Kantian Practical Reason Defended

Stephen L. Darwall

Ethics, Vol. 96, No. 1 (Oct., 1985), 89-99.

Stable URL:
http://links.jstor.org/sici?sici=0014-1704%28198510%2996%3A1%3C89%3AKPRD%3E2.0.CO%3B2-E

Ethics is currently published by The University of Chicago Press.

Kantian Practical Reason Defended

Stephen L. Darwall

There are two ways in which philosophical controversialists can approach a classical opponent of their views. They can attempt to refute him, or they can try to show that, while generally assumed to be an opponent, the philosopher really was not, at least when he was thinking clearly. Of these two strategies, the latter, if it can be pulled off, is dialectically superior. Directly attacking a classical figure only encourages current followers to defend him. Contemporary opposition is much likelier to be neutralized if the controversialist can show important elements of the classical figure's view to support his own. If successful, this can at least sow seeds of discord in the opposing camp. Internecine disagreements may emerge about what *is* central to the venerated figure's views, what peripheral, where conflict exists, and so on. At most, there is the hope that embracing a core doctrine can completely co-opt the opposition. If the core doctrine is one most cherished by opponents and if current opposition can be shown to be based on other, less central elements of the classical figure's view that are at odds with it, then opponents may be brought to see their error in a way that is nonthreatening and that does not risk philosophical parricide.

David Gauthier is a master of this second technique. In "David Hume, Contractarian," he sought to win over utilitarian opponents of contractarianism by showing that Hume, often regarded as a proto-utilitarian, actually held a contractarian theory of justice.[1] Such an argument bids to sow seeds of doubt in the utilitarian mind. After all, they might think, if David Hume is a contractarian, then why shouldn't I be one also?

In the present essay, Gauthier turns to Kant.[2] On the model of the earlier papers, an apt title might be "Immanuel Kant: Constrained

1. David Gauthier, "David Hume, Contractarian," *Philosophical Review* 88 (1979): 3–38. For a variation of strategy, arguing that a classical proponent of one's own view actually held a position less obnoxious to opponents than generally believed, see Gauthier's "Thomas Hobbes: Moral Theorist," *Journal of Philosophy* 76 (1979): 547–59. Gauthier there argues against the conventional wisdom that Hobbes is committed to an egoistic conception of rationality that is too slender a basis for morality. He maintains that Hobbes has the theoretical resources to hold a "doubly conventionalistic" view in which instrumental egoistic rationality is superseded by a conventionalistic rationality that can support conventional morality. This, of course, is Gauthier's own view.

2. David Gauthier, "The Unity of Reason: A Subversive Reinterpretation of Kant," in this issue; further citations are in parentheses in the text.

Ethics 96 (October 1985): 89–99

Bayesian." Here Gauthier's exegetical claims are less ambitious. Although he thinks that Hume was a contractarian, Gauthier does not pretend to show that Kant actually held his own favored theory of practical reason. But if his exegetical goal is less ambitious, his greater philosophical aim is surely more so. It is nothing less than to lead us "to a truer understanding of the role of reason in ethics" (p. 74). By demonstrating that the core elements of Kant's view lead to the conclusion that it is rational to maximize individual utility except when constrained by agreements that produce optimal outcomes, Gauthier hopes to undermine opposition to that conclusion based on other Kantian elements that conflict with the core doctrine. After all, he hopes his Kantian opponents will think, If that is what the fundamentals of Kant's view lead to, then why shouldn't I believe it also?

Faced with this strategy, how is a Kantian-inclined commentator to respond? One possibility is to fight fire with fire: attempt to reinterpret Gauthier on the same principles he aims to reinterpret Kant. Perhaps it can be shown that the fundamentals of Gauthier's own view lead not to a constrained maximizing theory of rationality but to some other surprising conclusion. Thoughts of such titles as "David Gauthier: Intuitionist" are tempting indeed.

What I propose to do instead is to consider Gauthier's treatment of Kantian themes itself with some care, since I believe that though it contains important insights it is open to criticism on a number of points. Like Gauthier, my aim will not primarily be exegetical. Since he does not claim that Kant held the view he attempts to derive from fundamental Kantian theses, I will not primarily be concerned to show that Kant did not. Rather, I shall argue that there are in Kant's position the resources for a view that neither conflicts with the fundamental Kantian themes Gauthier turns to his own use nor is beset by difficulties that riddle Gauthier's own attempt to demonstrate a relationship between rationality and morality.

There are three main claims in Gauthier's essay that I wish to discuss. First, that an account of practical reason paralleling Kant's treatment of theoretical reason would include both (*a*) the doctrine that pure reason cannot give rise to rational action by itself, any more than speculative reason can give rise to knowledge by itself, and (*b*) the doctrine that pure concepts of the will structuring the manifold of desire are conditions of the possibility of rational action in the same way that the pure concepts of the understanding structuring the manifold of intuition are conditions of the possibility of experience. Included in this second part is the suggestion that "the familiar ideas of the theory of rational choice correspond to the pure concepts of the will" (p. 79).

Second, Gauthier argues that practical reason's role in structuring the manifold of desire makes apparent the proper status of the desire for happiness. The desire for happiness is not simply given in or among our various desires, rather it is constructed a priori as the condition for the possibility of rational action. It is a desire that any rational but finite creature necessarily has. It follows that there is a practical requirement

that we seek our own happiness, Kant's official doctrine notwithstanding. Although Kant found it "astonishing how intelligent men have thought of proclaiming as a universal practical law the desire for happiness,"[3] this is in fact the conclusion dictated by some of his most central doctrines.

Third, and finally, Gauthier suggests an account of morality that is consistent with this picture of practical reason and that preserves Kant's ideas that (*a*) there is a "relation between morality and rationality, so that the ground of moral obligation must be found in the pure concepts of practical reason" (p. 81), and (*b*) the happiness of the moral agent cannot be the basis of moral motivation. This account is not developed very far, but we may assume that it involves the idea of constraints on maximizing that are the result of a rational bargain between individuals who are fully informed, where 'rational' is to be understood in terms of the formal theory of decision.

My plan is to discuss each of these claims in turn. I shall argue that, while Gauthier is right to stress that Kant's claims regarding reason in its theoretical employment in the first *Critique* have consequences for the treatment of practical reason, those insights do not provide any special support for a Gauthierian conception of reason over a view that, like Kant's official doctrine, ties practical reason directly to the Categorical Imperative. Moreover, I shall suggest that Gauthier's own picture of the "role of reason in ethics" is not entirely satisfactory.

Let us begin with Gauthier's suggestions for an account of practical reason that parallels Kant's account of theoretical reason. Again, there are two main elements: (*a*) Just as experience sets the limits to knowledge, so the manifold of desire sets the limits to rational action. Reason is incapable, by itself, of any action as it is of any knowledge. While Kant says that "pure reason of itself can be and really is practical" (*KpV,* p. 121), that conflicts with the "more critical" view of reason presented in the first *Critique.* (*b*) Just as there are conditions of the possibility of experience, namely, the transcendental unity of apperception and the categories, so there are conditions for the possibility of action, a unity of practical consciousness and the synthesis of the manifold of desire through the pure concepts of the will.[4]

Consider *a* first. The idea that desire sets the limits to action, as experience does to knowledge, is crucially ambiguous as between the two following propositions:

> i) Reason cannot determine the will entirely by itself without there also being a manifold of desire. (That is, rational agency

3. Kant, *Critique of Practical Reason,* trans. L. W. Beck (Indianapolis: Bobbs-Merrill Co., 1956), p. 27, Ak. p. 28. Further references to this work will be abbreviated (*KpV*) and placed parenthetically in the text. The page numbers will be those of the Preussische Akademie (Ak.) edition.

4. Similar suggestions can be found in Stephen L. Darwall, *Impartial Reason* (Ithaca, N.Y.: Cornell University Press, 1983), pp. 111–12; and in Wright Neely, "Freedom and Desire," *Philosophical Review* 83 (1974): 32–54, esp. p. 39.

requires not just reason but also what Kant called "the faculty of desire" or "the appetitive faculty"—what he defines in the *Metaphysics of Morals* as "the capacity to be by means of one's representations the cause of the object of these representations.")[5]

ii) Reason cannot determine the will except as the instrument of desire. (Reason plays no motivational role itself even when it is combined with the faculty of desire. It is, as Hume said, "the slave of the passions.")

To appreciate the difference between these two formulations consider a parallel distinction that Kant draws at the beginning of the first *Critique*. "There can be no doubt," he writes, "that all our knowledge begins with experience. . . . But though all our knowledge begins with experience, it does not follow that it all arises out of experience."[6] The distinction, of course, is that while, for Kant, there can be no knowledge of something that is not a possible object of experience, and so all knowledge begins with experience, not all knowledge is based on or "arises out of" experience. For some knowledge, Kant thinks, is synthetic a priori; its grounds are found not in experience but in conditions for the possibility of experience.

Analogously, from the proposition that action "begins with" and so cannot occur without a manifold of desire, it does not follow that all action "arises out of" or is based on desire in the sense that the grounds for any action, for any determination of the will, must be found in desires given in the manifold. For there is the analogous possibility to Kant's position regarding theoretical reason, one that he plainly also accepts with respect to practical reason. Reason can determine the will a priori, by providing the form necessary for the possibility of rational action, just as reason can ground certain beliefs a priori by providing the form necessary for the possibility of experience. It is thus open to Kant to believe that, while a manifold of desire is necessary for rational action, and hence reason cannot determine the will entirely by itself without any manifold of desire, still, given a manifold of desire, reason can determine the will a priori, by providing structure for the manifold necessary for rational action.

This way of describing things should remind us of Gauthier's own remarks about the constructive role of concepts of pure will as conditions of the possibility of action. On his account the desire for happiness itself is an a priori determination of the will by practical reason that is not given to us as part of the manifold of desire. Of course, this determination of the will of a rational agent could not be due entirely to reason. It must "begin" with a manifold of desire. No creature without a manifold of desire could be determined by her practical reason to seek her own happiness.

5. Kant, *The Metaphysics of Morals,* trans. J. Ellington, in *The Metaphysical Principles of Virtue* (Indianapolis: Bobbs-Merrill Co., 1964), p. 9, Ak. p. 211.

6. Kant, *Critique of Pure Reason,* B1, trans. N. K. Smith (London: Macmillan, 1964), p. 41. Further references will be abbreviated (*KrV*) and placed parenthetically in the text.

Kant's account of practical reason would exceed the strictures of the first *Critique* if he were to hold that practical reason could determine the will completely independently of the manifold of desire. "In both its speculative and its practical employment," he writes in the "Dialectic of Pure Practical Reason" of the second *Critique,* "pure reason always has its dialectic, for it demands the absolute totality of conditions for a given conditioned thing, and this can be reached only in things-in-themselves" (*KpV,* p. 107). This leads to the illusions, on the one hand, that speculative knowledge is possible through pure reason independent of experience and, on the other, that knowledge of the highest good, and consequent determination of the will, is possible through pure reason independent of the faculty of desire. But both, Kant holds, are impossible. Reason, by itself, is unable to furnish us either with knowledge of things in themselves or a knowledge of the highest good that could determine the will independent of the faculty of desire. Kant's Copernican Revolution in practical philosophy is as opposed to nonempirical rationalism as it is to nonrational empiricism.

It is significant in this connection that, when it comes to discussing practical laws, the only alternative Kant takes seriously to a practical principle's containing "determining grounds of the will because of [its] form and not because of [its] matter" (*KpV,* p. 27), is a principle's being grounded in some object of the faculty of desire. There is also the theoretical possibility that a practical law could be grounded in its matter, but not by its matter being the object of desire; rather, through its being directly discerned by reason as part of the highest good. But Kant rejects this possibility and for the reasons just mentioned.

What Kant believes pure practical reason does give rise to is the Categorical Imperative, what he calls in the second *Critique* the "Fundamental Law of Practical Reason." But far from proposing that this principle could direct the will without a faculty of desire, its very formulation makes clear its dependence on the existence of a manifold of desire: "So act that the maxim of your will could always hold at the same time as a principle establishing universal law" (*KpV,* p. 31). This is a principle for regulating action on maxims. Accordingly, it could only apply to a creature who is already inclined to pursue certain ends on certain conditions, to a being whose representations of objects are capable of motivating him to act, who has "by means of [his] representations" the capacity to cause "the objects of those representations"—a faculty of desire.

The Kantian thesis that reason can be practical through its capacity to regulate action on maxims by its Fundamental Law, is entirely compatible, therefore, with the strictures on reason argued for in the first *Critique.* Reason gives us, Kant maintains, not the matter of any practical law directly and independently of the manifold of desire. Rather it provides the form of a manifold of desire that makes rational action possible.

Now we can locate the real issue between Gauthier and Kant. The issue is not whether the Kantian doctrine that practical reason can de-

termine the will in accordance with its Fundamental Law, the Categorical Imperative, violates strictures on reason required by Kant's critical philosophy. It does not if acknowledging that Fundamental Law is a necessary condition of rational action. But whether that is so remains at issue.

Let us turn now to consider the second half of Gauthier's analogy with the first *Critique,* that action requires a structuring of the manifold of desire by concepts of the pure will corresponding to "the familiar ideas of the theory of rational choice" (p. 79). Rational action involves choice. "But a series of choices may not reveal a single preferential ordering of alternative possible actions. Taken together, they may not express the unified desires of an individual rational actor" (p. 79). Just as knowledge presupposes the possibility of unifying the manifold of intuition in a single consciousness, so rational action requires that the manifold of desire be unified in the consciousness of a single agent. There must be a transcendental unity of practical apperception and a synthesis of the manifold of desire through pure concepts of the will, "the familiar ideas of the theory of rational choice."

I agree that rational action requires an agent to be able to rank alternatives in a preferential ordering. Aristotle put the point this way, "Whether this or that shall be enacted is already a task requiring calculation; and there must be a single standard to measure by, for that is pursued which is *greater.* It follows that what acts in this way must be able to make a unity out of several images."[7]

Now Gauthier's proposal is that the way practical reason establishes a preferential ordering of alternatives is by ordering the manifold of desire. To provide a basis for rational action desires cannot simply be given in a manifold, rather they must be structured by practical reason to determine an ordering of alternative actions. But how is the ordering to be achieved? More to the point, is there any way for practical reason to structure the manifold of desire to determine an ordering without at the same time providing a standpoint from which specific desires in the manifold can be critically assessed? Gauthier does not discuss this matter, but it is important. For if desires cannot be rationally ordered without critical revision of specific desires, then the a priori desire presupposed by rational agency is not the desire for happiness, if that is the desire for the satisfaction of our given desires, but the desire for the satisfaction of critically revised desires. Note, by the way, that it is consistent with this second possibility that only an agent with a faculty of desire, with, that is, the capacity to be motivated by her representations of objects, could be capable of critically revising desires.

To see a particular instance of this problem consider how we are to describe a fully Gauthierian agent, that is, an agent who accepts Gauthier's "constrained maximizing" account of practical reason. Such an agent

7. Aristotle, *De anima,* trans. J. A. Smith, 434a 7–10, in *The Basic Works of Aristotle,* ed. Richard McKeon (New York: Random House, 1968), p. 600.

believes it overridingly rational for her to respect agreements in Prisoner's Dilemma situations even though alternative acts would maximize her happiness. Her choice is to respect the agreement even though the choice conflicts with what she desires most. Now if we take seriously Gauthier's Kantian remarks about the unity of practical consciousness, it would appear that there is a rationally disabling fissure in such an agent. Her structuring of her desires, necessary a priori for the possibility of rational action, leads to a best alternative that is in conflict with what she chooses and indeed regards to be the best choice given her Gauthierian conception of rationality. Something seems to have gone wrong.

I am not suggesting that a Gauthierian agent must be incapable of unifying her desires, rather, that she can be so capable only by critically revising her desires. In particular, she must be capable of critically revising her desires for outcomes given that they can only be achieved by violating optimal agreements. For on the theory of rationality she accepts she has reason to choose acts that maximize her happiness only if they do not violate such agreements. Only if she can critically revise her desires, then, will her unified manifold of desire not be at odds with her choices.

We can further appreciate this point by pursuing the analogy with the first *Critique*. The transcendental unity of apperception, the capacity to attach the "I think" to each representation, requires that experience be of objects, and hence, that a subjective/objective distinction be able to be drawn within experience. In order to see our experiences as ours we must be able to distinguish between the way things merely seem to us and the way they really are. If we cannot distinguish a perspectival or subjective element within experience itself, we will be totally and irretrievably absorbed in it in a way that makes it impossible potentially to step back from any experience to regard it as our own.

The same points apply in the practical realm. Beings who are totally and irretrievably absorbed in desires as they are given lack the capacity to unify them in one consciousness and to regard them as their own. To do that they must be able to distance themselves somehow from individual desires as they are given. The argument of the first *Critique* is that experiences can be unified in one consciousness, and as the experiences of one consciousness, only if individual experiences are represented to be corrigible in the light of each other: "If each representation were completely foreign to every other, standing apart in isolation, no such thing as knowledge would ever arise. For knowledge is [essentially] a whole in which representations stand compared and connected" (*KrV,* A 97). The analogous claim in the practical realm is that the manifold of desire can be unified in one consciousness only if individual desires are represented as corrigible in the light of others. But what could this mean? How can one desire be corrigible in the light of another? It plainly cannot mean simply that one desire is stronger. For that is possible even if each desire is "completely foreign to every other," and desires are in no way unified in thought.

One way in which a desire can be corrigible by another is by being a more or less considered desire. On representing to oneself the thought of a child's pleasure at having a toy, one desires to give it to him. But on realizing that the pleasure will consist in taunting his sibling, one desires not to. The second desire is not "foreign" to the first. It does not win out simply because it is a stronger desire. Rather, it wins out because its object is regarded to be more adequately represented than is the first. The first desire does not survive critical reflection.

Or consider a case in which I must choose between conflicting desires. I want to drink and to that extent scan the environment for potables, regarding anything else as a distraction or obstacle. I want to see the last inning of the ball game and to that extent am focused on the diamond and regard anything else as a distraction or obstacle. Rationally to choose between my two desires I must have available some standpoint that is not totally absorbed in the object of either. I must be able to represent both alternatives to myself and consider which to choose. But once I am not totally absorbed in either desire I am free to consider other aspects of their objects, or indeed, perhaps, of other objects. The standpoint necessary to choose between given desires is one from which each may be critically revised.

If we pursue the analogy with the first *Critique,* then, what emerges is that we can synthesize a unity of the manifold of desire only if we conceive of individual desires as corrigible, as more or less justified.[8] It follows that the a priori desire lying behind the unity of practical consciousness is not the desire for happiness, if that is the desire for satisfaction of given desires, but the desire for satisfaction of critically revised desires. Again, this is quite compatible with the internalist thesis that rational correction of desires can itself proceed only by engaging motivational susceptibilities. On the suggested Kantian picture, we rationally choose not simply by engaging our various desires as given but, rather, by engaging desires we would have were we to represent things to ourselves in various ways. And this, it is important to point out, is compatible with taking the Categorical Imperative to be the Fundamental Law of Pure Practical Reason. For correction of desires by this principle is correction that engages motivational susceptibilities, since it engages what one could will from a standpoint that we can in fact adopt. A Kantian agent can be governed by the Fundamental Law without threatening her unity as an agent, since the unity she constructs within her manifold of desire is a unity of critically revised desires.

The results of our discussion of the second half of Gauthier's first thesis threaten his second thesis: that the synthetic a priori desire for happiness generates a practical law that one seek happiness. If the argument above is correct, then the object of the relevant desire is not the satisfaction of our given desires, but the satisfaction of justified desires, of desires

8. I argue this at greater length in *Impartial Reason,* pp. 85–100, 108–11.

that would not be corrected by further rational representation or reflection. Nonetheless, this is not an insignificant result. It supports the thesis that there is a practical law that we do what would maximize the satisfaction of our critically revised desires, other things equal. Now, if Kant is right, part of what is involved in critical revision of the manifold of desire is submitting maxims to the test of the Categorical Imperative. That, of course, would have to be established by independent argument. Even so, it seems that an argument analogous to that of the first *Critique* supports the proposition that it is rational, other things equal, for agents to maximize the satisfaction of desires that would survive critical reflection, leaving it open whether critical reflection includes the Categorical Imperative test.

Gauthier's third thesis is that there is an account of morality, consistent with the view of practical reason he has sketched, that both preserves Kant's ideas that "the ground of moral obligation must be found in the pure concepts of practical reason" and that the happiness of the moral agent cannot be the basis of moral motivation.

Rational agents have need of morality because there are situations of interaction, illustrated by the Prisoner's Dilemma, where if each agent acts on the practical law requiring him to maximize his own happiness, then the outcome will be suboptimal for each. What is required is a new practical law, the moral law, presumably, that both constrains a person's maximizing "when she is among like-minded persons in PD situations" and "ensure[s] that the person who conforms to it may expect to maximize her happiness" (p. 86).

We can readily see why a Kantian who accepted the Categorical Imperative as practical reason's Fundamental Law would be led by reflecting on such situations to regulate the pursuit of outcomes she prefers independently. And indeed this is the argument of Kant's against "proclaiming as a universal practical law the desire for happiness," to which Gauthier refers: "If one . . . attributed the universality of law to this maxim, there would be . . . the most arrant conflict, and the complete annihilation of the maxim itself and its purpose" (*KpV,* p. 28). This is no more than to say that the maxim of unconstrained promotion of personal happiness conflicts with the Categorical Imperative since one could not rationally will that everyone act on it "at the same time" that one wills that one act on the maxim itself. In PD situations a contradiction in the will results—one wills both that the optimal and that the suboptimal outcome be promoted. The argument here is essentially that of Kant's fourth example regarding the duty of beneficence.

One of Gauthier's formulations suggests a similar argument. "We require first of all," he writes, "a principle that, suitably generalized, yields outcomes satisfying the second, optimality condition and does so in such a way that, *in any situation in which all persons follow the principle,* each may expect to do better than if all followed a maximizing principle satisfying the equilibrium condition" (p. 84; emphasis added). This suggests that

in assessing whether a principle really is a practical law we need to consider what it would be like if "all persons follow the principle." Now it is clear why that is relevant to a theory of practical reason that takes its Fundamental Law to be the Categorical Imperative. But it is quite unclear why it should be relevant to a theory of practical reason whose "fundamental law" is that each is to maximize his own happiness. Why should one care about what would occur if each were to maximize his own happiness? How is that relevant to how one should act?

Gauthier's answer begins with the idea that "an ideally rational actor is one who chooses in such a way that he maximizes the satisfaction of his desires" (p. 85). This is not the same thing as choosing *to* maximize his happiness and, in certain circumstances, it will lead to quite different results. Most significantly, in PD situations in which each participant knows on which principle each is choosing, it will maximize an agent's happiness not to choose on the maximizing principle. If he does maximize, others will know he is not cooperating in producing what would be an optimal outcome for each, and they will not try vainly to cooperate with him. Unless they are suckers they will also maximize and all will be worse off. The moral is: if others know how one is choosing, choosing to maximize in PD situations will be self-defeating.

Armed with these reflections, a rational agent will see that, according to the maximizing principle, he will do better in such situations to choose on some other principle, presumably moral in character, that requires agents to respect optimal agreements. If he is prepared to constrain his maximizing by this moral principle, he is likelier to be better placed actually to bring about his concerns.

This, I believe, is the substance of Gauthier's case for an account of morality that, like Kant's, preserves the relation between morality and rationality and maintains "that the happiness of the actor cannot be the basis of moral motivation." Indeed, it aims to show the connection between these two aspects. If the happiness of the actor were the basis of moral motivation, then it could not provide a rational solution to problems of human interaction typified by the Prisoner's Dilemma.

Plainly this argument for the rationality of being morally motivated, of being governed by principles of choice that constrain maximizing in PD situations, depends crucially on the assumption that others will know on what principles a person is choosing. In situations in which others believe that one is a cooperator, one will do better to choose on the maximizing principle than on the constraining principle. Gauthier writes, "Now if she is usually able to form correct expectations about others, and to enable them to form correct expectations about herself" (p. 85), the agent who constrains her maximizing by the appropriate moral principle may expect to do better in PD situations. But why should she enable others to form correct expectations about herself, if she can do even better by pretense? Of course there are many things that can be said here regarding stability of agreements, the relation between honesty with

oneself and others, the connection between honesty and relationships we value, the psychological cost of dishonesty, our social nature, and so on. An argument that an agent is more likely actually to maximize her happiness by constraining choices in PD situations by a moral principle could not hope to be convincing, I think, without a judicious weighing of these complex empirical matters. And there would be much to weigh on the other side. While a maximizing agent will do worse in PD situations when others know her character, there may still be ways available to her to transform the situation from a PD situation into something else, for example, by offering a deposit as security of her "good will."

Whether a conception of morality as providing an optimal solution to problems of human interaction encountered in situations typified by the Prisoner's Dilemma can be linked to a fundamentally maximizing conception of rationality, consequently, depends on very complicated empirical issues. Moreover, it is quite possible that no "universal" result can be established. While for many it might maximize happiness to choose on moral principles in PD situations, this may not be so for all. In any case, this matter cannot be decided by considering on what principles one would best choose if everyone knew on what principles each was choosing. Indeed, that question is tantalizingly close to Kant's question (On what principles would you choose to act if those principles were also to govern everyone else's conduct?), since in PD situations others will choose to maximize if they think one will and will choose to constrain themselves only if they believe that one will also. To the extent that Gauthier relies on the assumption that others are perfectly informed about each others' principles of choice,[9] he is making an idealization that makes more sense on Kant's view than on his own.

This leads to the pleasant thought that Gauthier might himself be represented as committed fundamentally to a Kantian approach that conflicts with his individual utility-maximizing conception of practical reason—that *he* could be reinterpreted as a Kantian moralist. However, since reinterpretion is perhaps safer and more seemly with "The Great Dead Philosophers," I shall demur.

9. As he does in David Gauthier, "Reason and Maximization," *Canadian Journal of Philosophy* 4 (1975): 411–33.