Reconstructing the Groundwork

Author(s): Marcus G. Singer

Source: *Ethics*, Apr., 1983, Vol. 93, No. 3 (Apr., 1983), pp. 566-578

Published by: The University of Chicago Press

Stable URL: https://www.jstor.org/stable/2380633

Reconstructing the Groundwork*

Marcus G. Singer

Robert Wolff's *The Autonomy of Reason* has been with us for some years and is increasingly referred to with respect and deference. In my estimation, however, it has not yet been adequately appraised. It is put forward as a commentary on *Kant's Groundwork of the Metaphysic of Morals*, and so it is regarded, but it is no ordinary commentary, scarcely one at all in any ordinary sense. Wolff is quite open about this: "In the Preface to my book *In Defense of Anarchism*, I remarked that . . . that essay presupposed an entire moral theory, which at that time I was quite unprepared to articulate. It was clear . . . that the first step in the development of such a theory would be an encounter with Kant, and this commentary has been that encounter."[1] Encounter indeed! Wolff's aim is not really to explain the text; it is, rather, to provide a "philosophical reconstruction" of it. For the *Groundwork* is said to be so riddled with confusions, unclarities, and contradictions that "it is simply not possible to fit all of [its] principal doctrines . . . into a single coherent chain of argument" (p. 1). Since Kant, despite the "incoherence" of what he actually said, was nonetheless "right in his original philosophical insights," Wolff sets out by this "reconstruction . . . to carry to completion the enterprise begun by Kant himself" (p. 4).

The nature of the enterprise, then, is clear enough. It is, to be sure, enhanced by Wolff's intriguing sense of drama. For we are told that "philosophical reconstruction is a gamble in which the reader wagers his time and energy on the text, hoping to win from it results that he could not have discovered by his own unaided efforts." Thus the scene is set for high adventure, in the manner of *Around the World in Eighty Days*, where the thrills of the journey only enhance and never supersede the suspense generated by the original wager. And thus

* I am grateful to Warner Wick for valuable suggestions regarding substance, style, and organization.

1. Robert Paul Wolff, *The Autonomy of Reason: A Commentary on Kant's "Groundwork of the Metaphysic of Morals"* (New York: Harper & Row, 1973), p. 223; all further references otherwise unspecified are to this. The *Groundwork* (hereafter *Gr.*) (1785) is cited by the page numbers of vol. 4 of the Prussian Academy (Ak.) edition of Kant's works, listed in the margin of H. J. Paton's translation, the one used here (London: Hutchinson's University Library, 1948), and also in the well-known translation by Lewis White Beck in his collection of Kant's *Critique of Practical Reason and Other Writings in Moral Philosophy* (Chicago: University of Chicago Press, 1949), later republished as *Foundations of the Metaphysics of Morals* (New York: Liberal Arts Press, 1959). The *Critique of Pure Reason* (abbreviated *KrV*) is cited in the translation by Norman Kemp Smith (London: Macmillan & Co., 1929), with the first edition cited as A and the second edition as B. The *Critique of Practical Reason* is abbreviated *KpV*.

Ethics 93 (April 1983): 566–578

Wolff says: "I am prepared to gamble on the *Groundwork* as a worthwhile text for the enterprise of philosophical reconstruction" (p. 4), which seems like a rather neat attempt at sleight of hand. The unquestioned importance and prestige of Kant's *Groundwork* is by this means to get transferred to Wolff's "reconstruction" of it. Sheer chutzpa of this order is relatively rare in philosophy, and one cannot but admire the élan of the enterprise and the incidental brilliance and even panache with which it is carried out. And it is certainly persuasive. Witness the respect in which it is held. But this respect is misplaced. What this *Autonomy* provides is not so much a philosophical reconstruction of the *Groundwork* as an imaginative expropriation of it for purposes Kant could not have recognized and could certainly not have adopted as his own. The result, though undoubtedly the expression of its author's autonomy, goes well beyond the limits of reason.

I. AUTONOMY AND ANARCHISM

Some key passages from some earlier essays are illuminating: "I have come to the conclusion that philosophical anarchism is true. . . . There could not be a state that has a right to command and whose subjects have a binding obligation to obey. . . . Every man has a fundamental duty to be autonomous, in Kant's sense of the term. . . ."[2] "The truly autonomous man gives laws to himself and thereby binds himself to what he conceives to be right"; "The primary obligation of man is autonomy, the refusal to be ruled . . ."; and "Insofar as a man fulfills his obligation to make himself the author of his decisions, he will resist the state's claim to have authority over him."[3]

I see no need to get involved here in any dispute about the meaning and validity of anarchism. We need simply to get clear on the present understanding of anarchism and autonomy and why it is supposed that Kant's principle of autonomy leads inexorably to anarchism, and the passages just reproduced help bring out just how Wolff's understanding of autonomy differs from Kant's. It is not Kant's view that we have an "obligation to make ourselves the authors of such commands as we may obey"—which is either the self-deception of supposing that each command we may obey is something we have "thought up" and then imposed on ourselves, or else the moral absurdity that we are never under any obligation to do anything other than what we want to do. Nor is autonomy, for Kant, the "refusal to be ruled." On Kant's view, we may act on, and only on, such maxims as we can will to be universal law, for it is only such maxims that we can regard as stemming from our own reason (*Wille*—rational will, practical reason). This is "the power of reason within us." And it is in this sense, and in this sense only, that we can, when we act morally—that is, on such maxims as we can will to be universal law—*regard ourselves* as acting on maxims that we have given ourselves as laws.

Wolff says, "Since the responsible man arrives at moral decisions which he expresses to himself in the form of imperatives, we may say that he gives laws to himself, or is self-legislating. In short, he is *autonomous*. As Kant argued, moral autonomy is a . . . submission to laws which one has made for oneself. The autonomous man, insofar as he is autonomous, is not subject to the will of another."[4]

2. R. P. Wolff, "On Violence," *Journal of Philosophy* 66 (1969): 601–16, p. 607.

3. R. P. Wolff, *In Defense of Anarchism* (New York: Harper & Row, 1970), pp. 22, 18, 17.

4. Ibid., pp. 13–14.

But this is not what Kant argued, and this is not autonomy "in Kant's sense of the term." Moral laws are not laws that one has actually and in fact "made for oneself." For no one can literally make a moral law. And the political condition of being subject to the will of another is not incompatible with being morally autonomous. Moreover, the autonomous person is subject to the will—the rule—of reason: the rational will or the practical reason (*Wille*) within. The rule one is subject to is just what one could impose on oneself, if one were wholly rational.

We are told, with no doubt unconscious humor, that "the responsible man is not capricious or anarchic, for he does acknowledge himself bound by moral restraints."[5] Since we are also told that "the responsible man . . . is autonomous," it would seem to follow that the autonomous person is not anarchic. Whether it does or not (it does not) is not something we need pursue. We need only notice that talk of "acknowledging oneself bound by moral constraints" is in this context incoherent. What is it to be bound by bonds that one can untie at will? This is not to be bound at all. If one is genuinely bound one is bound by bonds that one cannot untie at will, and whether one is bound or not is something one has to *recognize*, and is not something that can be either brought about or terminated by wish or want or choice. *This* is Kant's concept of autonomy. For, on Kant's concept, whether one is bound (obligated) is not a matter of anyone's choice (*Willkür*), but a matter of rational will (*Wille*).[6]

II. THE AUTONOMY THESIS

Wolff thinks that Kant is committed to "two incompatible doctrines. On the one hand, he believes that there are objective, substantive, categorical moral principles which all rational agents, insofar as they are rational, acknowledge and obey. If this is true, then the notion of self-legislation seems vacuous. On the other hand, he believes . . . that rational agents are bound to substantive policies only insofar as they have freely chosen those policies. But if this is true, then one must give up the belief in objective substantive principles" (p. 181). The conclusion here is a non sequitur. It is true that if there are substantive objective principles then the notion of self-legislation, as Wolff understands it (p. 178), is vacuous. But the Kantian idea of autonomy is not thereby vacuous.

Wolff wants to maintain that "men are bound by substantive policies only insofar as and only because they have legislated those policies themselves" (p. 181). But Kant is not talking about what men actually do choose or agree on. He is talking about rational beings and the rational nature of human beings (who are not perfectly rational). The key notion for Kant is what a rational agent *would* choose, and this is the key to understanding the difference between the Wolffian notion of autonomy and the Kantian notion of autonomy.

If it makes sense to speak of binding oneself by an act of choice or agreement (and it does), it is because there already exists, apart from any choice or agreement,

5. Ibid., p. 13.

6. The important distinction between *Wille* and *Willkür*, for some reason wholly ignored by Wolff, is developed in the *Metaphysik der Sitten (MdS)* of 1797 and is used in the *Religion* (1793). See the editorial note by T. K. Abbott in his collection *Kant's Theory of Ethics*, 6th ed. (London: Longmans, Green & Co., 1909), p. 268; also the introductory essay by J. R. Silber, "The Ethical Significance of Kant's *Religion*," in the 1960 edition of the translation by T. M. Greene and H. H. Hudson of Kant's *Religion within the Limits of Reason Alone* (New York: Harper Torchbooks), pp. xciv ff. and cxxxix.

a substantive standard that agreements or choices of certain sorts are binding. The very idea of a binding agreement cannot itself be brought into existence by a binding agreement, and it cannot be on the basis of a promise that promises are binding. This is not something that is a matter of choice, and we have here one of the significant and rational limits of autonomy.

Thus the idea that "men are bound by substantive policies only insofar as and only because they have legislated those policies themselves" is incoherent because it presupposes the notion of "bound by" and of "legislation" (which itself presupposes the notion of law and hence also the notion of "binding"), which such a view cannot, in turn, give an account of. But Wolff also tries to foist this idea on Kant, even though, as he himself sees in one place (only to forget it later), "acting only on laws that one has given to oneself and being bound by them only because one has so given them is not at all what Kant has in mind" (p. 178). Quite right. This is not at all what Kant has in mind. The notion that persons are bound only by those maxims that they in fact accept as binding (supposing this makes sense) is not at all the Kantian idea of autonomy, which instead provides a test of which maxims are morally acceptable and which are not. The laws that are "valid because self-legislated" are just those that would be accepted, and in this sense imposed on themselves, by perfectly rational beings and are just those that could be willed to be universal laws. The operative criterion is, if a perfectly rational being *could not* will otherwise, an imperfectly rational being *ought not* to. Therefore a maxim that could not be willed to be a universal law *cannot be* such a law.

Though Kant sometimes speaks of the principle of autonomy as "the supreme principle of morality," it is nonetheless evident that the principle of autonomy necessarily presupposes the principle of universality. Thus Kant says: "The principle of autonomy is 'Never to choose except in such a way that in the same volition the *maxims of your choice* are also *present as universal law*' " (*Gr.* 440, italics added). Here it is clear that although one's maxims can be freely chosen—it is one's maxims that are a matter of choice—it is not a matter of choice whether one's maxim can be willed to be universal law. This is something that cannot be *chosen*, and if it is "legislated," it is legislated by the rational will (*Wille*) not by the will (*Willkür*, the power of choice) that is free to choose its maxims as it pleases.

Kant says "the supreme law" is "act always on that maxim whose *universality* as a law you can at the same time will. This is the one principle on which a will can never be at variance with itself . . . " (*Gr.* 437, italics added). And "Every rational being, as an end in himself, *must be able to regard himself as also* the maker of universal law in respect of any law whatever to which he may be subjected" (*Gr.* 438, italics added).[7] That is, it must not be contrary to reason to suppose that a perfectly rational lawgiver could have given such a law. This condition holds just in case one could rationally and consistently accept such a law. For no one, not even an all-powerful being, can literally *make a moral law*. Talk of "making law" through one's will is only particularly potent metaphor. Though Kant's language on this point is frequently obscure and difficult, there are places where the point is made clearly and unequivocally. For instance: "Except in the case of

7. Abbott (p. 56) translates this as "must be able to regard himself as also legislating universally"; while Beck (in his collection *Kant's Moral Philosophy* [see n. 1 above], p. 95) has "and thus as giving universal law." Giving is not the same as making, and "giving law" seems a more accurate and sensible translation than Paton's "making law."

contingent laws, the lawgiver is not their author; he merely declares that they are in accordance with his will. It follows that no one, not even God, can be the author of the laws of morality, since they do not originate from choice but from practical necessity. If they were not necessary, it is conceivable that lying might be a virtue. But the moral laws can nevertheless be subject to a lawgiver."[8] It is clear from this that autonomy, as Wolff understands it, is not the "key to Kant's moral philosophy." Indeed, autonomy as Wolff understands it is no part of Kant's philosophy at all.

Autonomy, though a necessary condition for morality, is not identical with it. As Kant reminds us in a revealing and well-known footnote in the *Critique of Practical Reason* (Ak., vol. 5, p. 4; Beck, p. 119; Abbott, p. 88), while (1) freedom is the *ratio essendi* of the moral law, (2) the moral law is the *ratio cognoscendi* of freedom. It follows from this that (3) the moral law is not identical with freedom. But (4) autonomy is identical with (or, rather, equivalent to) freedom. It follows that (5) autonomy is not identical with the moral law. And the fact that Kant provides no examples of the application of the "principle of autonomy" supports this. What would be shown by examples is how in each instance no one could give (that is, take responsibility for) such a universal law, because no one could rationally accept it. But try to use the principle of autonomy as an independent formula and it is hopeless. The principle of universality must be presupposed throughout. Since the principle of autonomy must always be applied through the form of universalizability, when the law someone in this way gives to himself—that is, one's maxim—could not be willed to be a universal law, then the autonomy manifested in *this* species of self-government is, though a possible autonomy, nonetheless an immoral autonomy.

III. THE NO-MORALITY THESIS

We now know what to make of the autonomy thesis, that "the key to Kant's moral philosophy" is to be found in "the notion of autonomy," as Wolff understands it (p. 178). We must now take due note of its concomitant, that there can be no substantive binding principles of action; and that, in particular, the Categorical Imperative, with its test of universalization, cannot serve to establish "any substantive practical laws" (pp. 76, 50–51, 86–88, 90, 190, 198). The Categorical Imperative, from the point of view of anarchistic autonomy, is a purely formal negative test that serves only to rule *out* certain policies as inconsistent but cannot rule *in* any as binding or obligatory or valid (pp. 86, 198). There must be something more, it is said, "some appeal to the Idea of the Good or to a theory of obligatory ends," without which "the purely formal Categorical Imperative . . . " cannot "dictate a substantive policy" (pp. 87, 88). But, we are told, "there is nothing good in itself . . . there is therefore no valid theory of the objectively good . . . , and hence . . . there are not and could not be any obligatory ends" (p. 132). This idea, that there are no valid substantive principles determining what is right and wrong and no valid test of the morality of conduct apart from one's own "freely chosen commitments" (p. 219), I call, for obvious reasons, the no-morality thesis.

8. Kant's *Lectures on Ethics*, trans. (translation here somewhat modified) Louis Infield (London: Methuen & Co., 1930), pp. 51–52; see also *MdS*, Ak., vol. 6, p. 227; Abbott, p. 283. I am grateful to H. B. Acton's *Kant's Moral Philosophy* (London: Macmillan & Co., 1970), pp. 38–39, for its illuminating remarks on this point.

Wolff thinks that the Categorical Imperative is a "purely formal principle" (p. 48) wholly analogous to "the law of contradiction" and that, therefore, just as we cannot "derive the laws of nature from the law of contradiction alone," we cannot derive "substantive practical laws from the purely formal Categorical Imperative" (p. 76). Just as "the law of contradiction is a merely negative or necessary condition of the truth of a theoretical judgment," which "rules out those judgments which are self-contradictory but does not discriminate between true and false consistent judgments," so the Categorical Imperative can be used only "to *rule out* those maxims . . . which are . . . inconsistent" (pp. 76, 49–50).

Thus all Wolff in his *Autonomy* can make out of the Categorical Imperative is that it says to us, "Don't violate the law of contradiction." But the Categorical Imperative is not needed to rule out inconsistent maxims, nor is it put forward as having that function. If a maxim is inconsistent with itself then it *cannot* be acted on and there can be no question of whether it is right or wrong to act on it. Nor is there any hint in Kant's doctrine of the inanity that we ought to adopt consistent maxims. Kant's criterion is that we must be able to will the maxim of our action *to be a universal law*, and the inconsistency Kant is talking about is not in the original maxim itself but in its universalization. A maxim that cannot be universalized without contradiction is one that ought not to be acted on, though it can be, and if a maxim is one that ought not to be acted on then its contradictory ought to be acted on. It is true that this does not by itself determine just which of the many courses of action still open to one ought to be adopted, but that is of no matter. If your maxim is to make a promise without the intention of keeping it whenever it seems to be to your advantage to do so, then that you cannot will this maxim to be a universal law shows that you ought not to make a lying promise, and at the same time shows, not that you ought to make honest promises (for there is no moral necessity to make promises), but that you ought to make *only* honest promises.

If one is looking for the Categorical Imperative to tell one just what in particular to do in each and every circumstance of life, one is looking for the impossible. No moral principle can do this, and there is no reason why it should. The various casuistical questions in the *Doctrine of Virtue* make it clear that Kant did not suppose that the Categorical Imperative could serve by itself to determine what in detail ought to be done in any and every concrete situation.[9] Wolff's no-morality thesis is an echo of the old Hegelian view that the Categorical Imperative is a merely formal principle devoid of content. But it is plainly false that "Kant . . . tries to perform the manifestly impossible feat of deriving substantive practical laws from the purely formal Categorical Imperative" (p. 76). The Categorical Imperative is not a premise in practical inference, it is a criterion or test. One must always start with a maxim, and the content is in the maxim to be tested. If there is no determinate maxim to begin with, then there is no question for the principle to be applied to. The Categorical Imperative is in this respect not at all analogous to the "law of contradiction," which plays exactly the same role in practical reasoning as it does in theoretical.

9. These casuistical questions can be located quickly by means of the index to James Ellington's translation of the *Tugendlehre*—pt. 2 of the *Metaphysik der Sitten*—entitled *The Metaphysical Principles of Virtue* (Indianapolis: Bobbs-Merrill Co., 1964).

Why is it supposed that Kant must either independently of the Categorical Imperative establish a doctrine of obligatory ends,[10] or else derive obligatory ends from the Categorical Imperative itself? One can only conjecture.

Wolff is wholly captured by a means-end model of action according to which every action is undertaken as a means to an end and according to which a principle can be substantive and binding only if there are obligatory ends to which it prescribes an action as a means. Consequently he finds it impossible to distinguish, in the end, between hypothetical and categorical imperatives, since on the model he has in this way adopted, both must enjoin an action as means to some end, regarded as good. Since his own view is that "there are no good reasons for the choice of ends, save those which make reference to other ends already chosen and appeal to considerations of consistency, compatibility, and so forth," he thinks that "there are no ends which all rational agents as such have good reasons to choose" (pp. 137–38) and that therefore there can be "no substantive principles of practical reason which are valid for all rational agents as such." And he thinks that this must be Kant's doctrine as well. Wolff is thus led to foist on Kant the view that just as hypothetical imperatives have a "preamble" (hypothetical clause), namely, "If you want to attain *Y*, . . . ", so all categorical imperatives really have a preamble, which, though implicit, is the same for all, namely, "Being as you are a rational agent and having as your end the realization of the good . . . " (pp. 130, 149–50). All categorical imperatives, then, turn out to be really and truly hypothetical imperatives, enjoining the means that ought to be adopted to the "realization of The Good"; and with all imperatives the only problem will be the essentially calculative one of choosing effective means to antecedently determined ends.

We thus have presented to us the rather engaging spectacle of a Humean-utilitarian attempting to appropriate the Kantian moral philosophy. When Wolff says that "the passions are as dependent upon reason for a choice of appropriate means as reason is upon the passions for the impulse of a desired end" (pp. 119–20) and that desire is "a source of ends toward which reason guides us" (p. 126), he makes it amply though perhaps unconsciously clear that he is simply taking for granted a Humean view according to which reason can evaluate only the means to passionately determined ends while the ends themselves cannot be judged on any criteria other than purely formal ones of consistency. And by this means not only are categorical imperatives reduced to hypothetical ones, but to hypothetical ones having a somewhat peculiar shape. Thus, on Wolff's reconstruction of Kant, whereas "an imperative of skill says that an action is good to some *possible* purpose" and "an imperative of prudence says that an action is

10. The confusions in this *Autonomy* about obligatory or objective ends have been dealt with clearly and effectively, though of course only by indirection, by Mary Gregor, in *Laws of Freedom* (Oxford: Clarendon Press, 1963), pp. 85–94, 83–84, also p. 40. Curiously enough, Wolff reviewed this book: *Journal of Philosophy* 61 (1964): 226–32. The best account I have seen of the notion of objective ends is by John Atwell, "Objective Ends in Kant's Ethics," *Archiv für Geschichte der Philosophie* 56, no. 2 (1974): 156–71. (Atwell has since pointed out to me that "Kant never uses the expression 'obligatory end,' though he does of course use 'ends which are duties.'" I think this must be right, because I cannot recall any place where Kant actually uses this expression, and Atwell knows the Kantian corpus as well as anyone. "Obligatory end," then, may well have been an autonomous introduction by our autonomist, and I see that I must simply have fallen into the way of using it.)

good to some *actual* purpose" (happiness), "an imperative of morality says that an action is good to some *necessary* purpose" (p. 132), and this necessary purpose is "a necessary or obligatory end." Otherwise, muses Wolff, there is just no sense to be made of the notion.

There are just two things wrong with this reconstruction. One is that it misrepresents Kant; the other is that it involves an absurdity. Note first that happiness, on Kant's view, is not simply an actual purpose but a necessary purpose of imperfectly rational beings, a "purpose which they not only *can* have, but which we can assume with certainty that they all *do* have by a natural necessity," so it is "a purpose which we can presuppose *a priori* and with certainty to be present in every man because it belongs to his very being" (*Gr.* 415–16). Note further that a categorical imperative "would be one which represented an action as objectively necessary in itself apart from its relation to a *further* end," and it "declares an action to be objectively necessary in itself without reference to some purpose—that is, even without any *further* end . . . " (*Gr.* 414–15, italics added). Now this way of distinguishing categorical from hypothetical imperatives does not involve the sort of confusion ascribed to it. For Wolff, Kant's "habit of describing categorical imperatives as commanding actions without reference to ends" is "the principal source of confusion" (p. 130), and "this way of talking just doesn't make any sense." Why? Only because this "way of talking" violates the means-end model of action that Wolff has imported into the proceedings. Thus he says: "A categorical imperative cannot 'directly command a certain conduct without making its condition some purpose to be reached by it,' for that is the same as saying that it commands an agent to engage in purposive action which has no purpose" (p. 131). Of course "purposive action which has no purpose" seems self-contradictory, but that is not what Kant is talking about. Kant is talking about engaging in purposive action that has no further purpose or further end, that is, end other than itself; and there is nothing self-contradictory about this.

As to the notion of obligatory ends, what Kant says is that a categorical imperative represents an action "as good in itself and therefore necessary . . . for a will which of itself accords with reason." Kant says, it is true, that "an imperative . . . tells me which of my possible actions would be good," but he also says that "a categorical imperative . . . declares an action to be objectively necessary *in itself* without reference to some purpose—that is, even without any *further* end" (*Gr.* 414–15, italics added). Kant makes it abundantly clear, especially in the *Second Critique* but also elsewhere, that "the concept of good and evil must not be determined before the moral law . . . but only after it and by means of it . . . it is not the concept of good as an object that determines the moral law . . . it is the moral law that first determines the concept of good . . . " (*KpV*, Abbott, pp. 154–55; Beck, pp. 171–72; Ak., vol. 5, pp. 62–63). This makes it plainly false that "Kant is totally concerned with consequences, for he believes that a moral agent should be moved by the thought of the good at which his actions aim" (p. 85).

Consider now the treatment accorded to the second and fourth of what are called "the Famous Four Examples" (p. 161). Wolff disposes of the second (false promising) in a unique way, by adopting it into his own contractual theory. He first claims that "the argument actually given by Kant is hopelessly inadequate," on the ground that "the contradictory nature of the policy cannot possibly be demonstrated by appeal to the contingent fact . . . that people tend to disbelieve a persistent promise breaker" (p. 166). Nonetheless he accepts its conclusion (or

something very like it) on the ground that "since Kant's theory is . . . essentially a contract theory of obligation, false promising or breach of contract must necessarily go to the heart of his doctrine" (p. 165); and he thinks that false promising, where, as it is put, one has already "adopted a practice of promising" (p. 167), involves an inconsistency. This, however, misses the point. The example does not depend on anyone's antecedently "adopting a practice of promising." For the purpose at hand, one has adopted the practice merely by making a promise. In the mere act of borrowing money one makes a promise to repay; otherwise it would not be borrowing but would be either theft or a gift. And Wolff overlooks the essential point of universalization. One can *adopt* a policy of making promises one has no intention of keeping, but one could not (consistently) will such a policy to be a universal law, for on such a "law" there could be no such practice as promising.

The fourth example (beneficence) is neither dismissed nor appropriated, and it is evident that the idea that we have a duty to help others who are in need of help is one that cannot be readily accommodated into a contractual theory of obligation. But it is argued that it is invalid, on the following basis: "Suppose an individual adopts it as his policy never to set for himself an end whose achievement appears to require the cooperation of others and to forswear any ends he has adopted as soon as it turns out that such cooperation is needed. Under these circumstances, he could consistently will that his maxim of selfishness should be a universal law of nature, for he could be certain a priori that he would never find himself willing an end which that natural law obstructed" (pp. 170–71). The argument is ingenious, but nonetheless not sound. We are being asked to imagine someone who adopts the maxim of never setting for himself any end for whose attainment he might require the help of others and to give up any end as soon as it appears that for its attainment he would need such help. Now we certainly can imagine such an eccentric, and furthermore we can imagine such a being behaving consistently with this maxim, though he would have to cut himself off from all human intercourse to do so. But that is why such a maxim could not consistently be willed to be a universal law. For such a person might need the help of some in order to ward off the well-meant but unwanted assistance of others. We can imagine someone, even a rigid Stoic, who is prepared to renounce *any* end in order not to be assisted by someone else in the attainment of it. But we are not able to imagine anyone who is prepared consistently to renounce all of his ends in order to achieve *this one*. Because *one* of the ends of such a being is not to be assisted by anyone else in achieving any of his ends, and if he should need help—and he has no assurance that he would not—to achieve *this* end, he would have to renounce *all* his ends. This is what universalization of this maxim would lead to, and this is self-contradictory.

IV. THE CONTRACT THESIS

It is essential to the self-imposed task of reconstructing the *Groundwork* both that some sort of contract theory of obligation be seen to be true and also that the *Groundwork*, when seen in the right light, be seen to contain one. So it is said that "a contractual theory of moral obligation is the most plausible way of construing the argument of the *Groundwork*" and that "the most powerful passages of the text, such as those dealing with autonomy . . . clearly rely upon a contractual theory of obligation" (p. 168). It remains for us to see just what, on this understanding of the matter, a contract theory of moral obligation is to be taken to be and how much sense is to be found in it.

The theory seems to consist in the conjunction of these two distinct though related propositions: "The only rules I can be bound by are those that I impose on myself"; and "The only rules I can be bound by are those that I agree with someone else (promise, contract with, someone else) I will be bound by." The first can be called, for obvious reasons, the autonomy thesis; the second, the contract thesis.

The autonomy thesis is what is suggested by the understanding of autonomy as consisting in "acting only on laws that one has given to oneself and being bound by them only because one has so given them" (p. 178), and by such statements as "rational agents are bound to substantive policies only insofar as they have freely chosen those policies" and "men are bound by substantive policies only insofar as and only because they have legislated those policies themselves" (p. 181).

The contract thesis, which adds to the autonomy thesis the important element of contracting or agreeing with others and to the relatively simple task of choosing the more complex task of negotiating, is suggested by such statements as "moral obligations are the consequences of contractual commitments among agents who choose to bind themselves to one another" (p. 138); and "in a narrow sense of 'promising,' the practice of promising is only one of many practices I and others might collectively adopt. But in a looser sense, all contracting might be spoken of as promising. In this looser sense, immorality consists simply in some sort of breach of promise" (p. 169); and especially this: "I am persuaded that moral obligations, strictly so-called, arise from freely chosen contractual commitments between or among interaction with one another. [1] Where such contractual commitments do not exist, [2] cannot plausibly be construed as having been tacitly entered into, and [3] cannot even be supposed to be the sort that *would* be entered into if the persons were to attempt some collective agreement, then no moral obligations bind one person to another" (p. 219, enumeration added).

Now what rules out the autonomy thesis is that it confuses maxims with "valid" maxims. A maxim, after all, can be consciously self-imposed, so that a rule I impose on myself is a maxim. But not all maxims are binding, and some maxims are immoral. So I am not bound by *all* the rules I "impose on myself," and there is no way, on the autonomy thesis itself, in which the distinction between binding and nonbinding maxims can be made out.

The contract thesis has greater plausibility, for the notions of a promise, or an agreement or a contract, and of its being binding, are already to hand. But there is really no force at all in the notion that the *only* rules I can be bound by are those that I explicitly agree or contract to obey. Merely by living in society I incur certain duties and obligations, quite apart from any I have acquired specifically through contracting. Hence we get the notion that the contract in question need not have been (1) overtly and explicitly entered into, but can be (2) tacitly or implicitly entered into, or else (3) conjectured as what would have been entered into, if the attempt to arrive at an agreement had been made—in other words, it can be a hypothetical contract.

But, though these additions of tacit and hypothetical contracts constitute a welcome expansion of the view to cover moral realities, this expansion still does nothing to handle the fact that the rule that promises (contracts) are binding cannot itself be construed or understood as the sort of rule that one imposes on oneself or that derives from a contract. Furthermore, a hypothetical contract is not itself a contract that one has actually entered into or something that one has

actually committed oneself to, so a hypothetical contract is not at all the same as a binding contract.

It seems pretty clear that this extension of the notion of contract to include, in addition to explicit contracts, also implicit and hypothetical contracts, is undertaken because the obligations and duties we actually recognize cannot all be subsumed under or explained by the contract thesis. But in order to determine whether we can plausibly be supposed to have implicitly contracted, or whether we would have contracted if the occasion had arisen, we must already have a pretty clear idea of our duties and obligations as well as some rudimentary moral knowledge of when a contract is binding and when it is not. A habitual liar and cheat, or a professional hit man, would be prepared to enter into contracts that others, with a somewhat greater regard for the requirements of morality, would regard as abominable and of a certainty not morally binding. A professional hit man, after all, has *contracted* to kill; the intended victim, though not a party to the contract, is said to "have a contract out" on him.

Kant starts from the notion of universalizability and determines that what one could consistently be willing to have adopted as a universal law is a rule that one could be willing to have imposed on oneself, so that it is a rule that *can be construed as* self-imposed and as obligatory because it is *in this way* self-imposed. This is Kantian autonomy, and it is dependent on and derived from the notions of universality and rationality.

Wolff starts with the notion, which he thinks is involved in Kantian autonomy, of what a person actually in fact imposes on himself; and ends there as well, since in the process he rejects universalizability altogether. The contract thesis does not add to the autonomy thesis anything that essentially departs from this—it still comes down to what persons, in their full particularity, as they actually are, are prepared to impose on themselves. And in his rush to justify this sort of autonomy, on which nothing that an Idi Amin does is wrong as long as he doesn't break any of his agreements, Wolff in his Autonomy has failed to see that it can make no sense of morality.

APPENDIX: THE METAPHYSICAL BACKGROUND

The metaphysical presuppositions of this proposed reconstruction should at least be noted, though they cannot here be treated in detail. It is Wolff's contention that not only is the *Groundwork* beset by internal contradictions and confusions, but that large portions of it are inconsistent with the "deeper doctrines," as they are called, of the *First Critique*. It is said that even though "the most advanced teaching of the *Critique of Pure Reason*" is "too valuable to be ignored," nonetheless "Kant himself frequently did just that." So Wolff sets out to interpret "the argument of the *Groundwork*" by these "deeper doctrines" and to "reject as misguided whatever hopelessly conflicts with them" (p. 8). Thus some interpretations of the *First Critique* that Wolff has elaborated elsewhere are brought to bear to determine what, in Kant's moral philosophy, is "superficial" and to be rejected and what is "deeper" and to be adopted.

It is not, I think, very convincingly explained how Kant could have been so confused about the contents and implications of his own philosophy, which would seem, on this account, to have more things in it than Kant himself ever dreamt of. It is claimed that "Kant frequently makes false claims about the implications of his own theories" (p. 30). This, no doubt, is true, as it is true of practically every great thinker. But what is telling here is the way in which the point is used.

For it turns out to be one of the foundations of the claim, explicit later but implicit at the outset, that Kant's theory—when rationally reconstructed of course—"entails . . . anarchism," though the reader is quickly told that "Kant did not hold this view of the implications of his moral philosophy" (pp. 129–30n.).

Some of the features of Kant's moral philosophy which emerge from what is said to be "a sufficiently vigorous interpretation of the *First Critique*" (p. 8) are these. Although the key feature of Kant's moral philosophy, on the superficial level on which Kant is said to have understood it, is the conflict between duty and inclination, and the counterpart struggle between reason and desire for the governance of the will (pp. 67–68, 117–18), nonetheless there can be no struggle or opposition or conflict between duty and inclination (or between reason and desire, or reason and inclination), because it is a key precept of the Kantian metaphysics that there can be no interaction between noumena and phenomena. Furthermore, since "Kant is committed by his theory of knowledge to a doctrine of psychological determinism" (p. 66), he can give no account of moral evil (p. 211) or of how morally blameworthy action can occur (p. 122).

This idea, that Kant can give no account of how a person can be free and still act immorally, is one that has a respectable ancestry. And with some cause. For Kant's doctrine on this point is not only difficult but was not always available, since it was one that developed through time. But it is simply not true that Kant does not address the problem of evil—he does so explicitly in the *Religion* of 1793 and in the later *Metaphysik der Sitten*. His solution depends on his difficult and evolving distinction between *Willkür* and *Wille*—between choice (or elective will) and rational will (practical reason)—not present either in the *Groundwork* or in the *First Critique*. Thus Kant says: "The source of evil cannot lie in an object determining the will [*Willkür*] through inclination, nor yet in a natural impulse; it can lie only in a rule made by the will [Willkür] for the use of its freedom, that is, only in a maxim."[11] The details need not detain us. It is surely false that "there is no solution for this problem within the framework of Kant's moral philosophy" (p. 173).

As to the alleged impossibility of reconciling the conflict between duty and inclination with Kant's metaphysics, it is sufficient to notice a passage from the first edition of the *First Critique*: "Transcendental freedom is . . . as it would seem, contrary to the law of nature . . . and therefore remains a problem. . . . [But] it is a *merely speculative question, which we can leave aside so long as we are considering what ought or ought not to be done . . .*" (*KrV*, A803 = B831 [italics added]). Hardly a problem, and certainly not a consideration, that Kant was not aware of. To be sure, it was not dealt with in the *Groundwork*, but the *Groundwork* was simply that, a *ground*work.

Finally, in coming down, as it does, on the side of the antithesis of the Third Antinomy, and claiming that all events, including human actions and the outcome of human deliberations and negotiations, are both explicable and predictable, "at least in principle, by the natural laws of physiology, psychology, society, and history" (p. 221), this "reconstruction" manifests what Kant himself calls a "dogmatic empiricism" (*KrV*, A471 = B499) in attempting to pass off as proved scientific fact what is only metaphysical dogma. In the *First* and *Second Critiques*, it is true, Kant did maintain the thesis of the predictability, "in principle," of human conduct

11. Kant, *Religion*, trans. Greene and Hudson, p. 17 (Abbott, pp. 327–28). Cf. pp. 18–20, 31–32 (Abbott, pp. 330–31, 343).

(*KpV*, Ak., vol. 5, p. 99; Beck, pp. 204–5; Abbott, p. 193), though Kant expressly added: "Nevertheless we may maintain that the man is free." But *this* side of the Third Antinomy is no more Kant's theory than the *other* side and is no more deep or profound. And one who would comment on Kant's philosophy should surely be aware that Kant's confidence in this metaphysical faith was undermined somewhat by the time of the *Third Critique* (1790). If the law of cause and effect could not enable human reason to comprehend or predict the production of even a blade of grass,[12] then it cannot enable us to comprehend or predict the outcome of deliberation or negotiation. To be sure, the *Groundwork* needs interpretation. But it is not in need of this sort of *transformation*, in which it is divorced from the whole of Kant's developing and developed philosophy.

12. Immanuel Kant, *Critique of Judgment*, pt. 2, Ak. vol. 5, p. 400. Trans. J. C. Meredith in *Kant's Critique of Teleological Judgment* (Oxford: Clarendon Press, 1928), p. 30. Cf. Gregor, p. 133.