



Sacred Heart  
UNIVERSITY

Sacred Heart University  
DigitalCommons@SHU

---

Philosophy, Theology and Religious Studies  
Faculty Publications

Philosophy, Theology and Religious Studies

---

Spring 1991

## Kant's Dubious Disciples: Hare and Rawls

Edward Papa  
*Sacred Heart University*

Follow this and additional works at: [https://digitalcommons.sacredheart.edu/rel\\_fac](https://digitalcommons.sacredheart.edu/rel_fac)



Part of the [Philosophy Commons](#)

---

### Recommended Citation

Papa, E. (1991). Kant's dubious disciples: Hare and Rawls. *American Catholic Philosophical Quarterly*, 65(2): 159-175. doi: 10.5840/acpq199165224

This Peer-Reviewed Article is brought to you for free and open access by the Philosophy, Theology and Religious Studies at DigitalCommons@SHU. It has been accepted for inclusion in Philosophy, Theology and Religious Studies Faculty Publications by an authorized administrator of DigitalCommons@SHU. For more information, please contact [ferribyp@sacredheart.edu](mailto:ferribyp@sacredheart.edu), [lysobeyb@sacredheart.edu](mailto:lysobeyb@sacredheart.edu).

## Kant's Dubious Disciples: Hare and Rawls

by Edward Papa

---

The aim of this paper is to show that Kant's *Groundwork of the Metaphysics of Morals* has set a basic pattern of research for moral philosophy, a pattern established by the structure of the text, and one which implies an overarching thesis. The *Groundwork* is divided into three sections: (1) an analysis of ordinary moral judgments which identifies their formal rule, (2) the derivation of the supreme principle of morality, and (3) the presentation of a theory of reason which grounds the possibility of moral action. Kant seeks to explain the sense in which moral action may be said to manifest transcendental freedom: regardless of what may be the case in the world investigated by science, it is, Kant argues, both theoretically possible and practically necessary to affirm the idea of freedom.

The question of the manner in which contemporary moral philosophers such as Hare and Rawls continue the Kantian project has been much disputed. But philosophers identify themselves as Kantians, but the meaning of this identification is obscure since they are opposed in so many seemingly fundamental respects. Rawls is a deontologist, Hare a defender of utilitarianism. Rawls employs a basically empiricistic method; Hare denies that such methods are adequate for moral theory. What meaning could the Kantian epithet have beyond these determinations? The answer, I will suggest, is that it signifies a model of theory-construction patterned after Kant's *Groundwork*. Kant's *Groundwork* represents an ordered series of questions. These questions direct the human being to discover the grounds of moral determination within. By contrasting in some manner the "within" of the moral standpoint from the "without" of the standpoint of the world, Hare and Rawls retain something of Kant's notion of transcendental freedom. Yet neither conceives transcendental freedom in the metaphysical manner of Kant as signifying a domain which is in principle closed to empirical inquiry. For Hare and Rawls the moral standpoint is a purely formal determination, not in Kant's sense of constituting grounds for moral action regardless of the state of the world, but merely in the sense of constituting a pattern of thought that can be employed regardless of the state of the world.

After establishing the sense in which Hare and Rawls are properly classified as Kantians both with respect to the structure and general aims of their theories, the differences between them are specified in terms of the way in which they respond to charges of formalism. It is then argued that their theories fail to support the core Kantian doctrine of the autonomy of the moral subject. Finally, proposals are made concerning the requirements for the success of the *Groundwork* project, namely, how a positive conception of the autonomous subject may be maintained even while this subject is understood as rooted in the world.

## I

Kant's ostensive aim in the *Groundwork* is to vindicate what he takes to be the common belief in the objectivity of the moral judgment. The three parts of the *Groundwork* are ordered toward this objective. With the completion of the third part, Kant shows how morality, properly understood, can be rationally justified.

Part one of the *Groundwork* begins with an analysis of the moral judgment and is complete with the identification of the categorical imperative as the formal condition of moral reasoning. Since the categorical imperative is the formal structure that judgments or, more generally, acts must exhibit to be moral, to justify moral judgments and actions is to show that their principles, or maxims, in Kant's language, possess this form. The objectivity of moral principles consists in their conformity to laws grasped as necessary by reason. An objective moral principle is one that is valid for rational nature as such. The justification of moral principles consists in showing that acting on them is demanded by reason (*G* 400 fn.).<sup>1</sup> Since human beings are not perfectly rational, we can conclude that an act is required by reason *per se* rather than by another, perhaps antagonistic, motive only if it is done for duty's sake. In all other cases the act may in fact be ordered to the attainment of some end constituted by man's natural impulses or desires (*G*, 397-99). Kant's analysis of the moral judgment links the notion of acting for the sake of duty with acting on maxims that can be willed to be universal law, and both are understood in terms of the will's being abstracted from all motivation that lies outside its conformity with the requirements of reason.

But to specify moral action as action done for the sake of duty is not to reveal the full content of the moral law. The task of the second section of

<sup>1</sup> All page references to the *Groundwork* will be made in the body of the paper, with the *Groundwork* abbreviated by "G," and follow the standard pagination of the Berlin Academy edition. The translations are those of J. J. Paton (New York: Harper and Row, 1964). The primary texts of Hare and Rawls I also refer to in the body of the paper. These are Hare's *Moral Thinking* ("MT") (Oxford: Clarendon Press, 1981) and Rawls's *Theory of Justice* ("TJ") (Cambridge, Mass.: Harvard University Press, 1971). All other references will be found in the notes.

the *Groundwork* is to articulate the moral law as a principle of autonomy. Kant conceives duty as a requirement of reason imposed upon recalcitrant sensible nature (*G* 404-05). Sensible nature tempts us to restrict the function of reason in informing action to hypothetical or means-end thinking and to pursue ends constituted by sources external to the will. In testing maxims for their conformity to the moral law, Kant relies most heavily on the formulation of the categorical imperative which demands that a maxim must be capable of serving as a law of nature (*G*, 421). A maxim can fail the test of universal law in two ways: either (1) its universalization is self-contradictory and we cannot conceive of a system of nature which contains it as law; or (2) while self-consistent in the sense of conceivable within a system of nature, its universalization cannot be rationally willed. Since conformity to rational nature is the rule of moral action, and since Kant identifies rational nature with the autonomous will, the final implication to be drawn from the categorical imperative is "the Idea of the will of every rational being as a will which makes universal law" (*G*, 431).

Although as an imperative the moral law appears external to our imperfectly rational nature, it is in fact the issue of our reason. The idea that every rational being, as rational, is the author of universal law follows directly from the categorical imperative. Imperatives are either hypothetical or categorical. Since all hypothetical imperatives determine the will heteronomously, categorical imperatives must be authored by the will itself. Autonomy, Kant concludes, is "the sole principle of ethics" (*G*, 440). Since the task of ethics is to determine what the moral law requires, this amounts to determining the requirements of autonomous reason.

Still, Kant cautions against drawing substantive conclusions from an analysis of the meaning of morality. While the actual respects in which we are subject to sensible determination are properly items of scientific knowledge, the basis on which we may assert the autonomy of the will is unclear. A philosophical analysis of the meaning of morality does not establish that moral conduct is a real possibility for human life (cf. *G*, 445). To establish this possibility is the task of the third section of the *Groundwork*. Kant seeks to show that moral conduct is possible on the basis of a theory of reason which accounts for the fact that human nature is both sensibly conditioned and practically free.

Kant has shown that the concept of the moral law includes the idea of self-authorship. Furthermore, this implies, Kant argues, that the idea of freedom must ground the concept of autonomy. By definition, an autonomous will is free in a negative sense, free from "determination by alien causes" (*G*, 446). But reason has insight, Kant claims, into its own activity in a positive sense as well, since it must consider itself to be the author of the moral law. The will is thus perceived as a law unto itself, but only when it avoids heteronomous determination and therefore when its own structuring principle is universal law. Kant therefore concludes that "a free will and a will under moral laws are one and the same" (*G*, 447).

This is not, of course, to deny that we are also conditioned by sense. But Kant argues that the standpoint by which we conceive ourselves as free can

and must be distinguished from the standpoint we adopt when we view our actions merely as observed events (*G*, 450-52). As imperfectly rational agents, we are participants in two worlds: we belong to a purely intelligible world through our power of reason, a power capable of producing ideas transcending anything which could be contributed to the understanding by sense, and we belong to the sensible world in which we are determined by natural causes. Insofar as we exist under the laws of nature we are heteronomously determined, but insofar as we also belong to the intelligible world, and so exist under laws which "have their ground in reason alone," we have grounds to assert our freedom (*G*, 452). Since we are neither perfectly rational nor completely determined by sense, we "recognize," as Kant says, "the autonomy of the will together with its consequence—morality," and we recognize autonomy to be an "obligation" which we do not, as sensible creatures, by nature fulfill (*G*, 453). Our dual determination, then, shows freedom to be possible. For the purposes of moral action, this dual determination provides sufficient reason to affirm the principle of autonomy.

Once the theoretical possibility of freedom is vindicated, Kant's final point is not only that we make this possibility real in moral action but that it is necessary to do so if we wish to give evidence of our rational nature (*G*, 461). Although Kant does not offer his argument as knowledge of the intelligible world of reason (knowledge for us must be confined to the sensible world), nonetheless, insofar as we adopt the moral point of view we affirm our freedom by conceiving our wills as obedient to laws having their ground in reason alone. The positive freedom implicit in our authorship of the moral law justifies our belief in our autonomy. Autonomy, thus conceived, is not part of a description of the self but a principle necessarily affirmed in moral conduct.

## II

Hare's moral philosophy provides a linguistic foundation for Kant's claim that moral conduct evidences transcendental freedom. The three-stage argumentative sequence constitutive of Kant's *Groundwork* is pursued in Hare's trilogy of moral philosophy: *The Language of Morals*, *Freedom and Reason*, and *Moral Thinking*. Each work emphasizes one of the parts of the *Groundwork*. Like Kant, Hare begins with an analysis of the moral judgment in order to identify its formal rule. Hare argues that while moral judgments are not independent of the preferences of those who formulate them, they nonetheless have cognitive content. The use of moral concepts to express our preferences commits us to specific lines of conduct.

The logic of moral concepts is summed up in the requirement that moral judgments express universal prescriptions. Moral judgments are prescriptions insofar as their primary function is to guide choice rather than to merely describe what is the case. But moral judgments are not properly analyzed as imperatives since their purpose is to guide choice on the basis of object reasons. Objective reasons for making a particular moral choice

are contained in the principles which specify the relevant standard of value for preferences of a given kind. The objectivity of principles turns on their universalizability; principles constitute reasons for choosing something or acting in a particular way which purport to be valid for anyone in similar circumstances. Moral judgments not only express a preference for something, but commend that preference to others on the basis of standards which the other presumably acknowledges. With this analysis of the moral judgment, Hare believes himself to have articulated a basically Kantian conception of the moral judgment as irreducibly prescriptive and at the same time objective because universalizable (cf. *MT*, 4).

At the same time, Hare rejects Kant's contention that moral principles are categorical imperatives. Moral principles, for Hare, do not constitute objective reasons for action independently of those ends which can be attained through them. For Hare, states of affairs as well as actions may be subsumed under moral principles since these principles articulate reasons not merely for what should be done, but also for what rationally should be the case. Consequently, there is room in Hare's moral system for consequentialist moral principles, and indeed such principles take pride of place.

Although Hare insists that the analysis of moral terms be carried out independently of the question of the substantive principle of morality, he at the same time contends that utility is the first principle of morality on the basis of an analysis of moral terms. Utility occupies roughly the place in Hare's theory that Kant allocates to the categorical imperative. Like Kant, Hare believes that the identification of the supreme principle of morality is a matter of analysis. The employment of moral language, according to Hare, commits us to actions which maximize the satisfaction of desires of all persons affected equally—as Bentham put it, "Everyone to count for one, no one to count for more than one" (cf. *MT*, 4).

The core of Hare's method for normative ethics consists in the imaginative representation of the preferences of all persons affected by a given act. The difference between a prescription for myself in a particular situation and a moral judgment is a thought-process where the preferences of all those affected by the proposed judgment are considered one by one and weighed according to intensity of preference. After criticism by logic and the facts—the logic which directs us to consider the preferences of others equally with our own and the facts about those preferences—the objectively right moral judgment is that which tries to maximize the satisfaction of preferences. Since the requirement of universalizability entails that identical judgments be made about all cases identical in their universal properties (*MT*, 108), the standpoint of the individual who makes the judgment cannot be relevant in framing the judgment. Each person's preferences are relevant, but only in their universal features. Because universalizability precludes giving the perspective of the actual individual "I" more weight than the perspective of any other "I," the representation of the perspective of another is the hypothetical identification with the other as another "I." To represent another fully to myself, albeit in imagination, is to adopt the other's

motivations for myself (*MT*, 95). Hare does not believe that this involves, in principle, impossible comparisons. On the contrary, he claims that all intersubjective conflicts of preferences are resolvable in the same way in which a given individual settles his or her own conflicts. Just as anyone is able to rationally determine what he or she most prefers, despite conflicting leanings, so it is possible to pool the preferences of many persons through hypothetical identification and to weigh them in terms of intensity (*MT*, 110). Hare concludes that rational choice as determined by the logic of moral concepts is that of choice guided by the principle of utility.

Hare's contention that utilitarianism can be derived from the logic of moral words depends on his account of the moral viewpoint as a standpoint for critical thinking. Hare's moral theory is in this way structurally parallel to Kant's *Groundwork*. Just as Kant sought in the third part of the *Groundwork* to show that the principle of autonomy constitutes a ground for moral action rather than a purely logical possibility, so Hare's analysis of the moral point of view given in *Moral Thinking* is designed to show that the link between moral language and utilitarianism established in *Freedom and Reason* is not merely the result of one mode of analysis but reflects our actual moral thinking. Just as Kant was faced with the task of showing that moral autonomy is consistent with the myriad ways in which we are sensibly determined, so Hare is faced with the task of showing how utilitarianism is the basic requirement of moral thought, despite the fact that many of our moral sentiments are doubtlessly non-utilitarian and even anti-utilitarian. Finally, just as Kant succeeded by distinguishing between the standpoint of noumenal and phenomenal subjectivity, identifying morality with the former, so Hare distinguishes between two levels of moral thinking, the level of intuitive moral thought exemplified by most moral conduct and critical moral thought which is the product of thorough-going philosophical reflection.

Critical thinking, for Hare, is directed toward the question of justification. The question of justification generally arises when our intuitive moral principles give conflicting counsel in a given situation of choice (*MT*, 45). The only standpoint from which such a conflict can be resolved is that determined by the principle of utility. Hare conceives utilitarianism as an ideal of moral thinking represented by an "archangel" endowed with full knowledge of the facts, perfect powers of thought, and no human weaknesses (*MT*, 44). With perfect knowledge of the logical requirements of moral thinking and of the facts, the archangel will fully represent the perspectives of each person affected by a proposed act. Not being archangels, we have neither the knowledge of the facts nor the perfections of intellect and character which insure the archangel's success as a moral reasoner. For purposes of everyday moral choice, we are better off retaining *prima facie* principles which seem intuitively right. These principles are general yet clear enough to serve as guides for moral choice when time for reflection is not available or when character is weak (*MT*, 36-38). Since our moral education is based on *prima facie* principles, it is to be expected that many of our moral intuitions run counter to the principle of utility. There are,

nevertheless, Hare argues, good utilitarian reasons for continuing to cultivate such intuitions. On the whole, we are more likely to act in ways consistent with the method of critical thought by acting on intuitive principles: if we try to carry out critical thinking in the context of action, we are prone to rig the results so that they come out in our favor. When we are removed from the immediate context of action and have both the time and the dispassion to reflect clearly, we can better judge by means of critical thinking what the correct thing to do would be. Critical thinking always attempts to formulate principles for maximizing the satisfaction of preferences. While its principles can be entirely specific to the context of action when described in its universal features, given our human limits, the most valuable role critical thinking plays is in justifying *prima facie*, i.e., in determining which principles will on the whole satisfy the preferences of most of us when held with firm intuitive conviction.

The central argument of *Moral Thinking* has drawn a great deal of controversy. Hare argues that when reasoning is not flawed, all of us are constrained, on the basis of logic and the facts, to draw the same moral conclusions (*MT*, 6). A note of frustration at being misunderstood comes through clearly: "I shall be doing what will seem, to those who do not understand quite what is going on, to be very like retracting that view [i.e., that prescriptivity is always an element of moral thinking]" (*MT*, 6). Because the notions of preference and choice underlie Hare's concept of prescriptivity, Hare thinks his analysis has been misconstrued as a form of subjectivism. Hare notes that although it is often maintained as an objection to this theory of moral judgment that to be an objectivist one must hold that moral right is independent of the attitudes or inclinations of any particular speaker, it is, in fact, precisely in this sense that he is an objectivist (*MT*, 208 fn.). Given the logic of moral concepts, moral judgments must be universalizable; moral right, therefore, must be independent of the preferences of any individual. At the same time, given the requirement that moral right is prescriptive, it must essentially be conceived as a choice based on the preferences of all those affected; moral right can never, for Hare, be independent of preferences in general.

To insure that preference plays a constitutive role in moral thinking, despite its objectivistic character, Hare introduces two considerations about human freedom into his argument: first, it is possible to adopt an amoralist stance; and second, the concept of self itself contains an irreducibly prescriptive component. While perfect knowledge of the logic of moral words and of the facts compels universal assent to identical moral conclusions, this is the case only if one decides to engage in moral thinking at all. Logic does not compel us to acknowledge any universal prescriptions; amoralism is a real option (*MT*, 183). Moreover, the facts of primary relevance for moral judgments are themselves prescriptions rather than descriptive features of the context of moral action. To call someone "I" is to give particular weight to the satisfaction of that person's preferences (*MT*, 96-97). In the course of universalizing our prescriptions, we must hypothetically identify with another "I," and to do so entails that we prescribe that the other I's

preferences be satisfied. Since the word "I" is irreducibly prescriptive, Hare's central argument does not merely proceed from descriptive facts about the preferences of persons to prescriptive conclusions. The premisses concerning the preferences of persons themselves contain prescriptions.

Still, on Hare's critical method it appears that if we wish to act morally, we are only free to seek to maximize the satisfaction of preferences, and that if we are fully informed and rational, there will only be one way to do this. While conceding this conclusion, Hare insists that freedom is always present in the formation of preferences. Even if his arguments for the amoralist escape-route and the prescriptive nature of the "I" are rejected as unsound, we are in any case, Hare claims, "free to prefer what we prefer" (*MT*, 225). Apart from considerations of coherence, our rational choice at any particular time will only be "a function of our separate and perhaps conflicting preferences and their respective intensities and of nothing else" (*MT*, 225). Hare admits that universalizability does substantially curtail our freedom to prefer what we prefer in the sense that in moral thinking the preferences of others are given the same weight as our own; still, universalizability does not dictate what the content of preferences must be. Universalizability simply directs us to impartially maximize preferences. When we pursue the moral question, What ought I to do? we are ultimately referred back to those principles constitutive of our way of life, the principles which establish the limits of our awareness of our alternatives and, hence, the limits of freedom. For Hare, the limits of our ways of life are determined by reason, specifically by the interest of reason in justification. Freedom is not merely the possibility of choosing at random between alternatives; it is the awareness of the limits of the possibilities for action set by reason. The moral point of view for Hare, as well as for Kant, is the position of free persons who guide their actions on the basis of objective reasons. It is in the understanding of how reasons are justified as objective which separates them.

### III

Rawls develops an alternative to utilitarianism by reinterpreting Kant's concept of autonomy as the foundation of a theory of social justice. Like Hare, Rawls seeks to provide a method for resolving conflicts between our intuitive moral principles; unlike Hare, Rawls's method, far from maximizing given preferences, abstracts from these preferences to arrive at principles for regulating the social structures which give rise to preferences in the first place. Rawls's relation to Kant is complex. Rawls rejects Kant's attempt to ground the principle of autonomy in the structure of pure practical reason but gives the principle the central role allotted it in the *Groundwork*. Rawls puts the principle of autonomy to work in a basically empirical theory which, however removed from Kant's moral methodology, nonetheless follows the structure of Kant's *Groundwork*.

For Rawls, as for Kant, the examination of the moral judgment forms the point of departure for moral theory. The question of the truth or falsity

of the moral judgment, which for Kant motivates the philosophical inquiry into morality, is initially put into brackets by Rawls.<sup>2</sup> The question of moral truth, Rawls believes, has been long and fruitlessly pursued. Progress in resolving it, he suggests, depends on a more adequate investigation of the actual structure of our moral conceptions than has hitherto been undertaken. Once our moral framework has been explicated, we may return to the question of moral truth.<sup>3</sup> Rawls conceives this question in pragmatic fashion as a question of the adequacy of the moral framework as a whole to advance human interests. Unlike Hare, who takes the desirability of acting on moral principles for granted, Rawls takes radical critiques of our moral framework seriously, particularly those of Freud and Marx (*TJ*, 514, 489). To show that our moral conceptions are not malevolent, it is necessary to link the desire to act on moral principles with the advance of human interests conceived along the lines of what Mill called "utility in the largest sense, grounded in the permanent interests of man as a progressive being."<sup>4</sup> This link is established by means of an interpretation of the moral sentiments which shows their role in preserving a public conception of justice in a democratic society. Since the adequacy of any interpretation rests on its ability to account for the data, and since the data of moral theory consist in our moral judgments, the test of Rawls's moral theory is its adequacy in accounting for our moral judgments. Therefore, although Rawls, like Kant, starts with an investigation of the moral judgment, he is committed to viewing these judgments in such a manner that their truth-claim is bracketed.

Rawls's empiricistic method is so radically opposed to Kant's as to raise doubts about the propriety of Rawls's identification of his theory as "Kantian constructivism." Rawls's affinity to Kant is not methodological but structural, the structure being that of Kant's *Groundwork*. Rawls, like Kant in section one of the *Groundwork*, concludes his inquiry into the moral judgment with the identification of its formal rule. Like Kant in the second section, Rawls uses this formal rule to derive moral principles. Finally, following the pattern set in the third section of the *Groundwork*, Rawls grounds the moral point of view in an account of the structure of practical reason. Naturally, given Rawls's underlying empiricism, his account of practical reason, just as his accounts of the principles of morality and the rule of moral judgment, must be radically different from Kant's. Nonetheless, the employment of the *Groundwork* pattern guides Rawls's inquiry, justifying his characterization of his project as Kantian constructivism.

Kantian constructivism is explicitly posed by Rawls as an alternative to intuitionism, the doctrine which conceives moral judgments as implying

<sup>2</sup> John Rawls, "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association* 47 (1974-75): 7.

<sup>3</sup> *Ibid.*

<sup>4</sup> John Stuart Mill, *On Liberty* (Indianapolis, Indiana: Hackett Publishing Company, 1978), 10.

truth-claims of a unique sort. The problem of intuitionism is illustrated by Kant's inquiry in the first part of the *Groundwork*. There Kant treated the moral judgment as making a truth-claim that implies the requirement that one do one's duty for its own sake. This doctrine of what Rawls calls the purely conscientious act raises questions not only about the rationality of moral conduct but about its coherence since we are subject to a variety of duties which appear at times to commit us to conflicting lines of conduct. The consequence of this conflict, as Kant himself saw, was the conversion of moral requirements into mere hypotheses concerning the advance of non-moral ends and the seemingly hopeless dialectic of ordinary moral consciousness (*G*, 405). The transformation of moral requirements from universal prescriptions of prima facie assumptions concerning the general welfare is aptly illustrated by Hare. Rawls, by contrast, seeks to retain the strict character of moral precepts, while avoiding the dialectic between moral and non-moral considerations described by Kant. Rawls's solution is to conceive moral requirements as conditions for the choice of moral principles. These conditions will vary with the nature of the choice-problem, but they will remain the same insofar as they articulate a general conception about the nature of moral principles. Moral principles serve to resolve conflicts of interest between individuals. In order to serve this function, they must rest on conditions which could be agreed to in advance by the parties in the dispute, at least insofar as all are willing to be reasonable. The formal rule of moral judgment, for Rawls, is one of fairness. Moral principles are ones chosen under conditions which are fair, given the nature of the choice-problem at hand.

To identify fair conditions for the choice of moral principles, it is not sufficient to appeal to the moral precepts which we happen to have at hand since these may themselves reflect specific social interests. Rather, we must ultimately abstract to a situation in which the parties are ignorant not only of their social positions and natural endowments but of their place in history as well. The initial situation of fairness is reminiscent of the social contract of seventeenth- and eighteenth-century political thought, but as conceived by Rawls it is a pure hypothesis having to do with circumstances in which a fair agreement can be reached about moral principles. Autonomy is central to Rawls's account of these circumstances. Assuming, as Rawls does, that the identity of persons is indeterminable apart from their place in a given social structure, moral autonomy, if it were possible, would consist in detaching oneself from all knowledge apart from that necessary to affirm principles of justice. Autonomy represents the standpoint taken by any individual who seeks to regulate the social structure by moral principles collectively agreed upon.

The bulk of *A Theory of Justice* consists in Rawls's interpretation of the initial situation, his account of the principles that would be adopted within it, and his proof that the principles adopted are capable of regulating the social structure of a modern, industrial democratic state. The primacy of the question of social justice reflects Rawls's belief in the social nature of persons. The interests of persons as well as their characters are, Rawls

writes, determined by the institutional agreements within which they find themselves.<sup>5</sup> What we would be apart from our social institutions is, Rawls claims, unknown and probably unknowable.<sup>6</sup> Although Rawls's conception of the person would not seem to leave much room for private morality, Rawls contends that the same strategy by which principles of social justice are derived may be adapted to the problem of determining principles for individuals. Although it is not necessary for the purposes of this paper to examine Rawls's specific derivation of moral principles, it is important to note that Rawls provides a comprehensive account of moral reality, and we may inquire into the nature of that reality.

Rawls's employment of a pure hypothesis as the means for deriving moral principles gives his theory an almost fictional character. What interest, apart from the literary, can I take in the imaginary deliberations of hypothetical subjects in a highly contrived situation of choice? The problem is similar to that which Kant faced in the beginning of the third section of the *Groundwork*. Granted that our moral deliberations only make sense if they are conceived as the product of a self-regulating will, why should I believe that I have such a will? Kant's answer to this question is to explicate the moral standpoint as the perspective of noumenal subjectivity. But this solution is rejected by Rawls on both theoretical and practical grounds. Moral theory, Rawls holds, should not lead us to posit entities other than those required to explicate the data. Moreover, the coherence of moral theory with other sciences of the human being prohibits the employment of radically distinct methods of inquiry. Finally, Kant's distinction of standpoints fails to explain how the self-legislating activity of the saint is to be distinguished from that of the scoundrel since both would seem to equally express noumenal subjectivity (*TJ*, 254-55).

Rawls's critique of Kant allows us to highlight his own account of the grounds of morality. For Rawls, as for Kant, morality is a pure posit. But whereas Kant conceives this posit in terms of the self-legislating activity of the individual moral agent, Rawls conceives it in terms of the activity of moral theory-construction. The interest of Rawls's theory lies in the assumption that through moral theory-construction we can both overcome the limits of the social morality we are born into and, by doing so, represent new possibilities for reconciling the conflicting interests of persons. This assumption rests on Rawls's conception of the moral judgment. The moral judgment, for Rawls, may be the datum of moral theory, but this datum is "soft" in the sense that it can be revised as a result of theory-construction. The moral theorist is free to make any and all adjustments necessary to attain systematic coherence.<sup>7</sup> By going back and forth between explanatory

<sup>5</sup> John Rawls, "Kantian Constructivism in Moral Theory," *Journal of Philosophy* 77 (1980): 538.

<sup>6</sup> John Rawls, "The Basic Structure as Subject," in *Values and Morals*, eds. A.I. Goldman and J. Kim (Dordrecht: D. Reidel Publishing Company, 1978), 56.

<sup>7</sup> Rawls, "The Independence of Moral Theory", 8.

hypotheses about the conditions of moral choice and the judgments—employing what Rawls calls the method of reflective equilibrium—an adequate account of our moral framework is obtained.

But why should I care about the reflective equilibrium attained by moral theory between principles and judgments unless both happen to be mine? The data of moral theory, one might think, can change in light of knowledge of their principles only if both belong to the same person.<sup>8</sup> Rawls's reply is given in his Kantian interpretation of the initial situation. For Rawls, to the extent that moral theory explains our moral capacity, it must represent a conception that commands our allegiance. The task of the Kantian interpretation is to persuade us that Rawls's moral theory affirms a self-conception that we, in fact, desire to express. Rawls identifies three elements as essential to this self-conception, the ideas of morality, freedom, and equality.<sup>9</sup>

Rawls's theory claims to improve our moral competence: through reflective equilibrium our moral scheme comes to fit together as a coherent whole, and this whole is rightly seen as the expression of our Kantian self-conception. Moreover, moral theory provides the means by which social cooperation can be made to proceed on a fair basis. Therefore, although Rawls does not, like Kant, ground his moral doctrine in a theory of reason as a power of self-determination, he does view moral theory-construction as an exercise in human self-determination. Moral autonomy is the central concept in Rawls's theory not merely as it is represented in the thought-experiment articulating fair conditions of moral choice but as it represents the standpoint of the moral theorist—ultimately any individual—as constituting moral reality through the method of reflective equilibrium.

#### IV

Having reviewed the structural parallel between Hare's and Rawls's theories and Kant's *Groundwork*, it must be asked, in view of the extent to which Kant's original project is altered, in what sense Hare and Rawls may be said to belong to the Kantian tradition. The Kantian tradition in moral philosophy is characterized by an emphasis on moral freedom. Although this concept doubtlessly plays a central role in Hare's and Rawls's theories, it is unlikely that Kant would recognize their conceptions as having much to do with his own. For Kant, freedom is essentially connected with a concept of reason as a power of self-determination. Hare and Rawls offer alternative accounts of reason, but these fail to show the intimate connection, argued by Kant, between freedom and reason. At most they can establish the compatibility of freedom and reason under certain highly

<sup>8</sup> I owe this objection to Ernst Tugendhat, "Comments on Some Methodological Aspects of Rawls's Theory of Justice," *Analyse & Kritik* 1 (1979): 78-79.

<sup>9</sup> Rawls, "Kantian Constructivism in Moral Theory," 553-54.

idealized conditions. A defense of human freedom in moral life is the principal point of Hare's and Rawls's theories, but Kant would clearly not accept either theory as providing a justification for the belief in moral freedom. He would view their theories as elaborate rationalizations for a belief in freedom that is not rationally sustained. For Kant, freedom is tied to a concept of reason that consists essentially in self-determination. Hare's and Rawls's theories reinterpret self-determining reason in ways which, they believe, capture the essence of Kant's concept of reason, but neither articulates a moral theory based on principles of autonomy as such. Instead they offer substitutes and then argue that their alternative moral principles are compatible with a belief in autonomy.

Any attempt to ground morality on Kant's theory of reason is subject to an objection first formulated by Hegel and tacitly acknowledged by Hare and Rawls. According to Hegel, Kant's theory of reason gives rise to an ethics of inner intention so utterly divorced from the social world in which moral agents live and act so as to be unable to gain the foothold necessary to pronounce on the rightness and wrongness of concrete moral acts. Knowledge of right and duty are, Hegel believes, essential for moral life. But Kant, by abstracting the moral will from any possible content of moral acts, makes knowledge of duty impossible. Moreover, Kant's test of universalization, Hegel argues, merely serves to justify what is already assumed to be right since it presupposes the legitimacy of the institutional context to which the proposed act belongs. For example, in examining the maxim that a deposit for which there is no proof must be returned, Kant presupposes the legitimacy of property relations in the specific form of deposits. Kant's universalization test, Hegel concludes, yields definite results only if one assumes that the institutions, habits, and customs which form the context of moral action are legitimate in the first place. Any maxim can be justified by means of Kant's test, provided the agent makes the necessary assumptions. Consistency would provide some constraints on the will purporting to be rational—to accept the maxim of not respecting deposits has implications for the institution of property as a whole—but pure reason as a formal and abstract principle of the will does not dictate that any particular institution, with all the maxims proper to it, must belong to the moral universe.<sup>10</sup>

Hare and Rawls implicitly acknowledge Hegel's criticisms of the abstract and empty character of Kant's self-determining will. Both search for ways to provide their theories with determinate content and so avoid the charge of formalism. Yet they attempt to avoid this objection in very different ways: Hare by means of his version of utilitarianism and Rawls by emphasizing the social character of persons.

<sup>10</sup>Hegel, *Natural Law*, trans. T. M. Knox (Philadelphia: University of Pennsylvania Press, 1975), 77.

Hare's version of utilitarianism contains formal and substantive elements. The formal element rests on Hare's account of the logical properties of moral words and grounds the principle of utility. The use of moral language commits one to the principle that the preferences of all persons affected by a proposed act be determined and counted equally. The principle of utility provides, in Hare's theory, determinate content for moral deliberation. It does so, however, in a purely formal way: directing us to the relevant data, the principle does not, taken in isolation, determine moral choice. It is only after examining the actual preferences of persons that deliberation grounded by the principle of utility yields a specific result. Hare is quite clear that the determination of moral right is, ultimately, an empirical matter (*MT*, 5). The substantive element of moral theory on Hare's account rests on empirical investigation. As Hare writes, it is empirical inquiry that "brings our moral thinking into contact with the world of reality" (*MT*, 5). Apart from its ability to incorporate the results of empirical investigation into the actual preferences of persons, Hare's moral theory, by his own admission, would lack practical force. A purely logical inquiry into the meaning of moral language can never, on Hare's view, provide an answer to the question, What should I do?

Rawls characterizes Kant's concept of autonomy as "purely transcendent and lacking explicable connections with human conduct" (*TJ*, 256). Rawls suggests that his moral theory shows this connection can be made. Autonomy, in Rawls's theory, comes to be understood primarily in terms of the perspective of the moral theorist who is an empirical investigator seeking to identify the conditions for the choice of moral principles which will explicate our moral judgments. These conditions must, given their explanatory character, reflect the actual circumstances of human life. Moral theories such as Kant's which claim to establish principles which hold for all possible worlds must be empty for Rawls. "From the point of view of contract theory it amounts to supposing that the persons in the original position know nothing at all about themselves or their world. How, then, can they possibly make a decision?" (*TJ*, 159)

Rawls incorporates a Heglian element into his Kantian constructivism by acknowledging the primacy of social morality. It is this primacy which underlies Rawls's decision to begin his inquiry with the question of social justice rather than, as was the case with Kant, the question of individual obligation or duty. Rawls, moreover, assuming the social nature of persons, concludes that the Kantian conception of self underlying his conception of justice can only be realized when a just institutional arrangement exists.

[Kant] seems to think that the conception of ourselves as fully autonomous is already given to us by . . . our recognition that the moral law is supremely authoritative for us as . . . rational beings. Thus this conception of ourselves is implicit in individual moral consciousness, and the background social conditions for

its realization are not emphasized or made part of the moral doctrine itself.<sup>11</sup>

For Rawls, by contrast, no explanation of moral consciousness is complete which fails to delineate the structure of the institutional arrangement that makes it possible. It is not simply that Rawls acknowledges to a greater degree than Kant that our social relations have implications for our conception of self; Rawls contends that our self-conception arises in and through these relations.<sup>12</sup> By emphasizing the dependence of our self-conceptions on background social conditions, Rawls implies that the Kantian concept of the moral subject has emerged historically. The very possibility of moral theory as construed by Rawls depends on historical conditions.

## V

Hare and Rawls both attempt to avoid Kant's problem of formalism, but to do so in a way that acknowledges the freedom of the moral subject. Yet both theories abandon Kant's conception of the moral subject as an autonomous, rational agent and, from Kant's perspective, regress to a conception of the moral agent as subject to laws which proceed independently of his or her will. Whatever Hare's and Rawls's success in generating determinate answers to moral questions, it is doubtful whether their theories can be considered to advance the Kantian position since these answers cannot be construed as the dictates of an autonomous moral subject.

Hare comes close to conceding the necessity of heteronomy. Still, Hare's ideal moral reasoner, the archangel, deliberates in light of the *de facto* preferences of those affected by a particular course of action, regardless of their source. Heteronomy in Hare is not merely a concession to human limitation but an implication of moral reasoning as such. The limits of moral reasoning are set by desire, regardless of whether the desire in question is itself the product of autonomy or established in some other way.

In Rawls's case, the heteronomy of the moral point of view is less readily discernible, but only because his theory is more complex than Hare's. If one does not grasp the whole, Rawls's theory can be read as an elaborate model of how society can be understood as founded on the right of individuals to be recognized as free, equal, and moral, thereby preserving what Hegel, among others, took to be the core of Kant's concept of autonomy.<sup>13</sup> But in rejecting Kant's distinction between laws of nature and laws of freedom, Rawls implies the heteronomy of the moral will. Rawls's conception of moral theory as an explanatory theory like any other assumes that moral activity

<sup>11</sup>Rawls, "Kantian Constructivism in Moral Theory," 553-54.

<sup>12</sup>*Ibid.*, 568-69.

<sup>13</sup>Hegel, *Lectures on the History of Philosophy*, trans. E. S. Haldane and F. J. Simson (Atlantic Highlands, New Jersey: Humanities Press, 1974), III, 549.

can be conceived within the totality of nature. For Rawls, moral theory is part of an encompassing "theory of the natural order and our place in it" (*TJ*, 512). Rawls goes so far as to suggest that the "Kantian" concept of the self as a moral subject is ultimately a biological response to our environment.

The theory of evolution would suggest that. . .the capacity for a sense of justice and the moral feelings is an adaptation of mankind to its place in nature. . . It seems clear that for members of a species which lives in stable social groups, the ability to comply with fair cooperative arrangements and to develop the sentiments necessary to support them is highly advantageous. (*TJ*, 503)

Rawls's underlying naturalism is further illustrated by his claim that "the principles of justice are closer to the tendency of evolution than the principle of utility" (*TJ*, 503). Such statements show that although Rawls does not assert the heteronomy of the will in Hare's manner by allowing individual preferences to determine moral deliberation, he has nonetheless failed to develop a concept of society which can itself be understood as the concrete association of autonomous subject.

## VI

Although neither Hare nor Rawls succeeds in developing variants of the *Groundwork* project which at once avoid the charge of formalism and preserve Kant's conception of the moral point of view as the standpoint of the self-determining subject, Rawls in particular provides the clue for what is needed for success. Rawls's theory of justice, detached from his underlying naturalism, takes the moral subject as the foundation of all social relations. Whereas particular moral acts are conditioned by social practices, subjectivity, in the form of our Kantian conception of self, is, for Rawls, the principle informing our social institutions to the extent that they are just. What is needed to escape heteronomy, given this starting point, is the development of the concepts of the person and society as shaped by freedom, where freedom is not in turn understood as the product of natural forces but rather as the concrete unity of social life. Rawls's contention that moral theory can have practical import only if it has implications for our understanding of social institutions and practices is sound. But the social order foundational for our understanding of morality must itself be actually affirmed by subjects and not merely capable of being conceived as if it were affirmed. Unless we start with the actual affirmations of the social order by persons, through their practices, then we will have no grounds, given the social nature of persons, to affirm their autonomy. Only such a social conception advances the *Groundwork* project by overcoming the problem of formalism without, at the same time, abolishing the autonomous subject.

If understanding the freedom of human activity depends on understanding the social order which forms its ultimate context, then a positive conception of freedom might be developed through an account of the virtues proper to basic social practices. Such a concept, once elaborated, would enrich the three phases of the *Groundwork* project in the following ways. The account of moral rules given in the first phase could be supplemented with an account of rules exhibiting how virtues are developed within particular practices.<sup>14</sup> The account of the supreme principle constitutive of morality could be enriched by a concept of justification which assesses the role of practices within political traditions, i.e., which assesses whether the virtues practices require and promote are positive or negative forces in the development of a tradition.<sup>15</sup> And, finally, the theory of reason which establishes the possibility of the moral point of view should make sense of the perspective from which we shape complete lives within society, i.e., the perspective from which we shape our personal identities.<sup>16</sup>

Assuming that our conception of morality depends on our actual social relations, the notion of the Kantian subject must be understood in a social context. But our account of this context must be consistent with the freedom of the subject. This is only possible if the subject is represented as a positive determiner of its social relations and, ultimately, of itself. Hare and Rawls articulate alternative ideas of freedom. But the freedom to prefer what we prefer or to entertain a self-conception that we cannot posit as real does not make the possibility of autonomy as established by Kant's *Groundwork* determinate. On the contrary, the arguments of Hare and Rawls, if valid, would show Kant's conception of freedom as autonomy to be impossible.

*Sacred Heart University,  
Fairfield, Connecticut*

<sup>14</sup>Cf. Günther Maluschke, "Universalisierbarkeit und Institutionalisierbarkeit sittlicher und politischer Normen," in *Normenbegründung- Normendurchsetzung*, ed. Willi Oelmüller (Paderborn: Schöningh, 1978).

<sup>15</sup>Alasdair MacIntyre's *After Virtue: A Study in Moral Thinking* takes a step in this direction (Notre Dame, Indiana: University of Notre Dame Press, 1981).

<sup>16</sup>This problematic is discussed in Klaus Hartmann's "Moralität und 'konkretes Allgemeines'," *Archiv für Geschichte der Philosophie* 60 (1978). Cf. also his "Towards a New Systematic Reading of Hegel's Philosophy of Right," in *The State and Civil Society*, ed. Z. A. Pelczynski (Cambridge: Cambridge University Press, 1984).