moral domain (anyone, on this view, with a capacity for rational self-governance) participate as co-deliberators, each with an equal voice, in working out what the principle ought to be.

In evaluating candidate principles, each compares the implications for himself of each candidate principle being the principle regulating the type of situation in question. But the assessment by each of which principle she has reason to agree to is not based solely on which principle's implications are most favorable to her. Because each cares about the justifiability of his conduct to others, each takes into account in his own assessment of whether a candidate principle is one that he has reason to agree to both his own reasons for favoring or objecting to the principle and the reasons of other individuals for objecting to or favoring it. Convergence on a principle, one that each, as assessed from his or her own point of view, has reason to agree to, is arrived at through a process, one might say, of individuals comparing the strength of one another's objections to different principles. The converged-upon principle will be one of which it is true that the most serious objection that can be pressed from a particular point of view against it is not as strong as the objections, from other points of view, to any other candidate principle.

To say that a principle no one could reasonably reject is one that individuals authorize, or license, one another to be guided by in their practical thinking is to say, then, that the principle is one that any person, assessing it from his or her

own point of view, has reason to agree to for the regulation of the type of situation in question. By being guided by such principles in how one relates to others, one's conduct manifests respect for the value of others as human beings. For to be guided by principles that others authorize, or license, one to be guided by is to recognize others in one's practical thought as rationally self-governing creatures, capable of assessing reasons and justifications concerning principles regulating how it is permissible for individuals to relate to one another. As Kant remarked, making an analogous point, the dignity of a human being lies in the fact that he or she stands to the moral law as both subject and legislator.

Though respect for the value of human life requires that one be guided in one's practical thought by principles no one could reasonably reject, it is not this requirement that contractualism identifies as what moves the morally motivated. Rather, it holds that what moves one who takes the justifiability of his conduct to others to be specially important is the appeal of an ideal of moral community in which individuals stand in a relationship to one another of *mutual recognition*, living on terms of "unity with one's fellow creatures" (Scanlon 1998: 154). Individuals stand in this kind of intrinsically valuable relationship to one another when they regulate their conduct by principles justifiable to others on grounds no one could reasonably reject. The pain of guilt that we associate with having wronged another, Scanlon suggests, is best understood as a feeling of estrangement from another as a result of having violated the requirements of this kind of valuable relation with others (Scanlon 1998: 162). Living morally, for those who take doing so to be important, is integral, on this account, to their understanding of what it is to live well.

Assessing a principle's reasonable rejectability

As noted in the previous section, assessing the validity of a particular principle – assessing, that is, whether or not it is a principle that no one can reasonably reject – requires considering the general implications, from different points of view, of individuals being guided in their practical thinking, in the relevant type of situation, by the proposed principle. This is not, however, a matter of imagining what the implications of the proposed principle might be for particular individuals one happens to know. Rather, it is a matter of thinking about the proposed principle's implications from different individual standpoints. A "standpoint" is just an abstract description of a point of view characterized by a certain combination of things that people generally have reason to want and to care about (Scanlon 1998: 202–6). Any particular standpoint could (but need not) aptly characterize some actual person's point of view, and any actual person's point of view will certainly be aptly characterizable by many different standpoints.

So, let's say that what is in question is the validity of a principle that permits deception under certain circumstances. Assessment of its reasonable rejectability

will focus, not on the implications for actual people of it being permissible in some circumstances for individuals to deceive one another, but on its implications as assessed from different relevant standpoints, such as: someone who finds himself in circumstances in which deception is permitted, someone not permitted to deceive in circumstances in which he would have good reason to do so were it permissible, someone who stands to be deceived, someone whose interests stand to be affected in certain ways by others being deceived, etc.

One of the defining features of contractualism that starkly distinguishes it from any form of consequentialism is what has come to be known as the *individualist restriction*. It stipulates that the only implications of a principle that are relevant to its assessment are those that could be appealed to from a certain standpoint as implications of the proposed principle for a person being able to live his or her life. Ruled out as irrelevant, then, are the more impersonal kinds of consideration that consequentialists take to be important, such as the implications of a proposed principle for the aggregate value of the outcome that will obtain if it is generally complied with, or the aggregate value of its implications for several individuals. Neither of these is the type of consideration that could be appealed to from the standpoint of an individual objecting to a proposed principle *on his own behalf*.

On the same grounds, a principle's implications for those things a person might, for good reason, care about, but whose importance to him is not properly characterized as being important for his being able to live his life, have no bearing on a principle's assessment. A person with a certain standpoint might, for instance, take it to be important that certain species of wild fauna flourish, quite independently of the pleasure he takes in experiencing their flourishing, but that a principle will have bad implications for those species won't count as a relevant implication from his standpoint of the principle.

Critics of contractualism have forcefully argued that the theory would be more plausible without the individualist restriction. That may be, but the restriction does make sense as a central feature of the account. What it holds matters to the morally motivated person, after all, is that he respond to the value of the lives of others in a way consistent with standing in a relationship of mutual recognition to any other. Whether or not treating another in a certain way is an appropriate response to the value of that person's life is, intuitively, a matter that turns on the ways that he might object that one would be failing to take seriously the implications *for his life* by treating him this way.

Though the individualist restriction limits the range of considerations relevant to a principle's assessment, it is less restrictive than it might at first seem. First, as any principle is for the regulation of a certain type of situation, its assessment will have to take into account not just the implications of certain individuals being guided by it on certain occasions, but those of all individuals over time being guided by it in that type of situation. Second, implications of a principle having nothing to do with how well or badly off individuals will be are also relevant to its assessment. Consider, for example, a principle that licenses a

designated authority to periodically force randomly chosen individuals to serve as test subjects for dangerous medical experiments. It is of course relevant to assessing such a principle that it exposes everyone to the risk of serious harm. But a further objection to it that could be pressed from a certain standpoint is that such experiments being permissible undermine the exclusive authority of individuals concerning decisions about how their bodies are to be used (a prerogative that plays a fundamental role in an agent's understanding of his life as his own), turning them into a form of public property. This is a good reason for wanting to reasonably reject a principle that permits the envisioned form of experimentation, even though it has nothing to do with its implications for individual welfare.

That a certain principle has, from some standpoint, seriously negative implications does not, however, warrant the conclusion that, all things considered, it is one that could be vetoed from that standpoint as reasonably rejectable. Whether it can depends on the implications for other standpoints of the alternative candidate principles. If every alternative to the one that, from a particular standpoint, is seriously objectionable has implications for some other standpoint that are even more seriously objectionable, the principle cannot be reasonably rejected.

It cannot be reasonably rejected because standing in a relationship to other individuals of mutual recognition requires that one's rejection of a principle in favor of some alternative to it be justifiable to other individuals – especially to the one who has the most serious objections to this alternative principle. If the alternative to the principle one wants to reject is worse for someone from a different standpoint than the principle being rejected is for oneself, the proposed principle's rejection won't be justifiable to that person. What reason could one offer, after all, for why he should bear a greater burden in order that you are spared a lesser burden that would be compatible with the recognition of his equal value to oneself as a person?

Whether or not a particular principle is in fact reasonably rejectable is a matter of judgment about which morally informed and sensitive individuals may well disagree. Contractualism's aim is not that of providing a method for resolving moral disagreements; what it offers is a characterization of what the question is that conclusions about how it is morally permissible for persons to treat one another answer. But what the correct answer is to that question, in any given instance, may well be contestable.

The problem of aggregation

Part of contractualism's appeal lies in its value as a framework for developing plausible justifications for familiar convictions concerning how it is morally permissible for persons to treat one another that have proven difficult to articulate in consequentialist terms. Consider, for instance, the often-discussed

question of the permissibility of killing a healthy individual in order to transplant his organs to five others, whose survival depends on receiving a transplant. The problem for consequentialists isn't in saying something about why doing so is impermissible. It is, rather, that of saying something about why this is so that makes sense of the intuitive idea that its being impermissible has to do with there being something in itself objectionable about one person taking another's life as a means to even a worthwhile end.

Contractualism's strength lies in its ability to make good sense of this thought. On its terms, the question of the permissibility of killing a person for his organs turns on the strength of the reasons that could be pressed from the standpoint of that person against a principle permitting such killings, as compared with the reasons that could be pressed from the standpoint of those standing to benefit from such killings being permitted. Since the number of individuals who stand to benefit if such killings are permitted is not a relevant consideration in contractualist moral argument, it is fairly easy to see how an argument would go for the impermissibility of killing for organs, one that appeals, in particular, to the reasons individuals have for wanting the kind of control over their lives that killing for organs being permitted would undermine.

There are situations, however, in which the number of individuals who stand to be benefited or burdened is clearly morally relevant. Say, for example, you find yourself in a boat equidistant from two tiny islands. There is only one person on one of the islands, but there are six on the other. The tide is rising, so there is only time to reach one of the islands and save its inhabitants before all are drowned. Few doubt that the only morally permissible course of action is to save the six. And since the obligation in question is just an instance of a more general obligation to aid others when one can do so at little cost to oneself, the rationale for doing so ought to be explicable using the resources of the contractualist framework.

Doing so turns out, however, to be surprisingly difficult. Contractualism doesn't count aggregative considerations as morally relevant, so the amount of good that will be done by saving the six can't be appealed to in order to justify a principle that requires the rescuer to do so. Considering what might be justifiable to the standpoint of each of those whose life is at risk, however, leads to the conclusion that the only non-rejectable principle for the regulation of this type of situation is one that requires the rescuer to give each an equal chance of being saved by, for example, tossing a coin to settle the matter. The strength of the reason that each has for having the rescuer save him is exactly the same, after all. How, then, can it be justifiable to the one alone on an island that he not be given the same chance of being saved as each of the many?

Arguably, contractualism is not committed to the counter-intuitive conclusion that one should toss a coin to decide whether to save the many or the one (Scanlon 1998: 229–41). It's true that since there is no reason, *ex hypothesi*, to save any one rather than another, there are no grounds for thinking that a principle no

one could reasonably reject for the regulation of this type of situation requires any one, rather than one of the others, be saved. What can be reasonably rejected, though, is any principle that does not require the rescuer to take into account, in deciding whom to save, the reason of each for wanting to be saved. Tossing a coin, or any simple lottery procedure, violates this requirement, for if the rescuer is guided by such a procedure, it will make no difference to his thinking about how to proceed whether the situation is one of the one on the first island and of the six on the other, or two on the first island and twelve on the other.

A plausible principle for the regulation of the rescuer's reasoning in this type of situation that respects this requirement is one that directs the rescuer, in his thinking about how to proceed, to balance the reason to save one person's life against the reason to save the life of another that requires a course of action incompatible with saving the first person. By balancing competing reasons against one another, the rescuer takes into account each person's reason for wanting to be saved. What decides how the rescuer will proceed is the first reason to save someone that is not balanced by an equally weighty opposing reason to save the life of another.

This procedure directs the rescuer to save the six, but not because doing so will do the most good. And though the one who is not saved will no doubt curse his bad luck, the rescuer has not wronged him. The reason to save his life is given exactly the same weight in the rescuer's reasoning as the reasons to save each of the others. He is just unlucky that the reason to save his life happened not to require the same course of action as an equally forceful reason to save the life of another.

Whether this argument is successful as an example of how one can explain, in contractualist, non-aggregative terms, intuitive responses to cases which appear to be hard to make sense of without appealing to aggregative considerations is a contested issue (see Kamm 2007). Doubts have been raised about both the soundness of the above line of argument and, even if it is sound, whether it provides any guidance as to how a contractualist might justify the intuitive relevance of aggregative considerations in cases that do not have the structure of a conflict between equally forceful reasons. Whether contractualism can be defended as a compelling account of what we morally owe one another turns, to a large extent, on the development of convincing contractualist responses to these kinds of questions.

See also Consequentialism (Chapter 37); Contemporary Kantian ethics (Chapter 38).

References

- Kamm, F. M. (2007) "Owing, Justifying, and Rejecting," in *Intricate Ethics*, Oxford: Oxford University Press.

- Kant, I. (1997) *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor, Cambridge: Cambridge University Press. (The reference is to volume and page of the standard Prussian Academy edn, as given in the margins of this work.)
- Scanlon, T. M. (1982) "Contractualism and Utilitarianism," in A. Sen and B. A. O. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press.
- (1998) *What We Owe To Each Other*, Cambridge, MA: Harvard University Press.
- (2003) "Preference and Urgency," in *The Difficulty of Tolerance: Essays in Political Philosophy*, Cambridge: Cambridge University Press.

Further reading

- Kamm, F. M. (2007) "Owing, Justifying, and Rejecting," in *Intricate Ethics*, Oxford: Oxford University Press. (An important and sympathetic discussion of contractualism's credentials as a non-consequentialist account of moral reasoning.)
- Scanlon, T. M. (1982) "Contractualism and Utilitarianism," in A. Sen and B. A. O. Williams (eds) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press. (The original paper sketching the contractualist account. A must-read before tackling the book-length treatment of the view.)
- Wallace, R. Jay (2006) "Scanlon's Contractualism," in *Normativity and Will*, Oxford: Oxford University Press. (A comprehensive review article that critically discusses most of the major themes of *What We Owe to Each Other*.)