

STEPHEN DARWALL

## Contractualism, Root and Branch: A Review Essay

A theme running through T. M. Scanlon's essays in *The Difficulty of Tolerance* is an abiding concern to understand and adequately justify *rights*, including those involved in freedom of expression, toleration, punishment, promise, and contract. It is interesting and instructive to see the evolution in Scanlon's thinking as he returns to this theme repeatedly in various guises over a span of thirty years. In "A Theory of Freedom of Expression" (originally published in 1973), for example, Scanlon argues that governmental bans on expression that might result in false or dangerous belief conflict directly with autonomy of the individual, who must be presumed "sovereign in deciding what to believe" (p. 15). By the time of "Rights, Goals, and Fairness" (1978), however, Scanlon's thinking has moved more in the direction of a consequentialist or "instrumental" account, where rights are justified by their effects in protecting "urgent" or "important" interests (pp. 2, 31). "Preference and Urgency" (1975) defends what Scanlon then thought such an account needs: an "objective criterion" of well-being as against "subjective" criteria such as pleasure or preference-satisfaction.

By the time "Contractualism and Utilitarianism" appeared in 1982, however, Scanlon had become doubtful that the concept of well-being can bear the weight that consequentialist moral theories require, and he had seen his way toward another approach that, like objective forms of welfarist consequentialism, assesses candidate rights and principles by their consequences for "important" or "urgent" interests, but in a fundamentally different way. For contractualism, the basic issue is not the value of outcomes conceived in terms of welfare or impersonal value, but how affected interests give individuals grounds for reasonably

This review essay considers T. M. Scanlon, *The Difficulty of Tolerance* (Cambridge: Cambridge University Press, 2003).

accepting or rejecting principles “as a basis for informed, unforced general agreement” (p. 132). Considerations of welfare have moral relevance in this scheme, not intrinsically, as in consequentialism, but “because an individual could reasonably reject a form of argument that gave his well-being no weight” (p. 140). The fundamental notion is of a ground or reason for accepting or rejecting a principle “as a basis for unforced general agreement.”

Yet how is the idea of reasonable agreement to be understood? What makes something a ground of reasonable rejection (or acceptance)? Consider a philosophical utilitarian of a Sidgwickian stripe, who takes it to be self-evidently rational to aim at the greatest pleasure of all.<sup>1</sup> Why can she not hold that it is unreasonable for an individual to reject the principle of utility on the grounds that it may require her to bear significant burdens others do not if this will bring overall net gain? What Scanlon says in “Contractualism” is not wholly satisfying on this score, namely, that contractualism involves a “qualified skepticism” that denies the existence of “moral properties which have justificatory force independent of their recognition in any ideal agreement” (p. 140). The claim that the greatest overall welfare or pleasure is an intrinsically rational or reasonable aim faces, Scanlon argues, the problem that Mackie famously raised: how “intrinsic to-be-doneness and not-to-be-doneness” could be properties “in the world” (p. 140). It is not, however, clear how antirealist concerns are supposed to help at this point, since it is clear neither how they rule out a utilitarian-friendly conception of reasonable agreement nor rule in a definite rival to it. Could the Sidgwickian claim not be made consistently with holding an expressivist or quasi-realist meta-ethics? We shall return to this question presently.

In “Promises and Contracts” (2001), Scanlon brings the justification of rights directly into the contractualist framework. The question of whether a given moral right exists turns on whether a principle licensing its enforcement can or cannot not be rejected reasonably. Scanlon’s concern there is a contractualist justification of promissory and contractual rights, but grasping how he argues in this case can help us see what picture of reasonable agreement he must be assuming in the background. Moreover, I think it can be shown that the seeds and sometimes the substance of the contractualist approach is implicit also in Scanlon’s

1. Assume that he denies the other prong of Sidgwick’s “dualism of practical reason,” that is, that it is self-evidently rational to aim at one’s own greatest pleasure also.

earlier treatments of the justification of rights, including those that he characterizes as “instrumentalist.” My plan, therefore, is to consider the essays in *The Difficulty of Tolerance* from this perspective, hoping to shed light both on how to understand the idea of reasonable agreement within contractualism and on contractualism’s distinctive approach to the justification of rights. Even in Scanlon’s early essays, we can find what I will argue is contractualism’s root idea, namely, our equal standing to claim justifications *from* (and to give them *to*) one another. It is our *equal second-personal authority*, as I will call it, our standing in the relation of mutual accountability to one another, that ultimately rules the Sidgwickian claim out of the contractual framework.

In “A Theory of Freedom of Expression,” Scanlon defends what he calls the “Millian Principle”:

There are certain harms which, although they would not occur but for certain acts of expression, nonetheless cannot be taken as part of a justification for legal restrictions on these acts. These harms are: (a) harms to certain individuals which consist in their coming to have false beliefs as a result of those acts of expression; (b) harmful consequences of acts performed as a result of those acts of expression, where the connection between the acts of expression and the subsequent harmful acts consists merely in the fact that the act of expression led the agents to believe (or increased their tendency to believe) these acts to be worth performing (p. 14).

In sum, the state may not legitimately restrict expression on the grounds that it would give rise to false or harmful beliefs. Citizens’ rights to freedom of expression consists in this, that any attempt to restrict this freedom on these grounds would be illegitimate, a violation of the right.

What, however, grounds the Millian Principle and therefore this right? Why would any attempt to restrict freedom of expression on these grounds be illegitimate? “Theory”’s answers appear contractualist *avant la lettre*. Scanlon argues that “the powers of the state are limited to those that citizens could recognize while still regarding themselves as equal, autonomous rational agents” (p. 15). And again: “the harm of coming to have false beliefs is not one that an autonomous man could allow the state to protect him against through restrictions on expression” (p. 17). The underlying thought seems to be a test of rational or reasonable acceptance or rejection of some kind. Since an autonomous individual

could not reasonably accept the state's having the authority to violate the Millian Principle, he could not reasonably reject a principle that requires states not to violate it. It would therefore be wrong for the state to do so.

Why, however, could an autonomous person not reasonably accept (or decline to reject) violations of the Millian Principle when that would promote greater utility overall? Why, again, could an autonomous person not be a Sidgwickian utilitarian? In "Theory," Scanlon's argument proceeds directly from his conception of autonomy. To conceive of himself as autonomous in the relevant sense, "a person must see himself as sovereign in deciding what to believe" (p. 15). It therefore follows directly, in effect, by definition, that autonomous individuals could not "regard themselves as being under an obligation to believe the decrees of the state to be correct" or "concede to the state the right to have its decrees obeyed without deliberation" (p. 17). To accept either would be to deny their autonomy, conceived as their sovereignty over what they believe, even if either would promote utility. The Millian Principle, Scanlon concludes, is just "a refinement of these limitations" (p. 17). To concede to the state the right to limit expression on the grounds that it would give rise to false or harmful belief would be to renounce sovereignty over one's beliefs and hence autonomy, and this autonomous individuals could not reasonably do.

In later essays, Scanlon gives up the Millian Principle as well as this way of arguing for it and for grounding rights more generally. However, there are two different aspects of "Theory"'s argument for the Millian Principle that should be distinguished. One is the general (contractualist) idea that principles of right must be acceptable (or not reasonably rejectable) by autonomous individuals (together with the conception of autonomy as involving authority of some kind over oneself). The other is the more specific idea that accepting state authority of the sort the Millian Principle rules out would be inconsistent with autonomy *by definition* and, therefore, that autonomous individuals could not consistently accept it, regardless of the consequences of their doing so for their important interests (*including*, we should note, interests they might have as autonomous individuals). This distinction will be significant when we come to consider why Scanlon ultimately gives up "Theory"'s justification of the Millian Principle. I will suggest that it is possible for him to do so while still properly regarding his revised theory as contractualist rather than consequentialist.

Before we proceed, I would like to note an aspect of “Theory’s” conception of autonomy that is linked to some insightful remarks that Scanlon makes there about responsibility. In the course of discussing why it is unjustified for the state to restrict expression whose object is to persuade people to undertake harmful acts, although it might not be to restrict expression that has the same probability of causing harm but not via such persuasion (say, by providing relevant information to people who are already persuaded to do harm), Scanlon notes that this distinction is supported by our normal views of legal (and we might add, moral) responsibility. The difference between an accomplice to murder and someone who tries to convince someone to murder but is not otherwise an accessory is that the causal contribution made by the latter is, for purposes of assigning responsibility, “superseded by the agent’s own judgment” (p. 13). Unless the circumstances involve diminished capacity of some kind, we generally hold people responsible for acting on whatever reasons they end up acting on. It is no defense to say, “I would not have killed *A*, if *B* had not suggested the idea to me.” In this sense, a kind of sovereignty “in deciding what to believe and in weighing competing reasons for action” (p. 15) seems to be written into the very idea of holding people responsible for their actions. To hold people responsible just *is* to treat them as though they were sovereign over their own wills and deliberative beliefs in a sense.

It is useful to consider “Rights, Goals, and Fairness,” “Due Process,” and “Preference and Urgency” as something of a group. “Rights” has the latest publication date of the three, 1978, but Scanlon tells us that he has placed the essays in the order they were written, and “Rights” appears before “Process” (1977) and “Preference” (1975). Scanlon says he set out in 1975 to write a book on freedom of expression but that he abandoned the project, partly because he was becoming dissatisfied with “Theory”’s approach to justifying rights. (He also tells us that he was being pressed in the direction of consequentialist theories by Peter Railton and Samuel Scheffler in his Princeton seminars.) “Rights” was the upshot. As he views it now, Scanlon is no longer willing to call its account “consequentialist.” In the article itself, however, he calls it consequentialist “in a broad sense” (p. 34). At any rate, it seems clear that, although “Rights” is cast in terms that are friendly to consequentialism, the theory’s core idea, that rights are justified by their role in protecting “important” or “urgent” interests, can be interpreted within either a consequentialist or a

contractualist framework. The main point Scanlon seems concerned to make, departing from the line of argument in “Theory,” is that rights must be justified by considering the consequences of their recognition, and not as following by definition from a conception of the right holders as autonomous.

An adequate justification of a right, Scanlon now maintains, must include the following:

- (i) An empirical claim about how individuals would behave or how institutions would work in the absence of this particular assignment of rights. . . .
- (ii) A claim that this result would be unacceptable. This claim will be based on valuation of consequences. . . . taking account also considerations of fairness and equality.
- (iii) A further empirical claim about how the envisaged assignment of rights will produce a different outcome (p. 35).

Sensitivity to the consequences of protecting or recognizing a given right is not enough to make a way of justifying the right consequentialist, of course; a Rawlsian form of contractualism proceeds in this way also. What makes an account consequentialist is that it is fundamentally based exclusively on the value of outcomes. Still, even though an “unacceptability” standard (as in (ii)) need not depend on a relative assessment of overall outcome-value, this does seem to be the way that Scanlon is mainly thinking of it in “Rights.” So when, for example, he considers how distributive considerations might enter into justifying rights, the approach he favors is a “distribution-sensitive” valuation of outcomes of the kind featured in the consequentialist theory that Scheffler discusses in *The Rejection of Consequentialism*.<sup>2</sup> The idea here is that fairness and equality are “independently valuable states of affairs,” so the fact that a candidate right would promote such states counts in its favor.

The kind of “two-tier” theory that Scanlon puts forward in “Rights” clearly differs from direct or “act” forms of consequentialism, since any theory that “takes rights seriously” must place “limits on consequentialist reasoning at the level of casuistry” or individual cases (pp. 26–27). Such a theory may still be consequentialist, however. Like Mill’s account

2. Samuel Scheffler, *The Rejection of Consequentialism* (Oxford: Clarendon Press, 1982).

in *Utilitarianism*, it can understand rights as useful social devices. “To have a right,” according to Mill, “is to have something which society ought to defend me in the possession of.” If anyone asks, “Why it ought?,” Mill adds, “I can give him no other reason than general utility.”<sup>3</sup>

Scanlon evidently feels some pressure to distinguish the theory in “Rights” from rule utilitarianism since he believes that threatens either to collapse into act utilitarianism, which is inconsistent with respecting rights, at one extreme, or to lack sufficient “critical force” because it ratifies whatever rules the society actually accepts, at the other (p. 27). However, this may be a function of the varieties of rule utilitarianism (or rule consequentialism) that Scanlon considers. An injunction to act on “rules general *conformity* to which would have the best consequences,” does indeed reduce to act consequentialism, and a requirement to act on the rules actually accepted in one’s society just counsels supporting the status quo (p. 27, emphasis added). Surely the most plausible versions of rule consequentialism, however, are those, like Richard Brandt’s and Brad Hooker’s, that counsel action on rules the general *acceptance* of which has (or expectably will have) the best consequences.<sup>4</sup> Scanlon mentions Brandt’s version in a footnote. He remarks that it is like “Rights”’s theory in its form, but that it differs from it in being tied to a subjective theory of outcome-value and in being a maximizing account (p. 36, n. 7).

According to (i) through (iii), protecting a right must avoid “unacceptable” consequences, not bring about the best. As Scanlon himself points out, however, to be plausible (iii) should also require that unacceptable consequences be avoided at low (the least?) cost (p. 36). This seems to push back in the direction of a maximizing theory, albeit one with an “objective” notion of the value of outcomes that can include considerations of distributive justice.

On balance, then, “Rights”’s “two-tier theory” seems largely cast within a consequentialist tradition that derives from Mill. One thing, at least, is clear. Scanlon has definitely moved away from the kind of

3. John Stuart Mill, *Utilitarianism*, chap. V, para. 25.

4. Richard Brandt, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979); Brad Hooker, *Ideal Code, Real World* (Oxford: Oxford University Press, 2000). Hooker’s version did not, of course, exist at the time of “Rights.” Brandt’s version did, however, in “Some Merits of One Form of Rule Utilitarianism,” in *Mill: Utilitarianism, with Critical Essays*, ed. Samuel Gorovitz (Indianapolis, Ind.: Hackett, 1971).

contractualist account he advanced in “Theory,” according to which a candidate assignment of rights can be assessed in advance of considering consequences, simply on the grounds that it conflicts with or is required by the very idea of autonomous individuals. Nevertheless, even if a consequentialist interpretation is the most natural reading of “Rights” considered in its own terms, when we look closely at the notion of “importance” or “urgency” of interests, especially as it functions in “Preference and Urgency,” we may be able to see, not only why Scanlon can justly now say that his account in “Rights” “fits best within [a] contractualist moral theory,” but also that the seeds of this interpretation were there from the beginning (p. 3).

One thing that can throw us off is that Scanlon formally puts his claims in “Preference” in terms of the welfare consequentialist (and “philosophical utilitarian”)’s favorite notion of well-being. “Preference”’s official claim is that well-being cannot be accounted in subjective terms such as pleasure and preference satisfaction. So far, this would seem to amount simply to a plea for an objective, rather than a subjective, form of consequentialism. In the Introduction, Scanlon says that he later came to think that “the task of giving an overall account of well-being and that of explaining the basis of morally significant interests are less closely related than [he] at first supposed” (p. 5). If we look at “Preference” closely, however, we can see that Scanlon’s real task there was less to provide an (objectivist) account of well-being for its own sake (or to fit into a consequentialist theory of rights) than to say what features “criteria of well-being should have *if they are to play the role commonly assigned to them in moral argument*” (p. 70, emphasis added). It seems clear from the balance of “Preference” that the requisite role is to support “moral claims that some interests should be favored at the expense of others in the design of distributive institutions or in the allocation of other rights and prerogatives” (p. 73). Furthermore, Scanlon notes that it is a strike against subjective criteria that they “leave us open to being ‘held up’” by those with expensive tastes or who attach “inordinate importance” to minor concerns (p. 78). It seems clear enough, then, that what is at issue there is not well-being in the sense of what we appropriately want for someone out of benevolent concern,<sup>5</sup> or what might be

5. I argue that this is how we should understand the idea of well-being or welfare in *Welfare and Rational Care* (Princeton, N.J.: Princeton University Press, 2002).



good “from the point of view of the universe,” but what interests we should take *as warranting claims or demands* that we can legitimately make of one another.

Scanlon discusses Rawls’s thesis that we presume persons to have a degree of control over their preferences and remarks that one argument against subjective criteria might therefore be their manipulability in the service of expensive tastes.<sup>6</sup> He notes that this might warrant taking account of a taste’s actual pedigree in determining its importance, but then remarks that this just seems irrelevant in “determining the strength of the person’s claim on others for aid in satisfaction of the interest” (p. 79). For present purposes, it does not matter whether Scanlon is right about that. My point is simply what he takes to be at issue, namely, what criteria of well-being would have to be if the fact that something is necessary for, or would promote, someone’s well-being *can ground a claim against others*. In other words, the “role” that Scanlon takes well-being “to play . . . in moral argument” is to ground claims and demands we can legitimately make of one another. In the Introduction, Scanlon extends his remark that “Rights” can be understood as contractualist by saying that “defensible institutions must promote the well-being [important or urgent interests] of their citizens in certain ways because this is something citizens can reasonably demand” (p. 3). We can now see, therefore, that even in “Preference” Scanlon was concerned with what criteria of well-being would have to be for well-being to be intrinsically relevant to what we can reasonably claim from one another. In “Preference,” to be an “important” or “urgent” interest is, it seems, to be one that warrants a weighty claim.

Presently, we shall consider how the idea of legitimate claim can help us to understand the kind of reasonable agreement Scanlon must have in mind in “Contractualism and Utilitarianism” (as well as Scanlon’s reasons for rejecting the specific form of contractualist justification of rights that was implicit in “Theory”). First, however, it is instructive to note the role that legitimate demand plays in “Due Process.” The very notion of a right of due process itself involves, of course, the notion that those who are subject to authority have a right “to demand justification

6. Surely another is that, to the extent that expensive tastes are themselves the function of the distributive status quo, they can hardly be taken at face value in determining what distributive arrangements we *should* have.

for its uses" from those to whose authority they are subject (p. 46). To be sure, the underlying justification of that right might be thought to be consequentialist. However, Scanlon also says that his account is "grounded in a conception of the moral requirements of legitimacy for social institutions" and that due process is a "conditio[n] of moral acceptability for institutions that give some people power to control or intervene in the lives of others" (pp. 42, 43). I take this to suggest that morally *legitimate* institutions are not simply whichever legitimacy-defining institutions have the most instrumental value, but that the "power to control or intervene in the lives of others" is what sets the problem of legitimacy in the first place. The point is not just that especially important gains or losses, in terms either of well-being or of impersonal value, are in the balance, but that a justification of a special kind is called for when one person makes a claim or demand on the will of another, namely, a justification that could be given *to* and be found acceptable *by* that person. I will return to this theme shortly.

In "Freedom of Expression and Categories of Expression," Scanlon explains why he has come to reject the argument he gave for the Millian Principle in "Theory." He makes a distinction between appealing to the value of autonomy as an *interest* in "the ability to exercise independent rational judgment," and treating autonomy as a direct "constraint on justifications of authority" (p. 97). He rejects the latter since it "prevents us from even asking" whether urgent interests, including the interest in autonomy itself, might sometimes be better served by accepting violations of the Millian Principle (p. 98).<sup>7</sup> So far as I can see, however, we will be thus prevented only if we take autonomy to rule out such violations by definition. As I pointed out above, however, a justification of rights can avoid being insensitive to consequences by definition and still be a form of contractualism that assesses candidate rights and principles by whether persons can reasonably accept, or not reasonably reject, them in light of their interests as autonomous individuals (including their interest in having the ability "to exercise independent rational judgment").

It seems evident from the articles in the first half of *The Difficulty of Tolerance*, then, that the seeds of the distinctive moral theory that Scanlon sketches in "Contractualism and Utilitarianism" and develops

7. "Theory" permitted this only in cases of incapacity to rational judgment (p. 20).

in *What We Owe to Each Other* were there from the beginning. But what of “Contractualism” itself? What do we learn about its key claims by considering these other essays? Quite a bit in my view. In particular, we learn something about the character of the agreement that contractualism must involve along with its underlying motivation.

“Contractualism” describes three central elements of the theory: (i) an account “of the nature of moral wrongness,” which provides “a characterization of the kind of property which moral wrongness is” (p. 132); (ii) an associated account of “what morality is about” (p. 150); and (iii) an account of the “source of motivation” for avoiding moral wrong (p. 138). These are, respectively: (i) that “an act is wrong if its performance under the circumstances would be disallowed by any system of rules for the general regulation of behavior which no one could reasonably reject as a basis for informed, unforced general agreement” (p. 132); (ii) that morality is about “the idea of [such a] general agreement” (p. 150); and (iii) that the fundamental moral motive is “the desire to be able to justify oneself to others on grounds they could not reasonably reject” (p. 138). These contrast with the three central elements of contractualism’s foil: “philosophical utilitarianism.” For the philosophical utilitarian, the “fundamental moral facts are facts about individual well-being” (p. 120), so, (i) the rightness of acts is reckoned in outcome-value in welfare terms, whether directly (act utilitarianism) or indirectly (e.g., rule utilitarianism); (ii) morality is fundamentally about promoting and protecting well-being, disinterestedly considered; and (iii) the fundamental moral motive is universal or disinterested benevolence, a desire for greatest overall well-being.<sup>8</sup>

How exactly are we to understand “agreement” in this context? How can *agreement* be what morality is fundamentally about? If we stress the word too much, we are apt to be misled into identifying morality with what Kohlberg called “interpersonal concordance” in the “conventional” level (Level II) of his theory of moral development.<sup>9</sup> Scanlon is clear, however, that the basic contractualist motive is not to be in actual agreement with others (p. 138). What Mill called “the desire to be in unity with

8. For some, doubts that benevolence or concern for someone for her sake can be reduced to a desire for her welfare; see my *Welfare and Rational Care* (Princeton, N.J.: Princeton University Press, 2002).

9. Lawrence Kohlberg, “The Claim to Moral Adequacy of a Highest Stage of Moral Development,” *The Journal of Philosophy* 70 (1973): 630–46, at p. 631.

our fellow creatures”<sup>10</sup> may have some role to play in morality, but it is not what Scanlon is pointing to. Neither, however, can the fundamental contractualist motive be a desire simply to avoid *justified* disagreement. This might amount to no more than the desire to do (feel, believe, and so forth) as we should or as there is good reason to. Even moral forms of this concern, say the desire never to act in ways that are justly criticizable morally, need have nothing essentially contractualist about them.

The fundamental contractualist notion is rather of a kind of agreement that essentially involves justification *to* one another. What is it, however, to justify oneself *to* someone? It is not, it is important to see, simply to present a justification in someone’s presence, or even only to exhibit to someone good reasons for something one did. Suppose you are wondering where to spend your vacation and you ask me where I went and whether I would do it over again. I tell you about my experience camping at Kookamanga State Park and why I was pleased with my decision to go there. I have certainly given you reasons, and in that sense a justification, for what I did. Yet have I justified my choice *to* you? It seems ludicrous to suppose that I have. For that to be true, either you or I would have had to have taken you to have had some *claim* to a justification *from* me. Imagine saying, “Would you please justify your vacation choice to me?” or even something more polite that implies a request for such a justification. These seem clearly different.

To justify oneself to someone is to give her a kind of *second-personal authority*.<sup>11</sup> It is to regard and treat her as having a standing to claim a justification from one (and hence to address claims to others at all). Second-personal authority of this kind is essentially tied to accountability. Justifying oneself to someone is part of holding oneself responsible or accountable to her. So justification to one another is what constitutes mutual accountability. When I justify myself to you, I hold myself answerable to you, and treat you as having the standing to claim this from me. You reciprocate and accord me the same standing when you justify yourself to me. As I understand it, therefore, the root contractualist idea is that this standing is one that you and I share. We have an equal (second-personal) authority to make claims of one another, which

10. *Utilitarianism*, chap. III, para. 10.

11. Here and below I draw on ideas and arguments that I develop further in *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, Mass.: Harvard University Press, forthcoming).

we respect in seeing each other as beings to whom we should be able to justify ourselves.

Understanding morality in terms of mutual accountability illuminates why Scanlon can say in "Contractualism" that agreement (of this sort) is what "morality is all about." If moral self-regulation essentially involves making ourselves answerable to one another, then agreement on fundamental principles is not simply a collective epistemic achievement, or a standard of our each doing what is right individually; it is an essential element of the fundamental moral *relation* (responsibility to one another). This idea is suggested also by the passages from "Preference" and "Process" quoted above. Urgency or importance of interests is justificatory weight in warranting *claims* on others. The question of when practices or institutions are *legitimate* in light of their "power to control or intervene" turns on when this is consistent with individuals' legitimate claims.

This interpretation is confirmed further by the way Scanlon distinguishes the concepts of reasonableness and rationality in *What We Owe to Each Other*.<sup>12</sup> To bring out this "familiar distinction in ordinary language," Scanlon describes a case in which water rights are being negotiated and there is a wealthy landowner who can control the negotiations, who believes himself entitled to his vast holdings, and who does not like having "the legitimacy of his position questioned" (p. 192). Scanlon says that while it would not be unreasonable to propose that each person is entitled to some minimum supply of water, it might not be rational to make this claim, since that might enrage the large landholder and make the situation worse for everybody (p. 192). Unlike the concept of the rational, that of the reasonable "presupposes a certain range of reasons that are taken to be relevant" (p. 192). We can now see what this range must be. Scanlon must be taking it to be part of the very idea of the reasonable, and hence of the notion of reasonable agreement, that the relevant reasons concern, or are able to support, legitimate *claims*. They must be ones we could offer in justifying ourselves *to* one another. Furthermore, to play that role, they must be ones that we can accept consistently with what we assume in so justifying ourselves, namely, that we each have an equal basic second-personal authority. Since we all stand,

12. *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998). References to page numbers will be placed parenthetically in the text.

fundamentally, in the relation of mutual accountability and have an equal standing to claim justification from one another, unequal claims must be able to be justified within that framework. They must be supportable from a standpoint in which we regard one another, as Rawls put it, as “self-originating source[s] of valid claims.”<sup>13</sup>

For this purpose, philosophical utilitarianism provides reasons of the wrong kind; without some further argument that links well-being or desirable outcomes to what we can claim from one another it is impotent to support claims of reasonable agreement or rejection. This is the response to the Sidgwickian challenge that I raised above. That something would advance someone’s well-being is intrinsically relevant to its desirability from the perspective of benevolent concern for that person. It is a conceptual truth that someone’s well-being is what we should want for her from that point of view. But there is no conceptual link between well-being and what someone can legitimately claim (unless, that is, we use ‘well-being’ as Scanlon does in “Preference” to mean that which will support such a claim on someone’s behalf). Of course, we may well believe that people have a claim to having their well-being promoted or, at least, to its not being harmed.<sup>14</sup> However, that is a substantive judgment about legitimate claims, not about well-being. A theory, such as philosophical utilitarianism, that takes it that the “fundamental moral facts are facts about well-being” conceives of morality in fundamentally different terms. The same goes, *mutatis mutandis*, for forms of philosophical consequentialism, such as G. E. Moore’s, that take morality or ethics to concern impersonal value.<sup>15</sup> For this reason the Sidgwickian cannot justifiably hold that it is *unreasonable* to reject principles other than the principle of utility, at least he cannot without some further argument that links pleasure to what we can legitimately claim.

The conceptual link to legitimate claims is even more obvious in the case of rights. That rights involve, as a conceptual matter, what we have

13. John Rawls, “Kantian Constructivism in Moral Theory,” *The Journal of Philosophy* 77 (1980): 515–72, at p. 546.

14. Utilitarianism is sometimes argued for in this way on grounds of fundamental fairness. Part of Scanlon’s point in “Preference,” however, is that well-being, interpreted in subjective terms, is not plausibly held to support claims on others. This might be disputed, of course. What seems clear, however, is that contractualism conceives of the terms of moral debate in a fundamentally different way than does philosophical utilitarianism.

15. I take this to be what lies behind Scanlon’s ruling out impersonal value as a ground for reasonably rejecting a principle. See *What We Owe to Each Other*, pp. 218–23.

standing to claim is widely accepted. Feinberg put it best, when he said that “claiming . . . gives rights their special moral significance.”<sup>16</sup> It follows from this that without some further story that links welfare to a fundamental authority to make claims of one another at all, an account like Mill’s and (that of “Rights” interpreted along consequentialist lines) will be impotent to justify rights. At best, it may provide a reason why it might be desirable, maybe even from a moral point of view, for us to regard one another as though we had certain rights. It cannot, however, justify rights themselves. Contractualism, on the other hand, takes the standing or authority to make claims of one another (what I am calling *equal second-personal authority*) to be fundamental. This makes it a more promising approach to justifying rights.

A natural reaction at this juncture might be to agree with this last point about rights, and even to agree that the part of morality that concerns duties “we owe to each other” might require the kind of second-personal framework within which I am placing contractualism, but then to maintain that consequentialism might nonetheless be a correct theory of moral right and wrong, more broadly conceived. However, what then would be meant by calling an action “morally wrong”? If, as Mill thought, the concept of wrong is itself conceptually connected to what we are rightly held *responsible* for avoiding, hence what we can morally demand or expect, so that it *follows* from the fact that an act is wrong that it is blameworthy unless the agent has some excuse, then it will also be conceptually tied to accountability and justifiability *to*.<sup>17</sup> When we judge conduct morally blameworthy and hold people responsible morally, we presume an authority to claim or demand certain conduct on the moral community’s behalf.

As I interpret him, this was Strawson’s point in “Freedom and Resentment” and the source of his rejection of consequentialist accounts of responsibility.<sup>18</sup> “Reactive attitudes” like indignation, resentment, guilt, and others involving moral blame implicitly make a claim or demand of

16. Joel Feinberg, “The Nature and Value of Rights,” in *Rights, Justice, and the Bounds of Liberty* (Princeton, N.J.: Princeton University Press, 1980), p. 155.

17. “We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience” (*Utilitarianism*, chap. V, para. 14).

18. P. F. Strawson, “Freedom and Resentment,” in *Studies in the Philosophy of Thought and Action* (London: Oxford University Press, 1968).

their object. Indeed, Strawson says, “the making of the demand is the proneness to such attitudes.”<sup>19</sup> So although consequentialist considerations can give us reasons why it might be *desirable* to hold people responsible for a certain act or to think an act culpable, they will not be, without some further story of the kind mentioned above, reasons of “the right sort” actually to hold people responsible, to support judgments of blameworthiness, and therefore to think their conduct really is wrong.<sup>20</sup> According to contractualism as I propose we understand it, however, mutual accountability is what morality is fundamentally about. It is its root idea. Moral obligations just are what we have standing to hold one another to as equal members of the moral community, where features that make us apt for mutual accountability also qualify us for membership, hence give us an equal authority to hold one another responsible.<sup>21</sup>

In “Contractualism,” Scanlon says that the source of moral motivation is “the desire to be able to justify one’s actions to others on grounds they could not reasonably reject” (p. 138). In *What We Owe to Each Other*, however, Scanlon is loath to take desires as sources of reasons in general, so he there claims that the “priority” of moral reasons as contractualism understands them is grounded in the “value” and “appeal” of standing to others in the relation of mutual recognition, both in itself and as an ineliminable aspect of other valuable relations such as friendship (pp. 158–68). In my view, however, this also provides a reason of the wrong kind to account adequately for the authority of moral obligations we owe to one another. To put the point in the Strawsonian terms of the previous paragraph, such considerations concern the *desirability* of regarding one another as having equal basic second-personal authority, so they cannot adequately account for that authority itself or for its

19. Ibid., pp. 92–93. Gary Watson and R. Jay Wallace make similar claims. See Gary Watson, “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme,” in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, ed. Ferdinand D. Schoeman (Cambridge: Cambridge University Press, 1987), pp. 263–64; R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1994), p. 19.

20. “Freedom and Resentment,” p. 74.

21. This may or may not be exhausted by “what we owe to each other.” If there are moral obligations, say, to beings with respect to whom justification *to* and mutual accountability of any kind makes no sense, then morality conceived as mutual accountability will still entail that violating these is something we presume the authority to hold individuals responsible for as members of the moral community.



proper role in our practical reasoning. As I see it, there simply is no way to account for this independently of a second-personal standpoint. It can only be established within a theory of practical reason that can adequately register the authority of this point of view.<sup>22</sup>

In closing, I want to note two further places in which the second-personal dimension has relevance in Scanlon's later essays, one in which it is latent in what he says and a second where his claims may suffer from giving it insufficient attention. "The Difficulty of Toleration" gets at something significant about the importance of toleration that I have not seen pointed out before. Many people have noted that tolerance is difficult; Bernard Williams called it an "impossible virtue" because it seems to require a kind of cognitive dissonance.<sup>23</sup> In addition, that toleration has to do with equal respect and requires political rights, and liberties such as freedom of religion and separation of church and state are also familiar points. What is distinctive about Scanlon's contribution is the idea that societies have informal as well as formal politics, that these help determine significant features of the cultural landscape within which we attempt to find meaning in our lives, and that respecting one another as equals therefore requires democracy in our informal, no less than in our formal, politics. Formal political equality can be achieved, at least approximately, by political institutions. Informal politics, however, works through countless conversations and encounters, articulate and inarticulate. To treat people as equals in these settings is to give them the same basic standing in "the informal politics of social life" (p. 190). This is more a matter of attitude and expressive behavior than it is of formal procedures. As Scanlon stresses, fully accepting as equals in these informal ways others with whom we deeply disagree is difficult. He gives the following excellent example:

I firmly believe that "creation science" is bogus and that science classes should not present scientific theory and religious doctrine as alternatives with similar and equal claim to the same kind of assent. . . . But I confess to feeling a certain sense of partisan zeal in such cases, a sense

22. I advance such an argument in chapters X and XI of *The Second-Person Standpoint*. And I argue, in chapter XII, that contractualism is best interpreted in second-personal terms.

23. Bernard Williams, "Toleration: An Impossible Virtue," in *Toleration: An Elusive Virtue*, ed. David Heyd (Princeton, N.J.: Princeton University Press, 1996).

of superiority over the people who propose such things and a desire not to let them win a point even if it did not cost anyone very much (p. 196).

As this shows, it is *regarding* others as having equal standing in informal politics that is the hard part. That concerns how we *relate* to others second-personally, through our reactions and attitudinal approach no less than the content of our speech, recognizing others' standing to influence society's shape regardless of their views: whether, for example, we listen to them and take them seriously. As Scanlon emphasizes, this is different from thinking that all *views* are equally worthy of respect; toleration is a virtue of respect for persons, not for views.

Finally, in "Promises and Contracts," Scanlon attempts to account for promissory and contractual rights within a contractualist theory. The main thrust of his argument, familiar from his other writings on promises, is that the basic moral phenomena have nothing essential to do with whether promising is a social practice, and so, contra Rawls, are not to be accounted for in terms of a duty of justice or fairness to do one's part in supporting just practices and institutions. Scanlon's basic line of argument here seems to me to be largely successful, although less so than it might be were it to be more sensitive to the second-personal character of promising, namely, that promises are forms of address that presuppose a second-personal authority of promiser and promisee.

Scanlon's argument depends on the claim that it would be unreasonable to reject *Principle F*:

If (1) in the absence of objectionable constraint, and with adequate understanding (or the ability to acquire such understanding) of his or her situation, *A* intentionally leads *B* to expect that *A* will do *X* unless *B* consents to *A*'s not doing so; (2) *A* knows that *B* wants to be assured of this; (3) *A* acts with the aim of providing this assurance, and has good reason to believe that he or she has done so; (4) *B* knows that *A* has the beliefs and intentions just described; (5) *A* intends for *B* to know this, and knows that *B* does know it; and (6) *B* knows that *A* has this knowledge and intent; then, in the absence of special justification, *A* must do *X* unless *B* consents to *X*'s not being done (p. 245).

Let us stipulate that these conditions are all satisfied in standard promises. Scanlon's claim is that the wrongness of breaking promises is to be explained by the wrongness of violating *Principle F*. Unless,

however, we interpret the idea of “providing assurance” in a (second-personal) way that already presupposes the authority of the assured to claim compliance, or assume, at least, that assurer and assured present themselves to one another as presupposing this in common, I doubt that this is so.

Consider, first, cases where one intentionally or negligently causes someone to expect that one will do something. Scanlon argues persuasively that in these cases the person whom one has led to have the expectation acquires a claim on one. If she has not yet relied on the expectation, then she has a claim to one’s correcting the expectation if it is mistaken. Further, if she has relied on it, then she has some claim to compensation. If she has not relied on the expectation, however, she has no claim that one fulfill the expectation, only that one correct any mistake in it. Promises are, of course, different. If one promises to do something and the other has not yet relied on an expectation that one will do it, one cannot simply disabuse her of the expectation if one wishes not to do what one promised. The other has a remaining valid claim to one’s doing what one promised and not just to due notice of nonperformance or to compensation in the case of reliance.

*Principle F* attempts to capture this through a complex set of conditions connected to providing assurance. This idea can be interpreted in two different ways. A *causal* interpretation of providing assurance would be: causing someone to be assured that something will happen, say, that one will do something. On a *second-personal* interpretation, however, providing assurance would involve an act of *assuring*, that is, a putatively claim-giving address to another of the same species as promise (maybe the very same thing as a promise).

Scanlon could avail himself of a second-personal interpretation of providing assurance without rendering *Principle F* idle. We know from the literature generated by Searle’s “How to Derive an ‘Ought’ From an ‘Is’” in the 1960s that it is one thing to show that the idea of promise cannot be understood except in terms of putative undertakings of obligation and another to establish that a genuinely binding obligation is created in fact.<sup>24</sup> What I wish to argue now, however, is that if we interpret providing assurance in causal terms, *Principle F* is not compelling.

24. In addition to John Searle, “How to Derive an ‘Ought’ from an ‘Is,’” *The Philosophical Review* 73 (1964): 43–58, see, e.g., R. M. Hare, “The Promising Game,” in *Theories of Ethics*, ed. Philippa Foot (London: Oxford University Press, 1967).

To show this, imagine a fairly elaborate case in which all of the conditions in *F* are satisfied. Suppose you want me to attend your party. If I were to promise to do so, then I would have given you by that distinctive form of second-personal address a claim to my doing so; you would warrantably expect this *of* me, and not just justifiably expect *that* I would do so. Suppose, first, however, that I simply intentionally cause you to expect that I will do so, say, by telling your friends that I will. In that case, if you have not relied on that expectation, I can discharge any claim you have of me by simply telling you in a timely fashion, before you rely on the expectation, that I am actually not going to come, if that is my intention. Again, however, that is not so if I promise or assure you that I will come. In that case, you will have a claim that cannot be discharged by telling you of my real intentions should I decide not to come, even if you have not yet relied on the expectation that I will come. But what if *Principle F* is satisfied with the various assurance conditions being interpreted in causal terms?

Suppose it is common knowledge between you and me that you have access to my email and that you check it regularly. (Let's not worry about why *that* would be!) On Monday, I send an email to a friend saying that I firmly intend to go to the party unless you wish me not to (the consent dimension of clause [1]), but that I will not promise or assure you myself, since I desire that, *were* I, counterfactually, to change my intention, you would have no remaining claim to my coming to the party. My complex desire for this extremely unlikely counterfactual situation notwithstanding, my firm intention is still to come to the party unless you wish otherwise. This satisfies clause (1). Since I know that you want to be assured that I will come unless you wish otherwise, (2) is satisfied. I send the message in order to cause you to be assured, and since I know you will read the message and, let us assume, believe it, I have good reason to believe that you will be thus assured. You believe of course that *were* I to change my mind I would not regard you as having any claim to my coming (therefore, that I am not *assuring you* [second-personally] that I will come), but you also believe that the probability that I will change my mind is negligible, so (3) is satisfied.

Suppose, then, that on Tuesday, I send a second email to my friend saying that because you have likely read and believed my first email, as I had hoped, you are therefore likely now to be assured that I will come to the party unless you wish me not to. Since you read this second

message, you come thereby to know that I wrote the first message with the aim of causing you to be thus assured and that I believe you have been, so (4) is satisfied. Since I wrote the second message with the intention that you would read and believe it, and I know that you will read it, condition (5) is satisfied.<sup>25</sup>

Suppose, then, that on Wednesday, I send a third email to my friend saying that because you have likely read and believed my second email, as I had hoped, you must now know that I had written the first with the intention of causing you to be assured and that I know that you have been. You read my third email, so you now know that I had this knowledge and intent; so condition (6) is satisfied.

Suppose now that I *do* change my mind and decide not to go to your party *and* that you have in no way yet relied on the expectation that I will go. You have, of course, a claim that I notify you of this in a timely fashion before you rely on it, but do you have a claim to my actually going unless you consent to my not going? It seems that you do not. After all, although I intentionally created in you the expectation that I would go, because I had a firm intention to do so, I also intentionally created in you the expectation that in the utterly unlikely instance that were I to change my mind, I would not in fact come nor regard you as having any claim to my coming. In the relevant respects, therefore, this seems to be like a case in which I intentionally create an expectation and in which you have a claim to compensation if you rely on it and to timely notice of my real intentions if I come to change them, but not to my actually doing what you had been led to expect if you have not yet relied on that expectation and if I have given you timely notice of my changed intention.

Even though conditions (1) through (6) are satisfied, then, it nonetheless seems that, if I have informed you otherwise and you have not relied on it, you have no remaining claim to my actually coming to your party. I conclude that there is a case where it would not be wrong to violate *Principle F* on a causal interpretation. Furthermore, although I have not directly argued that *Principle F*, interpreted causally, is one someone could reasonably reject, I hope it is reasonably clear that whether that

25. Actually, as Barbara Fried has pointed out to me, the right way of my complex intentions might make the same point. I have brought in the complexity to make it vivid how each of Scanlon's condition might be satisfied.

is so or not, its not being so would not explain why it is wrong to violate promises.

Promises have an essentially second-personal character that *Principle F* cannot capture when it is interpreted in causal terms. It is part of the very idea of a promise or an assurance that the addresser gives the addressee to understand that she thereby has a claim to the addresser's following through. Moreover, when I promise or assure someone that I will do something, I must already be authorizing the other to presuppose that she and I share a second-personal authority to make claims of, and be accountable to, one another. In this last section, I have been arguing that *Principle F* is impotent to explain the wrongness of breaking promises unless we interpret it in second-personal terms. I have been suggesting throughout this review essay that contractualism is itself best grounded in the equal second-personal authority that you and I must presuppose whenever we take up the second-person point of view.<sup>26</sup> If this is right, equal second-personal authority is the root from which contractualism, and its impressively branching theory of rights, grows.

26. For an argument for this claim, see chapters X–XI of *The Second-Person Standpoint*.