



**October 9th, from 4-7 pm
Lester Pollock Room, FH, 9th Floor**

Colloquium in Legal, Political, and Social Philosophy

Conducted by

Liam Murphy and Samuel Scheffler

Speaker: T. M. Scanlon, Harvard University

Paper: Contractualism and Intuitionism



Colloquium Website: <http://www.law.nyu.edu/node/22315>

Contractualism and Intuitionism

T. M. Scanlon¹

The appealing idea that the justifiability of our actions to others is central to morality and provides the basis for an account of its content has led to a variety of views. I have made two particular contributions to the discussion of this idea, both of which may well have been mistakes. The first was to develop a specific moral theory based on this idea. The second was to give this theory the name, “contractualism.” This name is unfortunate insofar as it suggests that the basis of morality lies in some actual or hypothetical agreement between the affected parties, rather than, as I see it, in normative facts about the reasons that individuals have to accept or reject principles. So I somewhat regret giving my view this name, although I still have not thought of a name that would be better.

The particular formulation that I gave to the contractualist idea also seems to me now to have some serious problems, but I don’t regret offering it. Justifiability to others is an important, appealing, and, I still believe, fruitful idea, albeit rather vague. The way to make progress in thinking about it is to offer more specific formulations, see where they go wrong, and look for better alternatives. In this paper I will try to say what is distinctive, and to my mind appealing, about my particular version and discuss its implications, including, as far as I can manage it, an honest assessment of some of its shortcomings and some attempts to repair several of them.

¹ This paper began as a lecture presented to a conference on contractualism in Zurich in 2022. I am grateful to the participants in that conference for their comments and to Robert Audi, Selim Berker, Rahul Kumar, Hon-Lam Li, Peter de Marneffe, Richard Moran, Thomas Nagel, and Erik Zhang for comments on subsequent versions that have helped me to improve the paper.

The Questions of Content and Acceptance

Contractualism aims to answer two questions. It seeks to provide a characterization of the content of morality—of the subject we are thinking about in reaching conclusions about moral right and wrong—and to do this in a way that makes clear the reasons for taking morality seriously as a set of normally overriding standards of conduct.² I call these questions the question of content and the question of acceptance. Each of these questions needs some clarification.

I have stated the question of acceptance as the question of whether people have sufficient reason to accept moral principles as normally overriding *standards* of conduct rather than as *guides* to conduct. This is because the question is not only whether individuals have sufficient reason to comply with these standards themselves but also whether they have sufficient reason to accept them as specifying, and limiting, what they can demand of others. The question “Why be moral?” is misleading in focusing only on the former. Similarly, the question of the “demandingness” of a moral theory is commonly understood as a matter of the sacrifice it may require of agents, who must give up something important in order to comply with its demands. Putting matters this way focusses our attention on those who are in more advantageous circumstances and thus in a position to “sacrifice” in this way. But we need to consider also the position of those who would benefit from such sacrifices, and whether they would have sufficient reason to accept the fate that would result if the sacrifices of others were not required.

² As in “Contractualism and Utilitarianism,” in *The Difficulty of Tolerance*, pp. 127-128., substituting “reasons for acceptance” for “motivational basis.”

One kind of account of the subject matter of morality would just be a complete and consistent of which actions are right and wrong, stated in terms of ordinary first-order concepts such as rights, duties, and obligations. This would be a fine thing to have, but even if it brought these judgments into reflective equilibrium with each other, such an account would leave open the question of why we should regard them as normally settling questions about what to do and what we can demand of others. It could just be a consistent rendering of a socially inculcated set of requirements and prohibitions. What is needed is not just a systematic account of which things are morally right and wrong, but an account of what it *is* to be morally right or wrong, in the light of which we can see why being right or wrong should have the importance for us that it is commonly thought to have: a characterization of the content of morality that links up with a satisfactory answer to the question of acceptance.

The need for such an account lies behind Elizabeth Anscombe's claim that what she calls a "law conception of ethics" becomes empty once the idea of God as lawgiver is rejected. It is, she says, a case of "the survival of a concept outside the framework of thought that made it a really intelligible one."³ What she is claiming is that God's authority provides a convincing answer to what I am calling the question of acceptance, and that without this there is no basis for taking moral requirements to be authoritative.

Similarly, Jeremy Bentham wrote

Of an action that is conformable to the principle of utility one may always say either that it is one that ought to be done, or at least that it is not one that ought not to be done.

³ "Modern Moral Philosophy," in *Ethics, Religion and Philosophy: Collected Philosophical Papers, Volume III*, pp. 30-31.

One may say also that it is right that it should be done: that it is a right action; at least that it is not a wrong action. When thus interpreted, the words ought, and right and wrong, and others of that stamp, have a meaning: when otherwise, they have none.⁴

I take it that Bentham here agrees with Anscombe that an answer to the question of acceptance is needed. But he believes that utilitarianism, as account of the content of morality, provides a convincing answer. When he says that words such as ‘ought,’ ‘right,’ and ‘wrong’ have meaning only when understood in the way that utilitarianism offers, what he means is that utilitarianism is the only account of the content of morality that provides a convincing answer to this question. I agree with Anscombe and Bentham that an answer to the question of acceptance is needed, and I agree with Bentham that it can be provided by the right kind of account of the content of morality –an account not just of which things are right and wrong but of what it is to be right or wrong. But contractualism provides an answer to this question that differs from Bentham’s.

According to contractualism, the morality of right and wrong is a set of standards that can be justified taking the reasons of all into account, and it claims that we should accept these standards as authoritative because standards that can be justified in this way are a way of living together with others that we have good reason to accept.⁵ This answer to the question of acceptance may seem insufficiently strong or, as Nicholas Southwood puts it, insufficiently

⁴ “Introduction to the Principles of Morals and Legislation,” in Ryan, ed., *Bentham and Mill: Utilitarianism and Other Essays* (London: Penguin Books, 1987), p. 67.

⁵ I have described Anscombe’s view, Bentham’s, and my own as answers to the same question of why we should see *moral* standards as authoritative. It is important to note, however, that the “morality” they are talking about is not exactly the same thing. These different answers to the question of acceptance have different implications for the range of things that morality covers as well as for what it requires. The range of things subject to divine commands, for example, need not be the same as the aspects of our lives that we have reason to want to be able to justify to others.

fundamental.⁶ My problem is that I don't see how one can go deeper. I find Kant's attempt to ground morality in conditions of rational agency unconvincing. The appeal of his view seems to me to rest on the substantive appeal of his ideas of treating others as ends in themselves and as members of a kingdom of ends as desirable ways of living together with others—that is to say, on claims very much like those I am making. Ideas of interpersonal deliberation, as in Habermas and Southwood, seem to me not to go deeper than this.⁷ Their accounts of the content of morality and their answers to the question of acceptance seem to me to rest ultimately on claims about reasons that are similar to mine. Claims about reasons thus seem to me the bedrock (or perhaps I should say the shifting sand) on which we must stand.

It can of course be maintained that the question of acceptance is not one that we need to answer. This may seem to be what H. A. Prichard was arguing in “Does Moral Philosophy Rest on a Mistake?” Asking the question of acceptance is not the mistake that he was referring to, however. That mistake was the idea that some form of argument is required to get from a factual description of a situation to a moral conclusion about what an agent in that situation ought to do. He believes that this step is “immediate.”⁸ But Prichard does not even consider what I am calling the question of acceptance. He believes that once such a moral conclusion is reached there is no further question about what the agent ought to do. This is because Prichard takes a single sense of ‘ought’ as the basic normative notion, and does not distinguish between a moral ‘ought’ and a more general ‘ought’ of rationality.⁹ Since I take the idea of a reason as

⁶ Southwood, *Contractualism and the Foundations of Morality*, Section 3.2.

⁷ In Habermas, *The Theory of Communicative Action*, and Southwood, *Contractualism and the Foundations of Morality*.

⁸ “Does Moral Philosophy Rest on a Mistake?” *Mind* 21 (1912) pp. 21-37.

⁹ On this see Thomas Hurka, *British Ethical Theorists from Sidgwick to Ewing*, pp. 27, 132-134.

the basic normative notion, there is in my view room to ask, after the conclusion has been reached that something is what the agent morally ought to do, what reason the agent has to do it.¹⁰ I consider it a strength of a view that takes reasons as basic that it provides room for this question. But some may disagree, maintaining that such a view invites us to ask a question that is in fact illegitimate and impossible to answer, thus leading to skepticism.

Even though an account of the subject matter of morality of the kind we are looking for is not simply a set of first-order claims about which actions are right or wrong, a convincing account of the subject matter of morality needs to offer a plausible account of the kind of thinking that we engage in when we are deliberating about right and wrong. This does not mean that it must fit with all of our current first-order judgments. It would, after all, be surprising if all of our moral judgments turned out to be correct. Even firmly held judgments can turn out, on further reflection, to hold only under certain conditions, or only if certain factors are neglected.¹¹ But an account of the subject matter of morality should provide a plausible and informative account of the thinking we engage in when we are uncertain about what morality requires, and it should explain why many moral requirements that we are more certain about have the particular structure that they do.¹²

The idea of the “subject matter” of morality should not be taken to imply that this is a domain of facts in which every question of right and wrong has a determinate answer. According to contractualism, facts about what we owe to each other are facts about which

¹⁰ Here I agree with Wallace, *The Moral Nexus*, p. 68.

¹¹ Compare what Rawls says about “precepts of justice” in section 47 of *A Theory of Justice*.

¹² Compare Sidgwick on “The Relation of Utilitarianism to the Morality of Common Sense,” in *The Methods of Ethics*, Book IV, Chapter III, esp. pp. 453-454.

principles it would be reasonable to reject and which principles it would not. These are facts about the sufficiency of various reasons for rejecting a principle when various opposing reasons are taken into account. But the general domain of facts about reasons is not, as far as we can tell, fully determinate.¹³ If not every claim about reasons has a definite truth value, the same may be true of moral judgments if contractualism is correct.

Contractualism as a Reductive and Interactive Theory

The appeal of contractualism as an answer to the questions of content and acceptance depends on the reductive and interactive character of the form of justification that it involves. This form of justification is reductive in that it makes the principles that determine the content of morality depend on the reasons individuals have to want their lives to be affected, or not affected, in certain ways. It is interactive insofar as it depends on whether the reasons for rejecting a principle that a person has in virtue of being in one position, are *sufficient* reasons for rejecting a principle given the reasons those in other positions have for wanting the opportunities or protections that such a principle would provide.¹⁴

This interactive character of the contractualist idea of justifiability is crucial to its appeal as an answer to both of the questions of content and acceptance. But it brings with it a tendency toward moral requirements that are maximally demanding on agents. As Elizabeth

¹³ Or so I allow in Chapter 4 of *Being Realistic about Reasons*. See esp. pp. 84-86.

¹⁴ The interactive character of the justification for fundamental moral principles is explicit in my own version of contractualism but emphasized by others as well. See, for example, Samuel Scheffler on “reciprocal normativity” as a characteristic of moral requirements. (*Equality and Traditions*, pp. 56ff, citing Wallace.) See also Forst on the requirement of justification by reasons that are generalizable and reciprocal. (*The Right to Justification*, Chapters 1 and 2, esp. p. 42.) In Rawls’ original position the parties are seeking to maximize their own expectations, but the Veil of Ignorance forces them to consider for the reasons of those in other positions by considering the possibility that they themselves might be in any of those positions. [In a late section of “Contractualism and Utilitarianism” I questioned whether this is an adequate way of representing the idea of interactive justification.]

Ashford concludes, “the consideration of the comparative strength of the objections to principles governing obligations to help those in need, by those who will benefit and those who will be burdened by such principles, leads to an exceedingly demanding principle of aid that cannot reasonably be rejected.” Ashford does not regard this as an objection to contractualism but rather just “an appropriate response to morally salient features of the current state of the world.”¹⁵ She does, however, take it to be an objection to contractualism that it would be overly demanding even under more ideal conditions.

Other approaches within the broadly “contractual” category of moral theories, such as accounts offered by David Gauthier and Gerald Gaus,¹⁶ lead to a more minimal view of moral requirements because they adopt an interpretation of justifiability to each person that is not interactive. I will focus here on Gaus’s version.¹⁷ He says that justifiable principles regulating a given area of life must be ones that every citizen has sufficient reason to prefer to having no principles at all regulating that area. The reasons for preferring standards that he takes to be relevant are the ones that follow from that person’s evaluative standards, whatever these may be as long as these reasons are “intelligible” to others.¹⁸ These include, in Gaus’s account, reasons implied by whatever (intelligible) moral views that person holds.

So the form of justifiability on which the content of morality depends according to Gaus is not interactive.¹⁹ Other people’s reasons figure in the justifiability of moral principles to a

¹⁵ “The Demandingness of Scanlon’s Contractualism,” *Ethics* 113 (2003) p. 292.

¹⁶ Gauthier in *Morals by Agreement* and Gaus in *The Order of Public Reason* and other writings.

¹⁷ For a detailed critical assessment of Gaus’s view, see David Enoch, “The Disorder of Public Reason,” *Ethics* 124 (2013), pp. 141-176. Enoch takes some of his objections to Gaus to apply to my own view as well, insofar as it rests on an (idealized) idea of justifiability others.

¹⁸ *The Order of Public Reason*, pp. 279-283.

¹⁹ This would be true of Gauthier’s view as well.

person only if that person happens to care about them or is affected by what they are likely to do. It follows that if there is even one person who does not care about certain other people, or who thinks that the interests of some people don't matter because they do not deserve to be helped, then moral requirements requiring them to be aided are ruled out as not justifiable to this person in the required sense.²⁰

This rejection of interactive justification does not amount to a rejection of justifiability to others as a basic moral idea. Gaus accepts this idea, but he believes that justifiability to others in the sense he describes is what is required in order to be treating others with respect. I disagree. I am inclined to ask, "What about respect for those whose need for aid is dismissed simply because some people think them unworthy?" An account of the content of morality based on interactive justification seems to me a much more compelling answer to the question of acceptance. This disagreement brings out the fact that a view about what counts as a satisfactory answer to the question of acceptance is in a broad sense a moral view—not a view about which things are right and wrong, but a view about the way in which the way we have reason to want to live with others involves taking their reasons into account.²¹

The problem for contractualism, as I have said, is that although its interactive account of the content of morality offers an appealing answer to the question of acceptance, this account appears to entail conclusions about the content of morality that are more demanding than many of our considered judgments about what we owe to each other. This poses a problem

²⁰ Gaus, *The Order of Public Reason*, p. 363.

²¹ Thus Rawls writes that mutual respect requires that "When called for, reasons are to be addressed to those concerned; they are to be offered in good faith, in the belief that they are sound reasons as defined by a mutually acceptable conception of justice which takes the good of everyone into account." *A Theory of Justice*, 1st edn., p. 338. The question is how their reasons must be taken into account.

about how to reach a reflective equilibrium that takes into account not only our considered judgments about right and wrong, but also our considered judgment that these are standards that everyone has reason to be guided by.

This description of what we are doing may seem too respectful of common-sense morality. Instead of trying to find an account of the content of morality as commonly understood and then looking for answer to the question of why we should accept it, why not just inquire directly, from the outset, what standards we have most reason to accept?²²

This reasonable-sounding question reflects a misunderstanding. What an account of the content of is seeks to offer is not a portrait of what is commonly thought about right and wrong but rather an account of what we in fact owe to each other. We begin with our considered judgments about right and wrong because these are things that seem to us, at the outset, to reflect standards that we have reason to accept. The question is whether this is so, and if it is so what these reasons are. So we are, from the beginning, investigating the question of what standards we have reason to accept.

The phrase “considered judgments” can itself be misleading in a similar way. The search for reflective equilibrium starts with those judgments about the reasons we have that seem to us, on reflection, to be true. But the process of seeking reflective equilibrium is a process of rethinking these judgments in the light of the conflicts between them. There are two points here: the first is that what we are concerned with are conflicts with what has seemed to us to be true, not with what is commonly thought. The second is that even judgments that have seemed to us to be clearly true are open to revision.

²² I am grateful to Tyler Burge for pressing me to clarify the status of “common sense” moral beliefs in my project.

This revision can take a number of different forms. If a proposed account of the content of morality provides a sufficiently plausible and enlightening way of thinking about questions of right and wrong (as well as a satisfying answer to the question of acceptance), this can give us reason to revise what were previously our considered judgments about right and wrong. But the conflict with considered judgments about right and wrong can also lead us to revise a proposed account of the content of morality (for example, in the case of contractualism, possibly changing our mind about the class of reasons for rejecting a principle, or about the sufficiency of these reasons.) Third, conflict with particular considered judgments could lead us to reject an account of the content of morality altogether in favor of some alternative account, which would raise the question of what alternative answer to the question of acceptance it supports. Finally, as I said above, one might reject the need to answer the question of acceptance at all, holding that facts about what we morally ought to do are normatively self-supporting. The fact that an action would be wrong is in itself a sufficient reason not to do it which requires no further explanation. More darkly, one might conclude that the aspirations of morality to be justifiable to all are unfulfillable.

My present point is just that on the approach I am taking none of these alternatives is ruled out from the start. I will return to these questions in discussing the challenge raised by what Derek Parfit calls the Greater Burden Claim. In order to do that, I need first to examine the contractualist account of the content of morality in more detail.

Why Principles?

Contractalism holds that the rightness or wrongness of actions depends on the justifiability of principles that would permit them. So I need to say more about what these

principles are and why they have this role. Principles, as I am understanding them are what Kant calls maxims.²³ They hold that certain reasons are sufficient or even conclusive reason for acting in certain ways under certain conditions. According to contractualism the rightness or wrongness of actions depends on the principles that would permit such actions because individuals have reason to be concerned with effects on their lives other than the effects of particular actions and the cost of performing them. This is true in a number of different ways.

The first has to do with what might be called intrapersonal aggregation on the side of agents and on the side of those who are affected by their actions. As Parfit observes in discussing actions that add to the pollution of air, food, or water, “when we consider any one such act, the tiny effect on the many may seem trivial. It may seem not to matter if such an act imposes costs on others of less than ten cents or reduces the life expectancy of others by less than one minute. But when many people act in such ways, these small effects add up.”²⁴ What adds up here is not (or not only) the effects on different individuals but the cumulative effect on single individuals of the actions of many agents. This is what I will call intrapersonal aggregation on the victim side. A similar phenomenon also occurs on the side of agents. Making a sacrifice for the benefit of someone else on one occasion may not matter much for the the

²³ See *Groundwork of the Metaphysics of Morals*, 421, 427. The difference between the two is that what Kant is concerned with are the maxims of particular actions. Any rational action involves *taking* some consideration to be a sufficient reason for acting in a certain way in certain circumstances. What I am calling a principle claims in general that certain considerations are sufficient reason (for anyone) to act in a certain way in certain circumstances. It is thus like what Kant calls the universalization of a maxim.

²⁴ *On What Matters*, Volume. Two, pp. 205-206.

agent. But a principle that required agents to do this repeatedly, in situations that frequently arise, would have a serious impact on an agents' life.²⁵

In addition to these aggregative effects, it is important to consider principles, rather than only the effects of particular actions, because individuals have strong reason to want to have assurance that they will not be affected in certain ways, by the actions of others or by demands for assistance. This is an important aspect of privacy, for example. We not only have reason to want interventions into our house or our private life not to occur, but also reason to want justified confidence that such things will not occur.

The way of having confidence that is relevant for present purposes is to have a justified belief that others generally accept principles ruling out such interventions into our lives, or at least principles that limit the reasons that can justify such interventions. This also gives them the desirable status of having certain aspects of their lives regarded by others as generally off limits to them *because* of their reasons for wanting control over these matters.²⁶

These forms of assurance provide what Mill called "security," and regarded as a particularly important form of utility—or, as I would put it, as something that everyone has particularly strong reason to want.²⁷ I believe that security of this kind is also what Mill had in mind when he criticized Bentham for holding what Mill called "the doctrine of specific consequences" according to which the rightness of an action depends only on its consequences,

²⁵ See Murphy, "It is therefore appropriate, when offering reasons for and against a principle, to consider the possible cumulative 'intrapersonal' burdens it would entail. "Nonlegislative Justification," in *Principles and Persons: The Legacy of Derek Parfit*, p.

²⁶ See Kamm, "Non-Consequentialism, the Person-as-an-End -in-Itself, and the Significance of Status," *Philosophy & Public Affairs* 12 (1992), 354-389. But I see this idea of status as relevant to the question of acceptance rather than the question of content. That is to say, as an attractive feature of the relation with others established by morality as contractualism describes it rather than a reason within contractualism for rejecting a principle.

²⁷ Mill, *Utilitarianism*, Ch V.

or the consequences of similar actions if performed generally.²⁸ Mill's objection was that Bentham failed to see that rightness depended on considerations of security as well as intrapersonal aggregative effects. In my version of contractualism, reasons for favoring or objecting to a principle include general effects of both of these kinds.²⁹

Principles are often just schemas.³⁰ They identify a kind of case, specifying the different parties affected and the ways in which they are affected, and state that in a case of this kind certain considerations are normally sufficient reasons or even conclusive reasons for acting in certain ways unless there are sufficient countervailing reasons of a specified kind. Such a principle is merely a schema insofar as it does not specify when reasons of the latter kind are sufficient. Judgment about which is then required in order to reach definite conclusions about what may or must be done in a given case. We can see how this is so by considering the principles regarding fidelity to promises. This will illustrate a distinctive feature of contractualist reasoning about rightness and wrongness.

The basic principle of fidelity specifies that if one has given someone assurance that one will act in a certain way (a way that is, as far as one can tell, permissible) unless they consent to one's not doing so, and if one has good reason to believe that they want this assurance and may rely upon it, then this is normally a conclusive reason to act in that way. Contractualism explains this by claiming that individuals have strong reason to want to be able to offer and to

²⁸ Mill, "Remarks on Bentham's Philosophy," *Collected Works of John Stuart Mill Volume 10: Essays on Ethics, Religion and Society* edited by J. M. Robson (Toronto: University of Toronto Press, 1969), pp. 5-18.

²⁹ As I observed in *What We Owe to Each Other*, p. 205 and in note 15, pp. 394-395. The importance of these considerations is my main reason for advocating a principle-based contractualism rather than an act-based version, as advocated by, among others, Hanneke Scheinman in "Act and Principle Contractualism," *Utilitas* 23 (2011) pp. 288-315.

³⁰ I take the term from Wallace, *The Moral Nexus*, Chapter 5.

rely on such assurances, and therefore strong reason to reject a principle permitting violations of the principle of fidelity just stated. These are sufficient reasons for rejecting such a principle because individuals who would be burdened by the principle of fidelity could avoid being so burdened by refraining from giving assurances in the first place.³¹

As stated, this principle covers only what is required “normally.” It leaves open the possibility that, for example, keeping an obligation may be incompatible with giving aid to someone who very much needs one’s help. In some cases the cost to such a third parties may be so great as to justify not acting in the way agreed to. A principle that did not recognize possible exceptions of this kind could reasonably be rejected by someone in the position of such a third party. But it would also be reasonable to reject a principle holding that assurances are never binding whenever fulfilling them would benefit the promisee less than some third party could be benefitted by violating the assurance. This would render assurances meaningless. So the reasons that individuals have for wanting to be able to offer assurance, and to be able to rely on assurances given (a form of Millian security), would be sufficient reason to reject such a principle.

In between these extremes is a schematic principle of fidelity that recognizes that assurances given can fail to be binding when failing to fulfil them is the only way to aid someone who is in serious need of help. It does not seem possible to state this exception more exactly, in the abstract, by specifying how serious this need must be. But a judgment about this

³¹ This explains the requirement of voluntariness in terms of the value of having had the choice, as a protection against unwanted commitments, rather than in terms of an independent idea of the moral significance of consent. See my *The Significance of Choice*, and Chapter 6 of *What We Owe to Each Other*, and more recently in “Responsibility and the Value of Choice.”

in any particular case cannot be based simply on a comparison of the benefits of giving aid and of fulfilling the assurance. As I have said, in order for assurances not to be meaningless such a judgment must also take into account the degree to which recognizing an exception in cases of the kind in question would impair the general ability to offer assurance and to be able to rely upon it.³² It may be objected that I am just restating what we normally think about the obligation to keep a promise. My response is that I am offering an *explanation*, in terms of reasons, of why we should think this, and thus an explanation of what we are doing when we are assessing the limits of the obligation to keep a promise.

What is distinctive about this contractualist account can be seen in the way in which it resembles but also differs from W.D. Ross's view, according to which the rightness or wrongness of failing to fulfill a promise in order to help a third party is determined by balancing the *prima facie* duty of fidelity against the *prima facie* duty of benevolence.³³ The similarity lies in the fact that what I have called a schematic principle states what Ross called a *prima facie* duty. He wrote, "I suggest '*prima facie* duty' or 'conditional duty' as a brief way of referring to the characteristic (quite apart from that of being a duty proper) which an action has, in virtue of being of a certain kind (e.g. the keeping of a promise), of being an act which would be a duty proper if it were not at the same time of another kind which is morally significant. Whether an act is a duty proper or actual duty depends on *all* the morally significant kinds it is an instance

³² So, whether it would be wrong to violate a "death bed promise" for the sake of some important good depends on whether the reason for wanting to be able to make promises that are binding even in such cases is sufficient reason to reject a principle permitting them to be violated, taking into account the costs of this rejection. The answer may not be clear (that is why the question seems difficult) but this seems to me the right way to think about such cases. For a contrary view, see Scheinman, "Act and Principle Contractualism."

³³ *The Right and the Good*, p. 19

of.”³⁴ What I have called the schematic character of such a principle—its leaving open the possible justifiability of exceptions of a certain kind—captures what Ross called the “conditional” nature of a *prima facie* duty.

Where my contractualist view differs from Ross is that it offers a different, and I think clearer, account of how one gets from *prima facie* duties to conclusions about duty *simpliciter*. It may be true that the need to help a person would not justify failing to fulfill a promise unless, in the absence of the promise, it would be wrong not to help the person. But it does not seem to me that such cases are best understood as ones in there is a conflict between the duties of fidelity to promises and of benevolence, calling for a judgment as to which of these duties is “more incumbent” and “comes before” the other.³⁵ Rather, these are cases in which the “schematic” or “conditional” character of the relevant principles consists in allowing the possibility of certain exceptions, which must be recognized in order for the principle not to be rejectable. What is called for is not a judgment about the stringency of conflicting duties but a judgment about the sufficiency of the reasons for and against particular versions of these exceptions. Even though the answer may not always be clear, this seems to me a more informative way of thinking about such cases than the idea of conflicting duties.³⁶ This supports the case for contractualism as an answer to the question of content.

³⁴ *Op. cit.*, pp. 19-20.

³⁵ p. 19

³⁶ Indeed, it seems to me to fit well with the substance of what Ross himself says about promising. See *The Right and the Good*, pp. 34-35. Ross says (pp. 29-33) that, unlike conclusions about duties *simpliciter*, *prima facie* duties are self-evident. What I have called schematic principles may not be self-evident, since even they rest on judgments about the sufficiency of reasons. But they are more self-evident than conclusions about rightness and wrongness in virtue of the fact that they leave open less clear questions of sufficiency, to be settled later.

In its emphasis on principles that are justified in a reductive way, contractualism resembles my earlier view of rights³⁷ as well as rule utilitarianism and other forms of indirect consequentialism. But this resemblance is also misleading. One obvious difference is that my contractualist view assesses principles in terms of individuals' reasons rather than in amounts of utility or goodness. But a further important difference is that, as I have just pointed out, the reasons involved include reasons—such as reasons of security—that are not based on the consequences of individual actions. Responsiveness to these reasons for objecting to actions directly requires concern with the general effects of principles that would permit them. Principles are therefore not merely an indirect means of promoting or preventing effects of particular actions that individuals have reasons to want or to object to.

This bears on the familiar question of the “collapse” of indirect forms of utilitarianism into “direct” act-based versions, and explains why this problem does not arise for contractualism. The general form of that problem arises as follows. If the justification for requiring conformity to a principle is just that this is an effective way of promoting some value that can also be promoted by individual actions, and if in some case an action that violated that principle would better promote that value than adherence to the principle, then there would be sufficient reason to violate the principle and no good reason to object to doing so.

But things are different with regard to the permissibility of violating a principle that is justified by the fact that it provides some form of security, for example by requiring fidelity to assurances given or by giving individuals control over certain aspects of their lives. In

³⁷ See “Rights, Goals, and Fairness” in my *The Difficulty of Tolerance: Essays in Political Philosophy* (Cambridge: Cambridge University Press, 2003), pp. 26-41. I discuss the parallel between these views in “Contractualism and Justification.”

contractualist terms, the question is whether and when it would, or would not, be reasonable to reject alternative principles that allow exceptions to these principles in certain cases. As we have seen in the case of fidelity to assurances, an answer to this question must take in to account the degree to which security of the kind in question would be undermined by allowing an exception of a certain kind, and must consider whether this loss in security would be sufficient reason to reject a principle recognizing that exception, taking into account the reasons for wanting it.

There are two crucial points here. The first is that this question is a question about the general effects of principles that would permit or forbid certain actions, not simply about the consequences of individual actions. The conflicting reasons are, on the one hand, reasons for wanting a stronger form of security and, on the other, reasons for wanting, in general, to be able to rely on having benefits of a certain kind. The second point is that a principle-based account would “collapse” into an act-based account only if it would not be reasonable to reject a principle permitting an exception whenever the reason that the person whose security is violated has to object to the impact on him *of that particular action* are less strong than the reasons of others to want the benefit that *that particular action* would bring them. As we saw in the case of exceptions to a principle of fidelity to assurances, however, it would be reasonable to reject such a principle as long as that person’s security-based reasons have at least some weight, and certainly reasonable to do so if, as Mill maintained, these reasons are particularly strong.

Contractualism and Conventions

I have mentioned three reasons for concern with principles rather than just with the consequences of particular actions: the need for various forms of security, intrapersonal aggregation on the side of patients, and intrapersonal aggregation on the side of agents. These three cases differ in ways and are more complex than I have previously realized.

Consider first security. We have good reason to want it to be the case that individuals do not physically harm us or lie to us or fail to fulfill expectations that they have led us to form, and therefore good reason to want them to accept and be governed by principles that rule out doing these things because of the reasons we have not to want them done. But as I have said, reasons for wanting others to accept and be guided by these principles go beyond the effects on us of particular actions and include reasons for wanting security in the form of assurance that these things will not happen. The fact that they accept these principles provides this security and also, as I have said, a form of recognition that is also important to us. We would have the kind of security we would most like, of course, only if everyone, or at least everyone with whom we interact, accepted these principles. But our reason for wanting this in the case of a particular individual does not depend on the level of acceptance by others. Moreover, in the cases just mentioned, our reasons for objecting to the kinds of harm that we have reason to want assurance about—being killed or hurt, being misled, and so on—do not depend on the acceptance of principles ruling them out. These things are, as it were, morally *malum in se* not morally *malum prohibitum*.

But not all forms of security are like this. Privacy, for example, is importantly different. Privacy—the ability to avoid others observing or having knowledge about parts of our lives—is something we have reason to want independent of any social convention providing it. But we

need such conventions to define the ways in which we can have privacy, the ways in which others are not free to observe us, to read our messages, and come to have knowledge about other aspects of our lives, and these rights can be defined in different ways. Even if they do not have other objectionable consequences, invasions of privacy are wrong simply because they violate necessary and reasonable conventional norms. They are violations of what I have called the Principle of Established Practices.³⁸ This principle holds that if some practice is needed to avoid a result that we each have strong reason not to want, and if some practice of this kind that is not itself unnecessarily burdensome or unfair is widely accepted, then it is wrong not to comply with that practice. The wrongfulness of violating this principle has a contractualist explanation. Given that such a practice is needed to avoid results that each of us has reason to object to, each of us has reason to reject a principle permitting violations, and these reasons are sufficient if the practice in question is not unnecessarily burdensome or unfair.

In thinking about cases of what I called intrapersonal aggregation on the patient side I have failed to take account of the fact that most cases of this kind depend on conventions in the way just described. Actions that violate established norms for dealing with climate change or ozone depletion, for example, would not be wrong simply because of their individual consequences, which are generally trivial. Rather, they would be wrong because they violate established norms that are justified by the need to protect us against these harms. Individual acts violating these norms have only trivial consequences, so intrapersonal aggregation on the patient side is needed to explain why these norms are necessary and justified.

³⁸ *What We Owe to Each Other*, p. 339.

Classic examples of “what if everybody did that?” are instances of this. They are violations of the Principle of Established Practices. The question presumes that “everybody” is not “doing that,” and that they are not doing so because of their recognition of the problems that would occur if many people did.³⁹ That is to say, the question presupposes an established norm that is justified and being observed by many for that reason, despite the advantages to them of violating it. This last condition, about the reasons for complying, is what marks the difference between cases in which “What if everybody did that?” is a relevant objection and cases in which it is not.⁴⁰

Cases in which such norms are needed but have not been established are importantly different.⁴¹ What is wrong in these cases is not, or not just, failing to do what a norm would require but also failing to do one’s part in getting such a norm established. This is most obvious

³⁹ A point made by Liam Murphy in “Nonlegislative Justification,” *Principles and Persons: The Legacy of Derek Parfit*, text accompanying note 65.

⁴⁰ This is why it is important, in interpreting Kant’s moral theory, to see that what is “universalized” is not acting in a certain way, but rather accepting a maxim, that is to say a principle about the sufficiency of certain reasons for behaving in some way. Many unobjectionable acts, such as driving through Harvard Square at 5am every Wednesday, lead to difficulty if everyone tries to do them. But no problem arises from general acceptance of a maxim according to which convenience, habit, or pure amusement are taken to be sufficient reasons to do this.

⁴¹ Compare the two parts of what Rawls calls the Natural Duty of Justice: “This duty requires us to support and to comply with just institutions that exist and apply to us. It also constrains us to further just arrangements not yet established, at least when this can be done without too much cost to ourselves.” *A Theory of Justice*, 1st ed., p. 115. Like this second part of Rawls’ natural duty, what I am suggesting in the text to is what Kant called an imperfect duty. It requires individuals to take some steps to help get needed norms established, but does not specify what needs to be done or at what cost. One way to incorporate this within a contractualist framework would be to extend that framework to include the assessment of maxims as well as the wrongfulness of actions. A minimal version of my present point would be that it is wrong to hold a maxim of giving no weight at all to reasons for getting needed norms accepted. The claim would be that a “norm-regulating principle” that permitted this could reasonably be rejected. This fits with one interpretation of what Kant says regarding imperfect duties to promote the happiness of others and the development of one’s own talents: that adopting such a maxim would involve a “contradiction” since one needs developed talents and aid from others. Insofar as Kant’s argument depends just on what one can will without contradiction it seems to be restricted to this limited conclusion. Contractualism’s reliance on the sufficiency of some reasons given others opens the possibility of going beyond this. (See the contrast with Kant discussed on pp. 34-35.) Similarly, Barbara Herman suggests that the extent of Kantian imperfect duties to aid are defined by balancing reasons: some considerations are sufficient reason for not aiding while others are not. See her *Moral Literacy* (Harvard University Press, 2007), p. 297.

in cases in which there are many different ways of preventing the harmful effects, none of which is being followed. In such a case, silently and privately going along with one of these alternative schemes is no better than doing nothing. (And it is reasonable to reject a principle permitting one *simply* to do this.) What is required is to make at least some effort to get some such norm established by, for example, making known that one believes such a norm is needed and indicating one's willingness to comply, perhaps by publicly doing so as an example and indication of good faith. The contractualist argument for this requirement does not "legislate" which norm to follow. It just requires one to participate in the social process of selecting one and getting it established.

In cases in which there is only one obvious solution, such as simply refraining from a certain kind of action, it is easy to overlook the crucial importance of getting a norm established. The examples Parfit lists, "pollution of air, food or water" may appear to be cases of this kind.⁴² But in these cases too, if no norm of the needed kind is established then *simply* privately and quietly "doing the right thing" is not enough. It is reasonable to reject a principle permitting individuals merely to secretly comply with a norm that is not being followed by many others.

In cases of the kinds just discussed, the consequences of what many agents do add up to serious burdens on each individual. In cases of what I call intrapersonal aggregation on the agent side, actions that a principle requires an agent to perform, and the need to stand ready to perform such actions, add up to make that principle burdensome to live up to. These burdens may or may not add up to enough to make it unreasonable to reject a principle permitting non-

⁴² Parfit *On What Matters*, Volume. Two, pp. 205-206.

compliance. But even when they do not, they can be significant in ways like those we have just been discussing. The burdensomeness of a principle of aid may be increased simply by the increased number of cases in which aid is required. A principle requiring aid in Singer's pond case, for example, might be very burdensome for someone who lived near a pond where children often played unsupervised, and who frequently walked by it. This burden might be increased by the failure of other passers-by to do their part.⁴³ Whether this is so or not, however, and even if the burden is not sufficient to undermine the wrongfulness of failing to aid, this burden is a reason to reject a principle permitting people to take no steps to establish ways of dealing with the problem that would be less burdensome for those giving aid and perhaps more effective in protecting potential victims. These might include such things as a fence, a lifeguard, or more general acceptance of requirements that parents not allow young children to play there unsupervised. The relevance of such alternative measures, and the importance of willingness to contribute to getting them established, is even more obvious in cases such as international aid that Singer generalizes to.⁴⁴ Leaving aside the question⁴⁵ of when giving aid is still required, despite the burden, the point, following on from previous discussion, is just that compliance and non-compliance are not the only alternatives.

Reasonableness, Reasons, and the Question of Circularity

I said earlier that contractualism offers an appealing answer to the question of acceptance because it holds that moral standards are rules for living together that can be

⁴³ As discussed by Liam Murphy in "The Demands of Beneficence," *Philosophy & Public Affairs*, 22 (1997), 267-292.

⁴⁴ A way of arriving at a version of Murphy's view that duties to aid are collective obligations. See "The Demands of Beneficence" and *Moral Demands in Nonideal Theory* (Oxford University Press, 2000), esp. Chapter 10.

⁴⁵ To which Murphy's collective principle of benevolence is one answer.

justified taking into account the reasons of all to for wanting their lives to be affected in some ways and not in other ways. This idea may be appealing, but it is also vague, since it says nothing about how multiple and conflicting reasons are “taken into account” in justifying principles. My statement of contractualism was more specific. It said that the rightness or wrongness of an action depends on whether a person in some position could reasonably reject a principle that permitted it, if this person also had the aim of finding principles that others also could not reasonably reject. This stipulated aim means that the reasonableness of rejecting a principle must take into account the reasons of others to want the opportunities or protections that such a principle would provide for them. This interactive character of the account is crucial to its plausibility as an account of the content of morality and to its appeal as an answer to the question of acceptance.

There are, however, various reasons for being suspicious of the appeal to reasonableness in the contractualist formula. One is reason is the suspicion that reasonableness is itself a moral notion, the appeal to which may render my reductive strategy circular. There may be problem here, but it does not lie with the appeal to reasonableness. The idea of reasonableness that I am employing is simply the general idea of what a person has sufficient reason to do given a contextually specified range of considerations as reasons.⁴⁶ For example, what it is reasonable for you to believe in a given situation is what you have sufficient reason to believe given the information available to you in that situation. The claim that a person is unreasonable in refusing to believe the things that are clearly supported by the available

⁴⁶ For further discussion see my “Varieties of Reasonableness,” in *Conversations in Philosophy, Law, and Politics*, Ruth Chang and Amia Srinivasan, eds., (Oxford: Oxford University Press, 2024) pp. 265-276.

evidence is not a moral claim. 'Reasonable' can also be used to express moral approval or disapproval, as when a person is said to be unreasonable, meaning by this that the person does not give the interests of others the weight that one should. But the moral content in these expressions of approval and disapproval is provided not by the idea of reasonableness itself but by the set of reasons that are presupposed in that context and assumptions about the weight that a person *should* give to these reasons.

Similarly, the moral content in my statement of contractualism is not provided by the idea of reasonableness itself but rather by the specification of the relevant range of reasons and by assumptions about the importance they are to be given. The reasons in question are specified to be personal reasons that individuals have because of ways that their lives would be affected by living under a principle, and facts about the reasons of this kind that others have. The starting points of contractualist justification consists of facts about the sufficiency of various reasons of this kind as reasons for rejecting principles taking into account the opposing personal reasons of others. After saying more about the nature of these reasons and the resulting complexity of judgments about their sufficiency I will return to the question of circularity.

The ways of being affected that give individuals personal reasons to object to or favor a principle sometimes come in amounts, such as hours of pain or years of life. Many of the cases Derek Parfit discusses, for example, are stated in these terms. In such cases, it is natural to think of the relative strength of these reasons as determined simply by the quantity of these physical or psychological states. This is plausible in many cases. One has stronger reason to want to avoid five hours of pain than to avoid five minutes of the same pain, and more reason

to want to have five years of additional life than to have five days. The clarity and apparent explanatory power of contractualism as a reductive theory is maximized if it is seen as explaining facts about right wrong on the basis of judgments of this kind about the sufficiency of personal reasons.

But judgments about the sufficiency of reasons for rejecting principles can be more complicated. For one thing, because the question at issue is whether principles are to be accepted or rejected, reasons on both sides will include not only the desirable or undesirable effects of particular actions but also considerations of what Mill called “security.” The inclusion of reasons of his kind does not render the reductive theory circular insofar as judgments about the sufficiency of reasons are not in themselves judgments of rightness and wrongness. But the complexity of these judgments does reduce the clarity of the view and perhaps also its explanatory power.

It is important to see that even when the facts that reasons are are all of a simpler, quantitative sort, judgments about their strength or sufficiency involve a substantive normative step. A normative fact about the strength of a reason is not the same as a naturalistic fact about years of life or hours of pain. Moreover, the strength of such a reason for rejecting a principle is not always determined simply by the magnitude of the relevant underlying property. For example, as Parfit argues, and I am inclined to agree, the reason to bring about an increase in years of life or relief from pain may vary depending on how badly off the person will be without this help. It is natural to state this in terms of a relation between two quantities: on the one hand, the quantity of pain, or length of life, or of some material benefit, and, on the other, the strength of a reason to want this or to object to it.

This way of talking about reasons can involve a kind of doubling. We often say that some fact is a reason, as in “The fact that the stove is hot is a reason not to put your hand on it.” Many facts that are reasons are naturalistic, but normative facts can also be reasons. For example, the fact that my wife has good reason to want my help this afternoon would be a reason for me to stay home. We also commonly say that such facts, naturalistic or not, “provide” reasons for an agent to do something. It is reasons in this latter sense (the ones that are “provided”) that are said to be “balanced,” to “outweigh” each other, or to be “sufficient.”

But this second idea of a reason, as a normatively significant thing that has a certain normative weight or strength and can be sufficient or not sufficient, also seems to me confusing. Being a reason is a relation that the facts that “are reasons” in the first sense just distinguished stand to certain actions or attitudes under certain conditions. A fact’s having greater “weight” than another is just a matter of its being a sufficient or even conclusive reason for a certain choice under conditions in which the other fact is a reason for an incompatible choice. I don’t think that we have any clear idea of the weight or strength of a reason apart from hypothetical judgments about its sufficiency or conclusiveness in such cases.⁴⁷ Normative or non-normative facts do not “provide” reasons. They are (or are not) reasons, which may or may not be sufficient in a given context. “Balancing” is not a matter of *comparing* two normative entities or quantities (reasons in the second sense just distinguished.) It is just the process of deciding whether certain facts are or are not sufficient reasons for rejecting a principle taking into account other considerations that are in play.

⁴⁷ Or so I argue in *Being Realistic about Reasons*, Chapter 5.

Why would it be a problem for contractualism if judgments about the sufficiency of reasons for rejecting a principle involve moral judgments? It is natural to say that the problem is that this would make contractualism circular. But the objection is more general than that. The general form of the objection is that appealing to starting points of a certain kind makes an account unsuccessful (or less successful) in achieving its aims. This obviously depends on what those aims are.

This paper involves three forms of “justification” that have different aims. The aim of first-order moral argument is to determine whether certain actions are wrong, and if so to explain why they are wrong. Such an account is therefore not successful (indeed, “circular”) if at some point it appeals to an answer to the very question it aimed to answer. But the fact that such an account appeals more generally to moral notions, such as rights and duties, is not a problem. That is the kind of explanation that is needed. The fact that Intuitionism, for example, explains rightness and wrongness in terms of *prima-facie* duties may limit its explanatory power, but it does not make it circular.

The aims of contractualist justification, in terms of reasonable rejection, face in two directions. On the one hand, contractualism aims to provide an account of the content of the morality of right and wrong—an account of what it is to be right or wrong. As I have said above, in order for such an account to be plausible, the factors that it takes rightness and wrongness to depend on must be ones that strike us as morally relevant. It must explain why the most obvious conclusions of ordinary moral thinking seem obvious and fit with the way we think about unclear cases. Insofar as it does these things a contractualist answer to the question of content offers, or entails, a first-order moral theory. Looked at in this way, it is not a problem

that it appeals to moral claims. This may make it less complete, as in the case of Intuitionism.

But lack of completeness is not a fatal flaw.

The aims of contractualism as an answer to the question of content go beyond this, however. It aims to characterize what it is to be right or wrong in a way that answers the question of acceptance, making clear the reasons we have for regarding moral conclusions as normatively authoritative. This aim presupposes that such an answer is needed—that the fact that an action would be wrong or would violate some right is not in itself a complete explanation of the reason one has not to do it. So an account that appealed to facts about rights—for example, by holding that it would not be reasonable to reject a principle permitting someone to keep his organs because people have a right to their organs—would not succeed in this aim.

The fact that an account of the subject matter of morality has broadly speaking “a moral character” need not give rise to this problem. Requiring the personal reasons of all affected individuals to be taken into account and similar reasons of different individuals to be treated as equally important are plausibly said to be moral conditions. They are necessary in order for contractualist justifiability to be a plausible account of morality.⁴⁸ The fact that contractualist justifiability has this “moral character” does not undermine it as a basis for answering the question of acceptance. It would do this if the proposed answer were that others are morally entitled to treatment that can be justified in this way. But this is not the answer that contractualism offers. That answer, rather, is that treating others in ways that can be justified in

⁴⁸ As I said earlier in contrasting the appeal of contractualism’s interactive idea of justifiability with that of Gaus’s non-interactive version.

the way contractualism specifies is an appealing way of living with them. If they complain, one can say, “This is justified in a way that takes your reasons for objecting to how it affects you just as seriously as anyone else’s.” By contrast (it is assumed) the response, “But he has a right to do that” needs further explanation of what it is to have a right and why that matters. This is what contractualism seeks to supply.

Parfit on the “Greater Burden Claim”

The complexity of judgements about the sufficiency of reasons for rejection has implications for the question of demandingness. In particular, it has implications for the interpretation of what Parfit calls my Greater Burden Claim. This is the claim that “it would be unreasonable to reject a principle because it would impose a burden on you when every alternative principle would impose a much greater burden on others.”⁴⁹ As Parfit interprets it, this claim is very demanding. In his Case One, for example, if Grey gave one of his organs to White, Grey would shorten his life by a few years, but he would also give White many more years of life. As Parfit says, “Most of us would believe that, though it would be admirable for Grey to give his organ to White, Grey is not morally required to make this gift.”⁵⁰ But Parfit interprets my view to imply that White could reasonably reject a principle permitting Grey to keep his organ because the benefit that his keeping it would deny White is much greater than the benefit that the organ would have for Grey.

This case has special drama because it involves Grey’s bodily organs. But the conclusion Parfit attributes to contractualism is very general. He takes it to imply that one must always

⁴⁹ *On What Matters*, p. 192.

⁵⁰ *Ibid.*

give a benefit to some person whenever this benefit would be significantly greater than the loss to oneself. Call this the General Narrow Benefit Requirement. It is “general” because it would apply not only to a person’s bodily organs but also to any asset under an agent’s control, including financial assets, other forms of property, and the individual’s own time and energy. The benefits and burdens on which this requirement makes the rejectability of a principle depend are “narrow” because they include only the consequences of a particular action for the agent and others.

Contractualism does not imply that it would always be wrong to violate this requirement. It would have this implication only if it would be reasonable to reject any principle that permitted an agent to violate it. Parfit makes it appear that it would be reasonable to reject any such principle because he focuses only on reasons arising from the benefits at stake in a particular case.⁵¹ But the reasons referred to in the passage that Parfit cites from *What We Owe to Each Other* were intended to include reasons arising from the general effects of living under a principle.⁵²

Reasons to reject a principle permitting violation of the General Narrow Benefit Requirement are the reasons to want the general ability to demand benefits from others

⁵¹ This is the same mistake that Mill faults Bentham for committing in the passage I quoted earlier.

⁵² I make no claim to originality here. Many others since Mill have appealed to considerations of burdensomeness, or the importance of “space to live one’s life” as setting limits on moral demands. Peter Railton appeals to similar considerations within a consequentialist framework to show that consequentialism must allow room for special relationships. See Railton, “Alienation, Consequentialism, and the Demands of Morality,” *Philosophy & Public Affairs* 13 (1984) pp. 134-17, and Kumar, “Defending the Moral Moderate; Contractualism and Common Sense,” *Philosophy & Public Affairs* 28 (2000), pp. 295-296. Elizabeth Ashford also suggests that appeal to the general consequences of living under a principle, rather than just the effects of particular actions, is the most promising strategy for a contractualist response to this problem. See <https://plato.stanford.edu/entries/contractualism/#IrrIntent>, section 9. What I am doing is just developing a particular version of this strategy and considering some objections and alternatives. For discussion of this general issue, see Samuel Scheffler, *Human Morality* (New York: Oxford University Press, 1992).

whenever one would benefit more from having them. The reasonableness of rejecting a principle on this ground must take into account the opposing reasons for wanting the form of security that involves being free from the general demands that this entitlement would impose.

Focusing, as Parfit does, on the benefits and burdens in a particular case, it is natural to think of the opposing reasons at stake as reasons of different individuals in opposing positions: the reasons of a beneficiary such as White versus those of a potential benefactor such as Grey. But the general effects of living under a principle include being both a potential beneficiary and a potential benefactor. So the acceptance of a principle and its rejection have costs and benefits for each person. These costs and benefits will interact in different ways in different cases.

In Parfit's Case One the question is the reasonableness of rejecting a principle permitting individuals to keep organs that would give more years of life to others. The reason to reject such a principle is a reason to want the ability gain extra years of life by demanding organs from others. The question is whether this is a sufficient reason to reject the principle given the countervailing reason to want stable possession of one's organs. But if the principle permitting individuals to keep their organs under such conditions is rejected, and one must give organs up whenever others would benefit more from having them, then one could not count on any specific number of years of life from the organ one would acquire. One would be entitled to keep it only so long as no one came along who would benefit more from having it. This significantly weakens one's reason for rejecting the principle permitting individuals to keep their organs in such a case, and seems to me to make it insufficient, although this may be disputed.

There are three points here. The first is that Parfit makes it seem as if I am committed to the Greater Burden Claim by focusing only on reasons for rejecting a principle based on the effects of particular actions that it would permit, neglecting the general effects of living under such a principle. This is the same mistake Mill accuses Bentham of making. Second, as a result, he neglects the fact that when these general effects are being considered every individual is both a potential benefactor and potential beneficiary of a principle and thus has the same reasons for accepting or rejecting it. Third, this means that he also neglects the fact that in cases like his Case One the benefit to be gained by having the ability to demand organs from others is undermined by the fact that, *qua* potential benefactor, one might have to pass this organ on if there were someone who would benefit more from having it.

It would be a mistake to neglect this undermining effect and therefore reject a principle permitting potential benefactors to keep their organs, assuming that this would allow one to gain extra years of stable possession of an organ if one would benefit more than some potential donor. This would be the same mistake as the one involved in Kant's famous example of the lying promise.⁵³ Kant imagines a person willing the permissibility of gaining money by making a promise to repay it even though one has no intention of doing so. His point is that if this were permissible (and generally understood to be so) then making such a promise would not be an effective way of getting a loan. Kant describes this as a contradiction: willing the permissibility of acting in a certain way for a reason that presupposes that acting in this way is (at least generally regarded as) impermissible. (In Parfit's Case One, the contradiction would be willing

⁵³ *Groundwork of the Metaphysics of Morals*, Ak. 422.

the impermissibility of keeping one's organs for a reason that presupposes the permissibility of keeping them.)

But the impermissibility of an action is not settled by the fact that willing its permissibility (or rejecting a principle that would permit it) without taking these undermining effects into account would be involve such a "contradiction." Rather, permissibility depends on whether, taking these effects into account, an individual would have sufficient reason to reject a principle that permitted acting in that way.⁵⁴ It seems to me that in the present case, taking these effects into account, this reason is not sufficient.

Not all the cases fitting the general pattern of Parfit's Greater Burden Claim involve undermining effects of this kind. In Peter Singer's example of rescuing someone drowning in a pond, the benefit of having a claim to be rescued in such a case is not affected by a principle requiring a passerby to rescue such a person. This requirement does have costs for potential rescuers (and hence for those who are rescued, since that they may on some occasions be required to rescue others.) But it is stipulated that these costs are relatively minor, and the reason to want rescue to be required remains strong and is clearly sufficient reason to make it reasonable to reject a principle permitting a passer by not to rescue a drowning person under such circumstances.

This analysis applies also to other more specific duties to aid. The general point is that the wrongfulness of failing to aid does not depend only on the severity of the need and the cost

⁵⁴ Part of the appeal of Kant's view may be that, at least in the case of perfect duties, it seems to explain rightness and wrongness in terms of what one can consistently will, avoiding appeal to messier claims about the balance of reasons. This is a fundamental difference between Kant's moral theory and contractualism, according to which these claims cannot be avoided. For further discussion, see my "How I am not a Kantian" in Parfit, *On What Matters*, Volume Two, pp. 116-139, esp. pp. 122-123.

of providing it on a specific occasion. The question is whether the reasons a person has for wanting the general ability to demand aid in circumstances of that kind is sufficient reason to reject a principle permitting a potential benefactor not to provide it, taking into account the reasons that a person has to want the security of not being subject to the general requirement to give aid when those specific conditions are met. (This is parallel to what I said earlier about the permissibility of violating a promise in order to aid a third party.)

This is my best attempt to explain why contractualism does not entail the extreme demands of Parfit's General Narrow Burden Requirement but does allow for more specific duties to aid, which may themselves be quite demanding.⁵⁵ The grounds I have offered for rejecting these extreme demands, and the method I have offered for determining the demandingness of more specific principles both rely heavily on individuals' reasons for wanting various forms of security. This reliance may seem open to a number of objections.

One objection is that these reasons for wanting to be free from intrusive obligations may seem too selfish. One might, for example, have this reaction to Bernard Williams' view in which interference with appeals to one's "ground projects" may appear to trump any competing obligations.⁵⁶ This criticism is blunted by the interactive character of contractualist justification, which requires reasons for wanting the "security" of being immune to demands from others to be balanced against the reasons that these others have to want the form of security that lies in being able to rely on assistance of the required kind. So the interactive character that generates, or at least aggravates, contractualism's tendency toward

⁵⁵ Thus fitting with Ashford's view.

⁵⁶ See Williams, "Persons, Character, and Morality."

overdemandingness also makes a security-based response more plausible by providing a way of determining its limits.

But the balancing of reasons that this involves is invites further objections. One is that a judgment about the sufficiency of these reasons for rejecting a principle permitting a person not to provide aid in certain circumstances is really just a moral judgment about “how much we owe one another.” Some reassurance on this point might be provided by the fact, pointed out above, that when reasons concerned with the general effects of living under a principle are concerned, each person has reasons both as potential beneficiary and as potential benefactor. So the question of the sufficiency of one such reason given another opposing reason is question is not only about the balancing of reasons of different individuals but arises also as a question about the prudential reasons of a single person. If these two questions must have the same answer this would provide some assurance that an answer to the former (interpersonal) question is not a moral judgment.

But need these two questions have the same answer? If we focus, as Parfit did, simply on the costs and benefits of particular actions, then it would seem that they must. But this is less clear when what is at issue are the costs and benefits of living under different general principles. Consider the following two questions:

Q(inter): Is the benefit of having the ability to demand another person’s organ under conditions C a sufficient reason to reject a principle permitting a potential donor to keep his or her organ under those conditions, taking into account the reasons that potential donors have to want to be entitled to keep their organs under such conditions?

Q(intra): Thinking of him or herself alone, is the benefit of having the ability to demand an organ from another under conditions C sufficient reason for a person to reject a principle that would allow him or her to keep their organ under these conditions?

Why might one think that the answers to these two questions need not be the same? If the answer to Q(intra) is “no”—that the advantages of being entitled to demand an organ under conditions C are not sufficient reason for a person to give up, for him or herself, the entitlement to keep an organ in such a case—then it would seem to follow that the answer to Q(inter) is also “no.”

But if the answer to Q(intra) is “yes”—if the advantages of an entitlement to demand an organ under such conditions are sufficient reason to give up any entitlement to keep one’s organ in such a case—then it may not follow that the answer to Q(inter) is also “yes.” It is quite possible that while the advantages of being able to demand an organ are sufficient reason to give up the entitlement to keep one’s organ under such conditions, it is also the case that the advantages of being able to keep one’s organ are sufficient reason to forego the entitlement to demand an organ from someone else if the need should arise. Either personal choice may be reasonable.

If this is so, then it does not seem to follow from that fact that a person has sufficient reason in Q(intra) to opt for the entitlement to demand an organ under conditions C that the reason to want this entitlement is sufficient reason to reject a principle permitting a different potential donor to keep an organ under those conditions. In Q(intra) the person is deciding for him or herself which option to choose. But Q(inter) poses a different question. Rejecting a principle that would permit someone to keep their organ means denying them an option that

they might, as we just said, have sufficient reason to prefer. Q(intra) is a matter of the sufficiency of a reason for choosing for oneself. Q(Inter) is a matter of the sufficiency of a reason for legislating for all. The sufficiency of a reason for rejecting a principle in the latter case must take into account not only opposing reasons that individuals have for wanting the security of control over their organs, but also their reasons for wanting to be able to decide for themselves which of these sufficient reasons for opposite choices to act on. This reason would not be a factor if, returning to Q(Intra), the person did not, as I assumed earlier, have sufficient reason to choose either option.

This seems to me to explain the felt difference between Q(intra) and Q(inter)—between the sufficiency of a prudential reason given conflicting prudential reasons and the sufficiency of one person's personal reason given the opposing personal reasons of other individuals. The fact that there is this difference undermines the reassurance that the identity of answers to these questions might have offered that an answer to Q(Inter) is does not rely on a moral judgment. It does not show that an answer to Q(inter) must be a moral judgment—an answer to this question might just be a normative judgment about the sufficiency of certain reasons. But the question remains.

Even if the judgments of sufficiency on which contractualism relies are not moral judgments, however, the unclarity of these judgments is a further worry. In some cases this lack of clarity may not be a problem: in many cases the limits of duties to aid may just be unclear. But this is not always the case. As Parfit says, in his Case One it seems clear that Grey is entitled to keep his organ, although it might be very good of him to give it to White. One may question whether this seemingly obvious moral fact really depends on the less clear question of the

relative strength of reasons for wanting secure possession of one's organs and reasons to want the ability to demand help from others when one needs it and providing it would cost them less? This mismatch in clarity may seem to cast doubt on contractualism's claim to be a plausible answer to the question of content.

This raises the question of what the alternative answer is supposed to be. If the alternative is that the rightfulness of Grey's keeping his organ is simply a moral fact that is not dependent on the strength of conflicting personal reasons⁵⁷, then the question of emptiness raised by Anscombe and Bentham re-emerges. Why should those who need organs take this claim of moral entitlement as a justification for refusing to provide the aid they need? To respond to this question we need some alternative answers to the questions of content and acceptance—some account of what we are thinking about when we arrive at conclusions about what we owe each other that links up with an explanation of why we should take such conclusions as authoritative standards of conduct.

One account would be that what we are thinking about in such cases is just how best to understand a set of generally accepted standards about how we are expected to treat one another. We value the protections these standards offer us. And because we value this protection, we would like to believe that everyone has reason to accept these standards even though their content does not depend on facts about individuals' reasons in the way I have described. From this point of view, contractualism is an attempt to ground this hoped-for answer to the question of acceptance by giving a revisionary account of the content of morality that builds in dependence on individuals' personal reasons.

⁵⁷ As claimed, for example, by Judith Thomson, *The Realm of Rights*, p. 143.

Interpersonal Aggregation Reconsidered

The line of thinking about interpersonal aggregation that I set out in *What We Owe to Each Other* now seems to me mistaken in several respects. The first is that I was thinking of the problem as the apparent inability of contractualism to deliver conclusions that seemed, intuitively, to be clearly correct. It seemed clear, for example, that when faced with a choice between rescuing a single stranger on one rock or five strangers on a different rock it would be wrong not to save the greater number. So the problem for contractualism, as I was thinking of it, was that it failed to give the intuitively correct answer, undermining its claim to be an account of the content of moral right and wrong.

It now seems to me that this underestimated the problem. The conclusion that the rightness or wrongness of an action cannot depend on the number of individuals who would be affected is not only in conflict with common sense morality but also incompatible with a basic idea of contractualism itself, namely the idea that moral standards must justified taking the reasons of all affected parties into account. A principle that says we may save the smaller number in cases of the kind described is not justifiable in this way: it does not take into account the reasons of all. I failed to see that such a principle could reasonably be rejected because, contrary to this basic contractualist idea, I was interpreting the reasonableness of rejecting a principle in a strictly bilateral way, as a matter of the sufficiency of one person's reason for rejecting the principle taking into account only the single strongest opposing reason.

The "tie-breaking" argument that I offered for saving a group of two rather than only one recognized, implicitly, the need to take a broader view of the reasons involved in the

reasonableness of rejecting a principle.⁵⁸ In that argument I said that one of the two could object that a principle permitting the rescuer to save the single person failed to take his or her reasons into account as making a difference. There is something right about this. But the way I put it was mistaken in several respects.

As Michael Otsuka⁵⁹ pointed out, the argument I gave did not show that any individual had a sufficient personal reason to reject a principle permitting a rescuer to save the single person. Each of the two individuals on the same rock have a personal reason to reject such a principle, but neither of these reasons, taken alone, is sufficient reason to reject such a principle, taking into account the single person's equivalent reason to be saved. The complaint, by one of the two, that his or her presence is not making a moral difference, is not a further personal reason to reject a principle permitting the rescuer to save the one. It is rather a claim that the strictly bilateral interpretation of reasonable rejection does not live up to the contractualist idea that moral principles are ones that can be justified taking the reasons of all into account. I failed to see this more general implication of the objection.

A better contractualist explanation of how the number of individuals involved can make a difference needs to hold fixed certain basic contractualist ideas. The first idea is that moral principles must be justifiable taking the relevant reasons of all affected parties into account. The second is that an action is wrong only if any principle that permitted it would be one that an individual in some position would have sufficient personal reason to reject, taking into account the reasons of other individuals.

⁵⁸ *What We Owe to Each Other*, pp. 232-234.

⁵⁹ Otsuka, "Scanlon on the Claims of the Many versus the One," *Analysis* 60 (2000) 288-293.

What I will not hold fixed is the strictly bilateral interpretation of reasonable rejection, which holds that reasonable rejection is a matter of the sufficiency of an individual's personal reason to reject a principle, taking into account only the strongest individual reason against rejecting it. Although I have relied on this bilateral interpretation in the past, it is, as I have said, at odds with the basic idea that the reasonableness of rejecting a principle must take account of the conflicting reasons of all affected parties, giving relevant reasons of the same kind the same weight. This modification is consistent with, and even required by, the core ideas of contractualism: that only the personal reasons of individuals are relevant to the rejectability of principles, and that an action is wrong if any principle that permitted it would be one that an individual in some position would have sufficient personal reason to reject, taking into account the personal reasons of others.⁶⁰

The need to take into account this plurality of reasons requires recognizing the fact that the reasons a person has can "add up." That is to say, the number of reasons a person has for doing something or not doing it can make a difference to whether that person has sufficient reason to do it.⁶¹ For example, a father's reasons for going home early rather than staying to work on revising his paper might include the fact that his wife would like help with a project and that his daughter would like him to listen to her music practice. The fact that he has both of these reasons for going home can make a difference to whether his reason to work longer on his paper is a sufficient reason to stay. Taking these matters into account, we can explain how

⁶⁰ I believe this modification is therefore consistent with the proper understanding of what Parfit called my "individualist restriction," which he urged me to give up. See *On What Matters, Volume 2*, pp. 193ff, and pp. 113-117 of "Improving Scanlon's Contractualism," in Markus Stepanians and Michael Frauchiger, eds., *Reason, Justification, and Contractualism: Themes from Scanlon* (Berlin: de Gruyter, 2021), pp. 109-117.

⁶¹ As Ralph Wedgwood points out in "The Reasons Aggregation Theorem," *Oxford Studies in Normative Ethics Vol. 12*, pp. 127-148.

the number of individuals involved can make a difference in the case of a rescuer in a boat who has to decide between saving one stranger on a rock with the tide rising or saving a larger number of strangers on a different rock.

Consider first the *permissibility* of failing to save a single person in order to rescue a larger number of others. It seems obvious that this is permissible, and I need to explain how contractualism can explain this obvious fact. In contractualist terms, the question is whether someone in the position of the single person could reasonably reject a principle permitting a rescuer to save the larger number—that is to say, whether that person’s reason to want to be saved is a sufficient reason to reject such a principle, taking into account the opposing (individual) reasons of members of the larger group. In order to reach the conclusion that it would not be sufficient (and hence that saving the larger number is permissible) we need not combine these opposing reasons into one larger disvalue. Each of these reasons is itself already included in the set of considerations that contractualism requires the single person to take into account and on which the reasonableness of that person’s rejection therefore depends. As reasons for that person, they add up in the way that reasons generally do, as I pointed out above in my example about going home early. In this case the reasons add up as reasons for the single person *not* to reject the principle—that is to say, reasons that can make it unreasonable for the single person to reject the principle because the *cost* of rejecting it, measured in terms of the reasons against this, would be too high.

Why is this cost too high? I do not have a general explanation of how multiple opposing reasons can add up to make a single person’s reason for rejecting a principle insufficient. But in the case at issue these reasons are all the of the same kind—reasons to want to be saved—that

therefore all have the same significance in contractualist justification. Given that this is so, we can explain why these reasons make that single person's personal reason to reject the principle insufficient by employing the idea put forward in slightly different forms by Frances Kamm and Rahul Kumar, which I will put by saying that these reasons can counterbalance one another in contractualist justification.⁶² Given that they are equivalent, the single person's reason to want to be saved and the equivalent reason of any single member of the larger group "counterbalance" each other.⁶³ This leaves the single person with no uncounterbalanced reason for rejecting the principle. The uncounterbalanced reason of any other member of the larger group against doing so thus makes rejection unreasonable.

This analysis captures the thought behind the "tie-breaking" argument I offered in *What We Owe to Each Other*, but with several important differences, which respond to Otsuka's objections. First, my version of the tie-breaking argument slipped into considering reasons for saving the single person or the two others rather than reasons for rejecting or not rejecting a principle. By contrast, the argument I have just given stays within the contractualist framework by considering the sufficiency of one person's reason for rejecting a principle permitting others to be saved—specifically, the sufficiency of the single person's reasons for rejecting a principle permitting the rescuer to save the two. Within this framework the reasons of the two to want to be saved occur as reasons *for the single person*, namely reasons not to reject the principle.

⁶² This is a substantial assumption. It would not be true if, for example, the individuals on one rock had stronger reasons to want to be saved, perhaps because they were much younger.

⁶³ Kumar says that these reasons can "neutralize" each other. ("Contractualism on the Shoals of Aggregation," pp. 140-141) Frances Kamm made a similar point in "Is It Right to Save the Greater Number?" in *Morality, Mortality*, Vol. 1, p. 101. She prefers the term "balanced" rather than "neutralized," to make clear that "neutralized" reasons remain reasons, and I agree. The terms, "balanced" or "counterbalanced," make clearer that what is being claimed is something about the role of the reason in a particular contractualist justification. I believe that this was also Kumar's intent.

As reasons for this individual they add up in the way I have argued that diverse reasons of a single person normally do. This locates the aggregation in the proper place, a place where it should be uncontroversial. (By contrast, the strictly bilateral interpretation, requiring the person to consider only one of these reasons, seems irrational.) Finally, unlike the tie-breaking argument, this argument generalizes to cover cases in which there is more than one person in each group.

Now consider the question of whether it would be wrong not to save a larger group containing W, X, Y, and Z, rather than a smaller one, containing A, B, and C—that is to say, in contractualist terms, whether a principle permitting the rescuer to save the smaller group would be one that some individual in the larger group could reasonably reject. This may seem to pose a problem for contractualism.⁶⁴ In the previous case, when we were considering the *permissibility* of saving the larger number, that is to say, considering the reasonableness of the single person's rejecting a principle permitting this, the reasons of the members of the larger group to want to be saved were reasons that the single person had to take into account in assessing the sufficiency of his reason for doing rejecting the principle. As reasons for that person, these reasons therefore added up in the usual way that reasons do. When we are considering the rejectability of a principle permitting a rescuer to save a smaller group rather than a larger one, however, the reasons of the members of the smaller group add up as reasons against rejecting the principle, but the reasons of members of the larger group do not add up in this way, because they are not reasons of any single person.

⁶⁴ I confronted this problem in "Contractualism and Justification" (see pp. 30-31) but did not see how to solve it.

The reason that any member of the larger group has to want to be saved is equivalent to the opposing reason of any single member of the smaller group, and is thus counterbalanced by that reason. But the reasons of other members of the smaller group remain uncounterbalanced, and it might seem that, taking these reasons into account, it would therefore be unreasonable for a single member of the larger group to reject the principle permitting the smaller group to be saved. The member of the larger group cannot cite the reasons of others in the larger group as additional reasons to reject the principle because these are not personal reasons of that individual. To count these as reasons for him to reject the principle would be contrary to the contractualist idea that reasons for rejection must be personal reasons of an individual, not reasons of a group of individuals, or reasons of an impartial contractor.

The core of the individualist restriction—the idea that a reason for rejecting a principle must be a personal reason of some single individual—thus results in what may seem to be a surprising asymmetry in contractualist justification. Reasons against rejecting a principle add up but reasons for rejecting a principle do not.

Still, the contractualist idea of taking into account the reasons of all would seem to mean that the reasons of other members of the larger group must be taken into account in some way. But in what way? The solution to this difficulty is that the reasons of other members of the larger group are relevant not as additional reasons for rejecting the principle but as reasons among those that determine the cost of rejecting it. One effect of rejecting the principle permitting the rescuer to save the smaller group rather than the larger group would be that the members of that smaller group would not be saved. But another effect is that the

other members of the larger group *will* be saved. The cost of rejecting the principle (measured in terms of reasons) needs to take into account effects, and reasons, of both of these kinds.

Since the group including the individual whose reasons for rejection are being assessed is larger than the opposing group, there are enough other individuals in that group for their reasons to want to be saved to counterbalance *all* of the reasons of individuals in the smaller group. This makes the first person's reason to want to be saved a sufficient reason to reject a principle permitting rescue of the smaller group, since all the opposing reasons are counterbalanced.

Even if this argument removes a potential objection to contractualism by showing how it can account for the relevance of the numbers of individuals who can be saved, it may appear to be a weakness that this seemingly obvious conclusion requires what may appear to be a convoluted argument. In addition, it may seem not to capture the intuitively obvious fact that the larger number of individuals on one rock counts in favor of saving them. I believe, on the contrary, that it does capture this intuition, and that it does so in a plausible way.

First, what the argument shows is that in order not to be rejectable, a *principle* must require saving the larger number. So the principle that, according to contractualism, a rescuer is supposed to follow says that the rescuer should take the number of individuals on each rock into account and save the greater number. Second, the argument that I have just given shows that *every* member of the larger group would have sufficient reason to reject a principle that did not require this. Third, my argument that this is the case takes the reasons of members of both groups into account in a plausible way.

It would be implausible to claim that one person's objection to not being saved is made stronger by the presence of others on the same rock.⁶⁵ My explanation accepts this and explains how the presence of others on the rock can nonetheless make a difference to the claim to be saved of each member of the larger group. It does not make their reasons for wanting to be saved stronger, considered in themselves. What it does is to make their individual reasons *sufficient*⁶⁶ reasons for rejecting a principle that would permit them not to be saved, by reducing the cost of their rejecting that principle, measured in the force of opposing reasons.⁶⁷ The answer to the question in the title of Anscombe's paper "Who is wronged?" is thus that each of them would be wronged if the smaller number were saved.⁶⁸ They would each be treated in a way that would be permitted only by principles that each of them, individually, has sufficient reason to reject.

An important limitation of the argument I have just given is that it applies only to cases in which all of the benefits or burdens at stake are the same, such as lives saved, and the reasons for or against them are equivalent. To extend this argument to cases in which these

⁶⁵ Although not stated explicitly, I think that this is one idea behind Anscombe's argument in her paper, "Who Is Wronged?" *The Oxford Review* 5 (1967), pp. 16-17.

⁶⁶ Recall the point made earlier that the sufficiency of a reason in a given context, rather than an independent idea of the strength of a reason, is the fundamental notion.

⁶⁷ The strictly bilateral interpretation screens out these differences in cost, making making the reasonableness of rejecting a principle depend only on the opposing reasons of single individuals on each side. When, as in the present cases, these reasons are the same, this opens the door to several possible views: to Anscombe's view and Taurek's that no one is wronged by either choice; to the "hybrid" view put forward by Munoz-Darde, that since no one is wronged what one ought to do is to produce the most good by saving the larger group; and to the view defended by Otsuka and others (and suggested also by Taurek) that since everyone has the same claim, one should flip a coin, giving everyone an equal chance of being saved. At least within contractualism, the strictly bilateral interpretation is thus the deeper theoretical source of the plausibility of these non-aggregative views.

⁶⁸ As Kumar says, ("Contractualism on the Shoals of Aggregation," p. 146.) As he also says, this explains Frances Kamm's observation that "As the number of individuals in the larger group who outnumber those in the smaller group increases, there is a *greater* wrong done if we do not save the greater number." ("Owing Justifying, and Rejecting," in *Intricate Ethics*. p. 480.) There is a "greater wrong" just in that the number of individuals wronged is greater.

reasons are not the same, such as cases Parfit mentions involving deaths on the one hand and cases of paraplegia on the other, would require some specification of equivalence, in a given context, between these different harms or benefits, such as the number of cases of paraplegia that could counterbalance or be counterbalanced by a single death.

It should be noted that although these equivalences would allow numbers of “lesser” reasons to be part of what determines the cost of rejecting a principle, groups of such reasons would not count as reasons for rejecting a principle. Even if, taken together, these reasons are equivalent to—can counterbalance—a single stronger reason of a single individual, they are not, *taken together*, a personal reason of any single individual. They therefore cannot, taken together, be a reason for rejecting a principle, according to contractualism. (This is the asymmetry I noted earlier.)

It was a mistake in my earlier work not to see that the problem of interpersonal aggregation, and the proper response to cases such as my Transmitter Room example, lies here, in the question of when many “lesser” reasons can counterbalance a “weightier” reason.⁶⁹ And it was a mistake to try instead to deal with such cases by barring interpersonal aggregation altogether, via the strictly bilateral interpretation of reasonable rejection. Establishing equivalences between reasons of different kinds is not a problem only for contractualism. I do not know of any satisfactory general account of when many lesser burdens morally outweigh a smaller number of greater ones.⁷⁰ But aside from the general difficulty of

⁶⁹ In Hampshire, ed., *Public and Private Morality*, reprinted in my *The Difficulty of Tolerance*, pp. 124-150.

⁷⁰ This corresponds to a total utilitarian view’s need for cardinal interpersonal comparisons of happiness or utility. But contractualism needs only a partial ordering of reasons and sets of reasons: that is to say, a set of yes or no answers to questions of the form “Is R a sufficient reason to reject a principle given opposing reasons O₁, O₂ ...On?”

answering such questions there is a problem for contractualism about how the questions themselves are to be understood, and whether answers to them represent a form of moral judgment independent of contractualism itself.

Another mistake I made was failing to see the correct way of responding to the problem posed by cases such as my Transmitter Room example, in which aggregation of very small benefits seems to lead to implausible conclusions.⁷¹ I failed to see this partly because I did not fully appreciate the implications of the shift from taking amounts of utility, or the satisfaction of interests, as the basic elements of justification, as I did in “Rights, Goals, and Fairness,” to taking reasons as basic, as I do in *What We Owe to Each Other* and in more recent work. Although reasons add up in the way I have described they do not do so by forming what Sidgwick called “a mathematical whole,” to which every fact that is a reason for doing something has a weight (a degree of utility) that contributes in the same way to the justification of any decision in which it is at stake.⁷² Rather, reasons add up only if they are relevant to the decision in question.⁷³

⁷¹ *What We Owe to Each Other*, p. 235.

⁷² *Methods of Ethics*, p. 381. The claim that reasons can add up in the way I am suggesting thus does involve what Shelly Kagan called “the additive fallacy” since it rejects the idea that the contribution of one reason in a given situation is independent of other reasons that apply. See Kagan “The Additive Fallacy,” *Ethics* 99 (1988), pp. 5-31, esp. 17-18.

⁷³ The general phenomenon of relevance was first brought to my attention by Frances Kamm’s idea of “irrelevant utilities.” See Kamm, *Morality, Mortality* Vol. 1 (New York: Oxford University Press, 1993), p. 101., For recent discussion of the idea see Victor Tadros, “Localized Restricted Aggregation” in *Oxford Studies in Political Philosophy*, ed. David Sobel, Peter Vallentyne, and Steven Wall (Oxford: Oxford University Press, 2019), 5:171–204, Erik Khang, “Individualist Theories and Interpersonal Aggregation,” *Ethics* 134 (2024) pp. 479-511, and the many other works cited therein. As Zhang points out (p. 485), inspired by Kamm I recognized the significance of relevance in *What We Owe to Each Other* (pp. 239-240), But I failed until much later to see how this idea might be employed to provide an alternative account of aggregation within a contractualist framework. See “Contractualism and Justification, p. 33.

I do not have a general account of when a consideration is relevant to a certain decision, but it is at least clear that relevance does not depend merely on the relative seriousness of the harms or benefits in question. In the case I described earlier, for example, the significance of the reasons involved depend on the family relations between the parties. And something similar is true in the Transmitter Room example itself. If we think of the reasons for allowing the transmitter to remain on simply as reasons that the viewers have to want their enjoyment of the game not be interrupted, it seems that if these reasons are not relevant this must be due simply to their lack of seriousness relative to the pain of the technician.

But there is also the fact that the television company has led people to expect that they can watch the match uninterrupted, and viewers may have relied on this in deciding how to plan their day or evening. If it is a cable company, viewers may even have paid something for this assurance. So the pleasure that individual fans will get from watching the game without interruption (and the frustration they will feel if it is interrupted) are not the only reasons against rejecting a principle that would permit the transmitter to be left on. There are also reasons to object to the company's failure to fulfill expectations that it has created or even taken money for. When the case is seen in this way, the most basic question is thus like the question, discussed above, of the bearing of third-party interests on the obligation to keep a promise. It is the question of whether relieving the pain that the technician is suffering is sufficient reason to disappoint the assurance that viewers have been given. I think the answer is

that it clearly is sufficient. The obligations to viewers that the company would otherwise have are thus undermined, and there are therefore no reasons of this kind to add up.⁷⁴

This analysis explains why the reasons of viewers to watch the game without interruption, and the number of viewers who will be watching at a given time, would be relevant to the company's decision whether to shut down the transmitter for routine maintenance. The need for maintenance does not extinguish the company's obligations to individual viewers. Rather, the reasons to fulfill these obligations require it to schedule system maintenance so as to minimize interruption of the service they have contracted to provide.

More generally, this indicates that the question of whether interests of a certain kind are relevant to a certain decision is broadly speaking a moral question. What else could it be if the relevance of a reason depends not simply on the magnitude of that benefit or burden but also on the decision involved and the relation between the parties? This is not a problem for contractualism—not a problem of circularity—as long as the questions of relevance are ones that contractualism itself can answer, as it does in the Transmitter Room case understood in the way I have just considered, but may not do more generally.

It may seem surprising that strengthening the reasons of the viewers by elevating them to the status of (potential) claims of obligation also makes them more vulnerable to being disabled. But this should not be a surprise. The analysis just given involves two related steps. The first is the recognition that whether I owe it to someone to take certain reasons of hers into account in a certain decision is a moral question, the answer to which depends not only on the

⁷⁴ This explains why, as Tadros puts it, these reasons are disabled rather than outweighed by the reason to relieve the technician's pain. "Localized Restricted Aggregation," p. 175.

seriousness of the reason but also on the relation we stand to one another and the decision that is at issue.⁷⁵ The second step, which follows from this, is that this is a question about the kind of consideration that individuals owe to one another, a positive answer to which is presupposed by any adding up of reasons. It is therefore not surprising that if the assurance-based reason of the fans in the version of the Transmitter Room example as I have just described it does not make it unreasonable to reject a principle permitting the broadcast to continue, the same is true of the purely enjoyment-based reason in a variant of the case in which no assurance has been given and those whose enjoyment is at issue are unrelated strangers. Is the momentary enjoyment of an unrelated stranger, considered simply in itself, something that we must take into account in deciding whether it is reasonable to reject a principle that would permit us to fail to alleviate another person's great pain? The moral nature of this idea of relevance weakens the explanatory power of a reductive view. But on reflection it seems to me clearly the best way of understanding such cases.

Contractualism and Utilitarianism, and Intuitionism

Making numbers count in the way I have just described narrows the difference between contractualism and utilitarianism (and consequentialism more generally.) But important differences remain. Contractualism is like utilitarianism in being a reductive theory that makes the rightness and wrongness of actions depend on how the lives of individuals would be affected if certain actions were permitted, treating the lives of all individuals as equally

⁷⁵ The point I am making here may be an instance of the more general point Christine Korsgaard has in mind in saying that "The subject matter of morality is not what we should bring about but how we should relate to one another." "The Reasons We Can Share: An Attack on the Distinction between Agent-relative and Agent-neutral Values," in *Creating the Kingdom of Ends*, p. 275. See also Ross, *The Right and the Good*, p. 19, on the fact that our relation with others is not always merely that of potential benefactor and potential beneficiary.

Important. But it differs from utilitarianism in the way it takes these effects into account.

According to contractualism what matters, ultimately, are individuals' reasons for favoring or rejecting certain principles, based on the ways that living under these principles would affect their lives. According to utilitarianism what matters are these effects themselves and, ultimately, the total amount of happiness or welfare that they add up to.

The central role of reasons in contractualism leads to a difference in what it implies about the scope of morality. The form of morality that contractualism describes applies only to rational beings, because only rational beings can have reason to accept or reject moral standards. This narrowness is often seen as a problem for contractualism, and a reason for preferring utilitarianism. But this overlooks the fact that we stand in a different relation with other rational beings than with beings that are not rational.⁷⁶ It is only with other rational beings that we have, ideally, the possibility of living with each other in accord with principles that are mutually justifiable. This is the morality that contractualism aims to describe.

With non-rational beings we do not have this possibility. But in order to explain how our treatment of other animals can nonetheless be open to moral objection we do not need to strain, implausibly, to bring them within this relationship. The alternative is to recognize that what happens to them can give us reasons that are independent of the kind of rightness and wrongness that contractualism describes. As Bentham said, "The question is not Can they Reason? Nor can they talk, But Can they suffer?"⁷⁷ Even if some sentient beings are not rational, there are things that are good or bad for them which they seek to attain or avoid.

⁷⁶ As Christine Korsgaard also points out in *Fellow Creatures Our Obligations to Other Animals*, section 8.7.

⁷⁷ *Introduction to the Principles of Morals and Legislation*, p. 283. Korsgaard argues that it is the larger question of whether they have a good. *Fellow Creatures*, section 8.5.

What happens to them therefore can give *us* reasons that it is morally objectionable to ignore. I do not have an account of this wider dimension of morality, although I do not believe that utilitarianism is the correct view.⁷⁸ My present point is just that the fact that contractualism applies only to rational beings should not be seen as an objection. It reflects a real difference.

The basic role of reasons in contractualism also leads to substantive differences from utilitarianism in the considerations that the two views take into account, and in the way in which they do this. One difference is that contractualism takes reasons into account only if they are relevant in a given context. Utilitarianism might in principle distinguish between increases in happiness that are relevant and those that are not. But the aim of maximizing happiness does not supply any obvious rationale for doing this.

The most fundamental difference between contractualism and utilitarianism, however, is that even when they support the same conclusions about right and wrong their basic reasons for doing so are different. Contractualism starts from an idea of justifiability to each individual. It takes into account individuals' reasons for objecting to principles that would permit their lives to be affected in certain ways, and assesses the sufficiency of these reasons given the opposing reasons of others, treating the similar reasons of different individuals as having the same importance.

Utilitarianism, at least insofar as it is a teleological view⁷⁹ (a version of consequentialism), starts from the idea of promoting the greatest amount of good. It takes into account the benefit that an action or principle would provide for one person and the cost that

⁷⁸ As Korsgaard also argues in *Fellow Creatures*, pp. 156-166.

⁷⁹ On teleological and non-teleological forms of utilitarianism, see Rawls, *A Theory of Justice*, sections 27, 30, and Scheffler, "Rawls and Utilitarianism, in Samuel Freeman, ed., *The Cambridge Companion to Rawls*, pp. 426-459.

this would involve for others in order to determine the net amount of good that would result from accepting that principle. The fact that the reasons of some individuals for objecting to effects on their lives are counterbalanced by reasons of others is an irrelevant side effect.⁸⁰ The primary point is producing the greatest amount of happiness, wherever it occurs.⁸¹ As Rawls says, the teleological version of utilitarianism replaces the idea of valuing the good of different individuals impartially with the idea of the greatest happiness as an *impersonal* value, that is to say, something that is good, as Sidgwick put it, “from the point of view of the universe,” or from the point of view of an impartial spectator who feels the pleasures and pains of all individuals as his own.⁸²

I have described this contrast as one between contractualism and utilitarianism as accounts of the content of morality. But a central aim of such accounts is to provide the basis for answering the question of acceptance, and here the contrast is equally clear. In a passage that is frequently criticized as a non-sequitur, Mill wrote that “...each person’s happiness is a

⁸⁰ This difference is illustrated by the asymmetry pointed out at the end of the preceding section. According to contractualism, even when a number of lesser reasons, taken together, count the same as a single stronger reason of some individual in assessing the cost of rejecting a principle, only that single reason can be a reason for rejecting a principle. This is because the lesser reasons, taken together, are not personal reason of any single individual. By contrast, on a teleological view these reasons or, rather, the bits of utility corresponding to them, do contribute to the impersonal value of the state of affairs that results if the principle is rejected.

⁸¹ It might be thought that a teleological consequentialist answer to the question of acceptance derives support from a general rational requirement to maximize the good, as suggested by Samuel Scheffler (“Agent-Centered Restrictions, Rationality, and the Virtues,” in *Consequentialism and Its Critics*, p. 151.) I myself do not believe that there is such a requirement. (See the final section of my “Good of a Kind and Good for Understood in Terms of Reasons in Andrei Marmor, Kimberley Brownlee, and David Enoch, eds., *Engaging Raz: Themes in Normative Philosophy* (Oxford: Oxford University Press, 2025), pp. 3-21.) It seems to me rather that the considerable appeal of consequentialism derives not from a supposed rational requirement but rather from the consideration of particular cases in which an agent must decide whether to achieve some obviously good result, such as saving a life, at the expense of violating some deontological constraint such as an individual’s right. The reason for doing the former is obvious, and does not need the support of any general rational requirement. But it is much less clear what the reason is for not violating this constraint. The contractualist answer to the question of acceptance tries to respond to this problem.

⁸² See Sidgwick *The Methods of Ethics*, p. 382 and Rawls, *A Theory of Justice*, p. 190:

good to that person, and the general happiness, therefore, a good to the aggregate of all persons.”⁸³ Several things seem to be wrong about this argument. First, it is not clear that “the aggregate of all persons” even has a good. Second, Mill felt the need to make the step marked by this ‘therefore’ in order to provide what he called a “proof” of utilitarianism—to give an answer to what I am calling the question of acceptance. So what he needs to conclude in saying that the “the general happiness is therefore a good to the aggregate of all persons” is that *each* individual in this aggregate has reason to accept morality if it is based on utilitarianism.

Mill’s argument as he states it seems to be a non-sequitur because the general happiness is not a good to each individual in the same sense in which an individual’s own happiness is a good to that person. It is not something that is *good for* each individual person. But there is more than one way to understand the idea of “a person’s good.” On the one hand it can mean the things that are good for that person in the sense of benefitting the person by affecting the person’s life in a way that the person has reason to want. But a person’s good can also be understood more broadly, as encompassing all the things that the person has reason to want—things that are good from that person’s point of view. An end to war or famine in some distant part of the world can be something that a person has reason to care about, and thus part of that person’s good in this broader sense, even though it is not good for that person in the narrower sense—not something that benefits him or her.

So one way of understanding Mill’s remark, “...each person’s happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of all persons” is as the substantive normative claim that because each person’s happiness is good for that person, the

⁸³ *Utilitarianism*, in Ryan, p. 308.

general happiness (the sum of the happiness of all these individuals) is something that each person has strong reason to care about. This is a substantive normative claim that may or may not be correct, but it need not be read as a non-sequitur.

The contractualist answer to the question of acceptance is that we all have good reason to want to live with others under principles that can be justified taking the personal reasons of all into account in the way that it describes. This answer is fairly understood as a claim that living with others in this way is “a good to” every individual —something that every individual has reason to want. And Mill might, understood charitably, be understood as claiming the same for living with others under principles that lead to the greatest total happiness, namely that this is a way, as he put it, of “being in unity with our fellow creatures.”⁸⁴ It is not obvious, however, that maximizing the sum of happiness regardless of how it is distributed is a better realization of the idea of “being in unity with our fellow creatures” than living under principles that can be justified to all taking their personal reasons into account. Understood in the latter way Mill’s phrase would suggest that what is basic even for him is justifiability to others rather than maximizing impersonal goodness.

Mill interpretation aside, however, the idea that moral requirements depend on the sufficiency of individuals’ reasons for rejecting them given the opposing reasons of others seems to me a more informative account of the content of morality and a more appealing answer to the question of acceptance than a teleological view. This contractualist account can explain why it seems to many people that morality requires more of us than is commonly recognized, an implication that might have seemed to favor to a utilitarian or consequentialist

⁸⁴ *Utilitarianism*, in Ryan p. 299.

view. In this respect, the difficulty that contractualism has in limiting its tendency to be maximally demanding appears as a strength rather than a potential objection. In addition, the intuitions with which many arguments for a more demanding view begin, such as the pull we feel when we consider how much more we could do to relieve the plight of people in famine-stricken lands, seem to me best understood not as the thought that their suffering is a bad thing but rather that we owe it to *them* to do more to aid them when we could do so at so little cost.⁸⁵

As I have argued, contractualism can explain why the numbers count in some cases without assuming that maximizing happiness, or maximizing the good more generally, is the basis of morality. The contractualist explanation of both of these things—the pull toward a more demanding morality and the relevance of numbers in some cases—lies in the interactive character of contractualist justification and in the need to take into account the reasons of all affected, on both sides of a question.

This interactive character is thus central to the appeal of contractualism, and central to what distinguishes it both from a maximizing doctrine such as utilitarianism and from other broadly contractarian views. But it is also the source of significant difficulties since the reductive and interactive justification that contractualism involves pulls us toward very demanding conclusions. As I have said, this is not in itself an objection: morality may be very demanding. But it raises the question of whether contractualism provides a clear way of thinking about these demands and their limits. The reductive and interactive justification form of contractualist justification appears to have considerable clarity and explanatory power when

⁸⁵ See *What We Owe to Each Other*, p. 152.

the judgments of sufficiency on which it is based concern specific quantifiable ways of being harmed or benefitted. But, as I have argued, the relevant questions of sufficiency are much more complicated than this, and judgments about them are not so clear. Indeed, they may seem much less clear than the moral conclusions that they are supposed to explain.

These problems might be claimed as a strength of contractualism. Its interactive character forces us to make explicit the reasons that are supposed to support commonly held judgments and to defend the importance that must be attributed to these reasons in order for these judgments to be supported. If the balance of these reasons is less clear than we would like, the fault does not lie with contractualism. If, on the other hand, this lack of clarity is taken to undermine contractualism's claim to be a convincing account of the content of morality, the question then is what alternative account there might be, and in particular what answer to the question of acceptance this account can offer, assuming that that question requires an answer.