Originally Processed With FOIA(s):
S

FOIA Number:

FOIA MARKER

This is not a textual record. This is used as an administrative marker by the George Bush Presidential Library Staff.

Record Group/Collection: George H.W. Bush Presidential Records
Collection/Office of Origin: Speechwriting, White House Office of
Series: Snow, Tony, Files
Subseries: Subject File, 1988-1993

OA/ID Number: 13896
Folder ID Number: 13896-005

Folder Title:
Morality and Economics, [1988]

Stack:	Row:	Section:	Shelf:	Position:
G	18	29	2	4

SOCIAL
PHILOSOPHY
& POLICY CENTER

MORALITY AND ECONOMICS
DE MORIBUS EST DISPUTANDUM

Viktor Vanberg

Transaction Books
New Brunswick (USA) and London (UK)

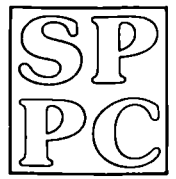
Editor of Original Papers: Ellen Frankel Paul
Managing Editor: John Ahrens

The Social Philosophy and Policy Center, founded in 1981, is an interdisciplinary research institution whose principal mission is the examination of public policy issues from a philosophical perspective. In pursuit of this objective the Center supports the work of scholars in the fields of political science, philosophy, law, and economics. In addition to the Original Papers Series, the Center hosts scholarly conferences, and publishes both a professional journal, SOCIAL PHILOSOPHY & POLICY, and a book series. For further information on the Center, write to the:

Social Philosophy and Policy Center
Bowling Green State University
Bowling Green, OH 43403 USA

About the Author:

Viktor Vanberg is an Associate Professor of Economics at George Mason University in Fairfax, Virginia, and a Research Associate at the Center for Study of Public Choice. Before coming to the United States in 1983, he taught at universities in West Germany. His current research interests are constitutional economics and the economic analysis of rules and institutions. He is the author of numerous publications (in both German and English) on the individualistic tradition in social theory, and particularly the application of this theory to the study of social institutions. Among his more recent publications are articles in *Economics and Philosophy* and the *Journal of Law, Economics, and Organization.*



SOCIAL
PHILOSOPHY
& POLICY CENTER

MORALITY AND ECONOMICS
DE MORIBUS EST DISPUTANDUM

Viktor Vanberg

Transaction Books

Original
Papers No. 7

Published by the Social Philosophy and Policy Center
and by Transaction, Inc. 1988

Library of Congress Catalog Number: 87-19213
ISBN: 0-912051-19-1
Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Vanberg, Viktor.
Morality and economics.

(Original papers; no. 7)
At head of title: Social Philosophy and Policy Center.
1. Economics—Moral and ethical aspects. I. Bowling Green State University. Social Philosophy & Policy Center. II. Title. III. Series: Original papers (Bowling Green State University. Social Philosophy & Policy Center); no. 7.
HB72.V268 1987 174'.26 87-19213
ISBN 0-912051-19-1 (pbk.)

MORALITY AND ECONOMICS
DE MORIBUS EST DISPUTANDUM

tion and those who misperceive reality can "rationally" be moral is not an attractive alternative either. What one would certainly prefer is to live in a society which is structured in such a way that morality is a rational choice, at least for most of the individuals involved. And to the extent that appropriate structures do not emerge spontaneously, it will be a matter of deliberate political action to make sure that conditions prevail which allow for a *rational morality*.

Table of Contents

IX. Conclusion

It has often been suggested (e.g., by Edmund Burke in the epigraph to this essay) that there exists some kind of substitutability between individual morality and social enforcement.[70] In some sense this may be true. As I have argued in this essay, however, individual morality is essentially conditioned by social enforcement. In order to be viable, morality must be perceived as an efficient strategy, and this will only be the case if rewards for moral conduct and punishments for rule violations are appropriately allocated in a social community. The current level of morality in a community always reflects past enforcement of moral rules, and it will only be maintained by ongoing enforcement. The level of morality can, in this sense, be viewed as *social capital* that, in the absence of sufficient reinvestment, will depreciate over time.[71] It is the crucial function of social enforcement—whether formal or informal—to make sure that morality can, in fact, be perceived as a successful strategy. Where informal enforcement ceases to be effective, formal enforcement may serve as a substitute mechanism, though it may well be that in various respects it is only a poor substitute.

As the foregoing analysis implies, the proper purpose of formal enforcement can only be to affect people's constitutional, strategic choices, not their situational choices. Its purpose cannot be to make sure that to act morally is always the rational choice. It is simply impossible and certainly undesirable to organize large-scale social interaction in such a way. Neither is it necessary. All that is required for a beneficial social order is that the individual participants have good reasons to assume that morality is a promising strategy overall. To make sure that always being moral is a rational strategy is, however, a much easier task than to make sure that being moral is always rational. Life in a society where the latter is achieved would probably be extremely unpleasant. And to live in a society where only those who believe in transcendental remunera-

[70] See James M. Buchanan, *The Limits of Liberty—Between Anarchy and Leviathan* (Chicago: The University of Chicago Press, 1975), pp. 117f., and "The Moral Dimension," pp. 1, 5.

[71] It should be kept in mind that throughout this essay, all issues are ignored which arise with dissent on what "moral rules" ought to be enforced in a social community. The discussion herein proceeds on the assumption that there exists a set of rules to which obedience is generally considered desirable, and morality is viewed as the disposition to follow those rules.

forcement mechanisms have changed the relevant payoffs, one should expect individuals to adjust their "moral strategies" and to redefine their own demarcation line between the kind of social setting where they consider morality a reasonable strategy, and those social settings where this is not the case.

If these adaptive changes in people's moral dispositions are considered a "social problem," moral appeals will hardly be a promising remedy. As David Hume argued two centuries ago, it is "impossible to change or correct any thing material in our nature, the utmost we can do is to change our circumstances and situation."[69] What would seem to be required are structural changes which—if they are a matter of choice at all—are a matter of social, constitutional choice. Such structural changes may include institutional reforms, such as various forms of "decentralization," which aim at reactivating to some extent informal enforcement mechanisms. To a large extent, however, such constitutional measures may simply be confined to using the means of formal enforcement to compensate for the loss of informal enforcements.

[69] David Hume, *A Treatise of Human Nature* (Oxford: Clarendon Press, 1967), p. 537.

I. Introduction[1]

> "Men are qualified for civil liberty in exact proportion to their disposition to put moral chains upon their own appetites. . . . Society cannot exist unless a controlling power upon will and appetite be placed somewhere, and the less of it there is within, the more there must be without."
>
> Edmund Burke

Ever since Adam Smith reasoned about men's *Moral Sentiments*, as well as the nature and causes of the *Wealth of Nations*, the relation between ethics and economics has been an issue of scholarly interest; and—judging from publication rates—it seems to be an issue which recently has attracted considerable renewed attention.[2]

Current and received contributions on the issue seem to fall roughly into two categories. The first category includes contributions which, in one way or another, discuss the potential role of ethical judgments—judgments about what ought or ought not to be—in economic theory itself. Such contributions are essentially concerned with the limits and scope of normative economics and the rational justification of the ultimate criteria of evaluation upon which normative economic statements are based.[3] The second category comprises contributions which, in one

[1] I am indebted to James M. Buchanan, Bunshiro Ando, Hartmut Kliemt, and an anonymous referee for helpful suggestions and comments on an earlier draft, which I presented at Professor Karl Brunner's Interlaken Seminar on Analysis and Ideology, May 1986.

[2] For a selection of recent publications on the issue, see Allan Buchanan, *Ethics, Efficiency, and the Market* (Oxford: Clarendon Press, 1985); James M. Buchanan, "Ethical Rules, Expected Values, and Large Numbers," *Freedom in Constitutional Contract—Perspectives of a Political Economist* (College Station and London: Texas A&M University Press, 1977), pp. 151-168, and "The Moral Dimension of Debt Financing," *Economic Inquiry*, vol. 23 (1985), pp. 1-6; Georges Enderle, ed., *Ethik und Wirtschaftswissenschaft* (Berlin: Duncker & Humblot, 1985); Jack High, "Is Economics Independent of Ethics?" *Reason Papers*, vol. 10 (1985), pp. 3-16; David Levy, "David Hume's Invisible Hand," *The Wealth of Nations—The Public Choice of Moral Information, Hume Studies*, 10th Anniversary Issue (1985), pp. 110-149; Michael S. McPherson, "Limits on Self-Seeking—The Role of Morality in Economic Life," D.C. Colander, ed., *Neoclassical Political Economy* (Cambridge, MA: Ballinger, 1984), pp. 71-85; Amartya Sen, et al., "Ethics and Economics," *Social Philosophy & Policy*, vol. 2 (1985); George J. Stigler, "The Ethics of Competition: The Friendly Economists," *The Economist as Preacher and Other Essays* (Chicago: University of Chicago Press, 1982), pp. 14-26, and "The Ethics of Competition: The Unfriendly Critics," *ibid.*, pp. 27-37; Manfred Tietzel, "Moral und Wirtschaftstheorie," *Zeitschrift fuer Wirtshafts-und Sozialwissenschaften*, vol. 106 (1986), pp. 113-137.

[3] On the rational discussion of ethical principles, see Hans Albert, *Treatise on Critical Reason* (Princeton, NJ: Princeton University Press, 1985), pp. 71ff.

way or another, discuss ethics or morality as an aspect of human behavior, i.e., as part of the empirical reality studied by economists. Such contributions are essentially concerned with the explanatory status and significance which theoretical economics, as a positive science, is to attribute to ethics or morality in its analysis of human behavior.

This essay belongs in the second category. Its purpose is to analyze how ethics or morality can be incorporated as a potential explanatory variable into economic theory. Section II specifies the particular issue to be addressed, and Section III discusses some common conceptions of how economics might account for a variable "morality." Sections IV, V, and VI propose an alternative conception, based on the logic of Gary Becker's and George Stigler's reformulation of the theory of consumer choice.[4] Section VII compares the approach advocated here to the game theoretical argument on recurrent PD games and to David Gauthier's arguments concerning "rational morality." Some implications are discussed in Section VIII and a conclusion is drawn in Section IX.

[4] Gary S. Becker, *The Economic Approach to Human Behavior* (Chicago and London: The University of Chicago Press, 1976); George J. Stigler and Gary S. Becker, "De Gustibus Non Est Disputandum," *The American Economic Review*, vol. 67 (1977), pp. 76-90.

environment. One factor which will tend to make a person's moral dispositions somewhat less responsive to environmental changes is the emotional reaction which we describe as "bad conscience," a reaction which apparently is learned concomitantly with the process in which rule-governed behavior emerges. Once a general behavioral pattern has been conditioned in a person, deviations from the rule typically do not leave the person emotionally unaffected but, rather, tend to cause a feeling of uneasiness. As a *learned* response, however, these emotional reactions may also be unlearned, and they will be if a person experiences directly or indirectly that deviations remain without external sanctions. That is to say, internal sanctions cannot be considered an autonomous, independent source of moral enforcement. They are a source that ultimately derives from, and ultimately remains dependent on, the direct or indirect experience of external enforcement. In the absence of any external reinforcement, people can hardly be expected to develop a conscience. And an individual's conscience will typically not remain unaffected if changes in relevant environmental characteristics occur. If such changes occur, individuals will presumably not respond with instant adjustments in their "moral strategies." Neither, however, can they be expected to permanently preserve a moral disposition irrespective of systematic changes in the relevant payoffs.[67]

What is often described as a process of "moral erosion" in modern industrial societies may—to the extent that it actually occurs—simply reflect a rational adaptation of people's moral strategies to systematic changes in their social environment. The most obvious changes that occurred in these societies over, say, the past two centuries can be summarized in analytical terms as a progressive replacement of small-group interaction by large-group interaction, with the typical consequences which result from the fact that the mechanisms of *informal enforcement* which constrain the former are not effective for the latter.[68] A particularly significant aspect of this process has certainly been the change in the structure and role of the family. It is a person's learning history within his family of origin which obviously has a crucial impact on what one may call his moral development or, in more behaviorist language, his moral conditioning. To the extent that the changing role of the family and the general weakening or even elimination of informal en-

[67] Scott points out, on pp. 106f. of *Internalization of Norms*, that people's "alertness" to environmental changes will vary with the "strength" of their moral commitment: "Moral learning, like all learning, is a matter of degree. Extinction or termination of a response is slow when learning is relatively complete. But learning varies in its completeness. . . . There is. . .a continuum of moral learning, which may be represented as running between two poles: complete calculation and complete commitment."

[68] See, e.g., Roland N. McKean, "Economics of Trust, Altruism, and Corporate Responsibility," E.S. Phelps, ed., *Altruism, Morality, and Economic Theory* (New York: Russel Sage Foundation, 1975), pp. 29-44.

does not mean that individual morality is entirely uncalculated. The real issue is not whether the individual calculates or not, but at what level benefits and costs are compared. The "calculus of morality" is, as I explained above, a constitutional calculus, the very essence of which is to "compare" the overall benefits and costs of adopting a moral routine with those of case-by-case choices.

In contrast to what is sometimes suggested in sociological treatments of the subject, people's morality or commitment to "internalized norms" is never unconditional. It is a disposition which individuals will not acquire if their direct or indirect experience does not tell them that moral behavior in general "pays," and which they will not preserve if their experience systematically changes. Or, in other terms, "moral commitment requires sanctions for its maintenance as well as for its origins."[65] Whether or not individuals have reasons to learn that morality is an "efficient" strategy essentially depends on the character of their environment. An individual's social learning over a considerable initial time span typically takes place in small group settings (notably the family), a crucial characteristic of which is that individual behavior is closely monitored and rule enforcement is quite effective. In this type of environment there are generally good reasons for an individual to acquire a general disposition to respect "the rules of the game," and individuals tend to carry this disposition with them as they enter into other social settings. This tendency to *generalize*, to transfer a behavioral strategy that has proven to be "successful" in one context into a new context, is, however, typically tentative and conjectural. It is controlled by the opposite tendency to *discriminate* between those kinds of environment in which a certain activity or strategy is in fact advantageous, and those environments where this is not the case. That is, depending on the relevant characteristics of different social settings, people's *morality* may well be restricted to specific environments only.[66]

Just as people will tend systematically to discriminate among social settings in which morality is or is not experienced as an efficient behavioral strategy, they can also be expected to respond to changes in their environment which affect the relative payoffs from morality on the one hand and expedient case-by-case choices on the other. Although morality, like all rule-governed behavior, is to some extent necessarily unresponsive to environmental variation (perfect responsiveness would, in fact, be incompatible with the very notion of a rule), morality is a matter of an individual's adaptation to relevant characteristics of his

[65] *ibid.*, p. 103. See *ibid.*, pp. 92f.: "A norm is learned or internalized when the actor has been sufficiently conditioned by sanctions that his behavior conforms to the norm at a spatial or temporal distance from sanctions. Such learning is, however, never complete; it remains ultimately dependent on subsequent sanctions, even if the prospect is remote."

[66] See the aforementioned notion of "honor among thieves."

II. Morality and Moral Behavior

The notion of ethics or morality apparently implies the idea of some normative standards or rules of behavior, the general observance of which is typically considered to be conducive to the existence of a beneficial and properly functioning social order.[5] More specifically, one of the important characteristics of moral rules seems to be that they ask for a kind of behavior that is considered socially beneficial—i.e., beneficial to all individuals in some defined social community—but that appears to be in conflict with the immediate or direct interests of the actor himself. Moral rules impose "impartial constraints" on the pursuit of individual interests, constraints which are socially desirable in serving interests that individuals share as members of a social community.[6] This may not be perfectly descriptive of all kinds of rules which everyday language classifies as "moral." But it is certainly true for a relevant subset of what is commonly perceived as rules of "morality," and it is this subset which is of interest for the purposes of this essay. The rules that require us "to keep promises" or "not to cheat," for instance, will typically be considered standard examples of moral rules, while a social convention, such as driving on the right side of the road, will probably not be, presumably because the latter does not involve (at least not under standard circumstances) a conflict between individual and common interest.[7]

[5] See, e.g., Stigler's "Unfriendly Critics," p. 35: "What we call ethics. . .is a set of rules with respect to dealings with other persons, rules which in general prohibit behavior which is only myopically self-serving, or which imposes large costs on others with small gains to oneself. General observance of these rules makes not only for long-term gains to the actor but yields outside benefits." If it is intended to be a definition of ethics, Stigler's statement is certainly burdened with a number of factual assumptions which can be and, in this essay, will be separated from purely definitional considerations.

[6] In *Morals by Agreement* (Oxford: Clarendon Press, 1986), David Gauthier defines moral principles as those "that constrain the actor pursuing his own interest in an impartial way" (p. 3) and "the conception of morality as a set of rational, impartial constraints on the pursuit of individual interest" (p. 6).

[7] The argument can be stated more precisely in game-theoretical terms: typically those rules seem to be classified as moral rules which provide solutions to recurrent Prisoner's Dilemma (PD) situations as opposed to recurrent coordination-type situations, to which conventions like "driving on the right side" provide general solutions. See Viktor Vanberg, "Spontaneous Market Order and Social Rules—A Critical Examination of F.A. Hayek's Theory of Cultural Evolution," *Economics and Philosophy*, vol. 2 (1986), pp. 91ff. The potential conflict between common interest and individual interest which characterizes PD rules, in contrast to the largely self-enforcing character of coordination rules, seems to

Discussions of ethics or morality as a potential determinant of human behavior apparently are concerned with how moral behavior can be explained. That is, why do people behave in accordance with moral rules—if and to the extent that they actually do so?[8] Our everyday interpretations of human behavior, as well as standard social theoretical accounts, typically refer to some kind of expected sanctions as factors which explain why a person obeys moral rules. Those sanctions may range from the more obvious formal sanctions imposed by some special enforcing agency, over the sometimes less visible informal sanctions people spontaneously impose on each other, to the totally intangible internal sanctions people's consciences impose on them (including the imagination of sanctions expected from some transcendental entity).

There seems to be, in principle, no reason why the economic model of rational choice should not incorporate such expected sanctions as explanatory variables, viewing them as part of the opportunity costs and, consequently, among the determinants of people's choices. Yet, de facto, economic analysis apparently concentrates—to the extent that it pays attention to the social-institutional context at all—on formal sanctions, paying less attention to informal enforcement mechanisms and being extremely hesitant to refer to *intrapersonal* sanctions as explanatory variables. It is particularly the role of such intrapersonal behavioral determinants, however, to which arguments on ethics or morality typically seem to refer.[9] And the specific issue on which the following analy-

be central to our common notion of *moral* rules. This is not meant to imply, however, that all principles which make actors in PD situations choose the "cooperative" alternative qualify per se as moral principles. See, e.g., Edna Ullmann-Margalit, *The Emergence of Norms* (Oxford: Clarendon Press, 1977), pp. 41ff.

[8] A separate issue which will not be discussed here is why certain kinds of rules—rather than some conceivable alternative rules—are considered moral rules in particular social groups. If, as indicated above, moral rules are viewed as providing solutions to recurrent problems (in particular of the PD type) in social interaction, it seems plausible to assume that moral rules will share (through history and across cultures) certain common basic characteristics, to the extent that the general structure of typically recurrent interaction problems is similar in all social groups. Such an assumption seems to be a central underlying premise of the "individualistic-evolutionist conception of social institutions" that I describe in "Spontaneous Market Order," the modern versions of which can be traced back to David Hume and Carl Menger. See Menger, *Investigations into the Method of the Social Sciences with Special Reference to Economics* (New York and London: New York University Press, 1985), pp. 228f.: "Certain conditions resulting from general human nature and thus appearing everywhere produce similar institutions of law everywhere by their nature, while tribal differences and variety of external conditions. . .result in differences in law." For an excellent discussion of Hume's theory of "moral institutions," see Hartmut Kliemt, *Moralische Institutionen—Empiristische Theorien ihrer Evolution* (Freiburg/Muenchen: Karl Alber, 1985).

[9] See, e.g., John L. Mackie's *Persons and Values* (Oxford: Oxford University Press, 1985), p. 159, which considers the "internalization of norms" to be a crucial ingredient of "morality": "The association of moral sentiments with the practices. . .and a sense of guilt about one's own transgressions, is a major part of such internalization. Only when this stage is reached can we speak properly of a morality."

VIII. Morality: Individual Adaptation and Collective Choice

In sociology, a common objective against a rational choice interpretation of moral conduct is that "social order" would not be possible if people were to actually calculate the costs and benefits of rule obedience and rule violation in each and every choice situation, because there are too many situations in which the benefits from "cheating" would clearly outweigh the costs. Therefore, the standard argument goes, one has to assume that people typically tend to follow social norms in a habitual, noncalculating manner. Such a general inclination towards uncalculated rule obedience is typically explained by the assumption that in a process of *socialization*, individuals *internalize* the social norms of their community.

What, in sociology, is labeled "internalization of norms" obviously corresponds to what, in the context of this essay, is viewed as the process in which individuals learn or acquire the dispositional trait *morality*. The foregoing analysis suggests that, if the level of individual "constitutional" choice is taken into account, a rational choice perspective can indeed explain the kind of *uncalculated* rule obedience at the level of particular situational choices which sociologists have stressed as a crucial ingredient of social order.

To analyze the phenomenon of "internalization" in terms of an individual's "constitutional choice" will certainly appear to most sociologists as an inadequate rationalization of what, in fact, may be a largely unintentional process. Actually, this should not be a crucial issue. Whether the acquisition of the dispositional trait *morality* is modeled as a process of rational constitutional choice or as a process of unreflected behavioral conditioning,[64] an individual can be expected to acquire and to preserve this disposition only if it is experienced as generating sufficient advantages or rewards. What is crucial for the general research strategy—i.e., for the issue of where one looks for an explanation—is not whether the process is described in terms of rational choice or of unreflected habit formation. What is crucial is whether or not the relevant costs and benefits to the *individual* are viewed as the central explanatory variable. That individual moral acts may be uncalculated

[64] For an excellent discussion of the notion of norm internalization in the terms of behavioral psychology, see Scott, *Internalization of Norms*.

particular actions in one case, routines in the other—and the kind of expected consequences that are of relevance in making a choice—specific outcomes in one case, patterns of outcomes in the other. In contrasting his concept of "constrained maximization" with the "traditional" notion of "straightforward maximization" Gauthier, on the other hand, seems to suggest that more is involved than just applying the same notion of rationality to different levels of choice. It is the additional claims which Gauthier attaches to his concept of "constrained maximization" which mark the essential difference between his approach and the one taken here. Measured against the latter, Gauthier's approach appears to try to prove too much or, in any case, more than is needed in order to show that there can be a rational basis for morality, understood as a general disposition to act in accordance with moral rules.

sis will focus concerns the nature and potential explanatory significance of those intangible, intrapersonal determinants of choice which account for a person's moral conduct, in addition to and separate from actual threats of formal or informal external sanctions.[10]

Before commencing the discussion of this issue, a point of conceptual clarification is in order. *Moral behavior* and *morality*, at least as the concepts are used here, should be carefully separated. Moral behavior is understood here in a purely descriptive sense, as behavior that is in fact in accordance with moral rules—irrespective of what the actor's actual motivation might be. Morality, on the other hand, is understood as a dispositional trait which accounts for a person's moral conduct. Whether a person's behavior qualifies as *moral behavior* in this sense is properly a matter of observation and classification. Whether a person acts out of morality is an explanatory issue. To use, as is sometimes done, the term "moral behavior" in a sense that makes "morality" one of its definitional attributes is to blur the distinction between what is to be explained (moral behavior), and how it is to be explained (by reference to morality or to other factors). As it is used here, the notion of morality does not refer to any observable behavior as such, but to some intrapersonal, dispositional variable which may account for observable moral behavior.

Authors who argue in favor of incorporating an intrapersonal determinant of behavior like morality into economic explanations generally justify their approach by pointing out that there is ample evidence that people respect moral rules even in situations in which—so far as an observer can discover—the external incentive structure does not seem sufficiently to motivate such "moral behavior."[11] Examples of such moral conduct, which seem to reflect the genuine impact of some intrapersonal, moral force, range from more trivial instances like people not littering when unobserved at a remote beach, over more significant in-

[10] To be sure, the distinction between "external" and "internal" sanctions or behavioral determinants is not totally unambiguous since, strictly speaking, *expected* sanctions are the directly relevant variable for a theory of individual choice. And a person's expectations of "external" sanctions are, in principle, no less intrapersonal and subjective than what is explicitly classified here as intrapersonal determinants. However, if one rules out, as is typically done in economics, the possibility that people permanently and systematically misperceive relevant characteristics of their environment, it can be reasonably assumed that people's expectation on the whole reflect the "actual" threats of formal and informal sanctions which the observer–analyst can discover in the respective setting. This does not require postulating that people never misperceive their environment. It only requires ruling out that such misperceptions are a relevant source of systematic differences and changes in people's behavior.

[11] Richard B. McKenzie, "The Economic Dimensions of Ethical Behavior," *Working Paper Series* (Blacksburg, VA: Center for Study of Public Choice, 1976, Mimeographed), pp. 7f.; McPherson, "Limits on Self-Seeking," p. 77; Robert H. Frank, "If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience?" (Ithaca, NY: Cornell University, 1985, Mimeographed), p. 3.

stances like honest behavior towards strangers, to the most spectacular instances like people heroically sticking to certain principles despite considerable costs to themselves.

Certainly, for each and every one of those particular examples one may suppose that a more careful analysis should be able to disclose some subtle or hidden "external" determinants which would allow for a satisfactory explanation without any need to bring in a special intrapersonal variable like morality.[12] But as long as those "hidden external determinants" are not identified and shown to provide a satisfactory explanation, it is legitimate and worthwhile to ask whether and how an intrapersonal variable *morality* can be incorporated consistently into economic theory. Which one of the two alternative strategies—to search for "hidden external determinants" or to incorporate *morality*—will turn out to be the theoretically more fruitful one is a matter of fact that cannot be decided on a priori grounds.

[12] It should be obvious that the morality issue as defined here is different from the issue of informal versus formal sanctions, though the two are sometimes confused. That people often respect moral rules in the absence of formal enforcement is a trivial and uncontroversial observation. What is controversial is not whether informal sanctions matter (who would want to deny this?), but whether moral behavior occurs in the absence of (sufficient) external sanctions of any kind, formal or informal.

So far, Gauthier's argument is perfectly analogous to the argument put forward in this essay. The crucial difference between the two approaches arises from Gauthier's supposition that the rationality of a constitutional moral choice directly carries through to the level of particular, situational choices.[61] No such claim is made by the present approach. All that is claimed here is that it can be rational—provided that certain conditions are given—to make a constitutional moral choice, i.e., to adopt a moral routine and renounce the attempt to make an "optimal" choice in each and every particular choice situation. In arguing that a moral routine can be rational, what I argue is that it can be rational *not to calculate* in particular choice situations whether or not to follow a moral rule, but rather routinely to act morally without any such situational calculation. I definitely do not mean to imply that, once a constitutional moral choice is made, it must be rational for a person to act morally in each and every choice situation, *even if* he should identify a particular choice situation as one in which—when all consequences are properly accounted for—the immoral choice appears to be the "maximizing" one.

The point is subtle, but crucial. In part, Gauthier also argues as if the rationality of *not* calculating at the level of particular choices is what he considers to be the relevant implication of a rationally adopted moral disposition.[62] But he clearly seems to make a stronger claim when he argues that, once a rational choice to adopt a moral disposition is made, rationality requires a person to comply even if, in her judgment, this means "acting in a way that results in real disadvantage to herself."[63] In making such a claim, Gauthier seems to give to his concept of *rational moral choice*—which he calls "constrained maximization"—an interpretation which not only shifts the notion of rational choice from the situational to the constitutional level, but also changes the meaning of the concept of rationality itself.

The distinction between situational and constitutional levels of choice, as it is used in the approach advocated here, does not require any presumption that the meaning of "rational choice" is somehow different in both contexts. Rationality means, in both cases, to choose the alternative which is expected to yield the largest payoffs. What is different is the character of the alternatives among which a choice is to be made—

straightforward maximization, and of adopting constrained maximization, as his disposition for strategic behaviour."

[61] *ibid.*, p. 187: "Our argument identifies practical rationality with utility-maximization at the level of dispositions to choose, and carries through the implications of that identification in assessing the rationality of particular choices."

[62] This aspect is stressed when Gauthier argues on p. 167 of *Morals by Agreement* that a "constrained maximizer" is disposed to cooperate "without considering whether some individual strategy would yield her greater expected utility."

[63] *ibid.*, p. 169.

Gauthier's approach and the one adopted here are much closer in their central thrust, and the differences between them are much more subtle. Like the present approach, Gauthier's approach puts its main emphasis on the distinction between a "constitutional" and a "situational" choice level. As Gauthier stresses, in contrast to an interpretation which "identifies rationality with utility-maximization at the level of particular choices,"[52] his own approach identifies "rationality with utility-maximization at the level of dispositions to choose."[53] The choice to be moral is analyzed as "a choice about how to make further choices,"[54] as "a choice among dispositions to choose."[55] And the "rationality of morality" is viewed as a matter of *rational dispositions*, not of rational actions;[56] the capacity to choose one's dispositions is considered "an essential part of human rationality."[57]

According to Gauthier, the crucial reason why a disposition to act morally can be rational is that it allows a person to get access to cooperative arrangements, and thus to potential gains from cooperation, from which a person lacking such a disposition would be excluded.[58] Though in a particular choice situation violation of a moral rule may be advantageous compared to moral conduct, the disposition to act morally may benefit a person more than a disposition which allows for such transgressions.[59] What is relevant to Gauthier's "calculus of morality" is the comparison of the overall benefits that can be realized from alternative dispositions, rather than the comparison of the benefits from particular alternative choices.[60]

[52] *ibid.*, p. 182.

[53] *ibid.*

[54] *ibid.*, p. 158.

[55] *ibid.*, p. 183.

[56] *ibid.*, p. 182f.: "A disposition is rational if and only if an actor holding it can expect his choices to yield no less utility than the choices he would make were he to hold any alternative disposition."

[57] *ibid.*, p. 183: "But there is a further significance in our appeal to choice among dispositions to choose. For we suppose that the capacity to make such choices is itself an essential part of human rationality."

[58] *ibid.*, p. 162: "The disposition to keep one's agreement. . .makes one an eligible partner in beneficial co-operation, and so is itself beneficial." See *ibid.*, p. 183: "The essential point in our argument is that one's disposition to choose affects the situations in which one may expect to find onself."

[59] *ibid.*, p. 162: "A person disposed to violate his covenants. . .cannot rationally expect to reap the benefits available to co-operators. Even if his particular breaches of covenant would benefit him, yet his disposition that leads him to such breaches does not."

[60] With respect to "constrained maximizers," as he labels morally disposed persons, Gauthier states on p. 15 of *Morals by Agreement*: "Of course, constrained maximizers sometimes lose by being disposed to compliance. . . . Nevertheless, we shall show that under plausible conditions, the net advantage that constrained maximizers reap from co-operation exceeds the exploitative benefits that others may expect. From this we conclude that it is rational to be disposed to. . .internalizing moral principles." See *ibid.*, p. 170: "To demonstrate the rationality of suitably constrained maximization we solve a problem of rational choice. We consider what a rational individual would choose, given the alternatives of adopting

III. Morality: Preference or Constraint?

The personal attribute of morality can be characterized as a dispositional variable. It reflects a person's general disposition or propensity to act in accordance with moral principles, relatively independently of the specific incentive structure inherent in particular choice settings.

Attempts to account for a dispositional variable like morality in economic explanations seem to be at odds with the standard economic classification of explanatory variables into either *preferences* or *constraints*. As normally understood, this classification corresponds to the "subjective vs. objective" distinction: preferences are the subjective, intrapersonal determinants of choice, and constraints are the objective, external determinants.[13] Given this system of reference, introducing morality into economic explanations as a variable would seem to require us either to classify morality as a subjective, intrapersonal preference-variable or as an objective, external constraint-variable.

Since classifying morality as an external constraint apparently is ruled out by the fact that morality has been explicitly defined here as an intrapersonal, dispositional variable to be strictly separated from external sanctions, the only alternative seems to be to view morality as a preference-variable. This view appears to be implicit in the dichotomy between morality and self-interest which is quite common in discussions of the relationship between ethics and economics.[14] The proponents of such a dichotomy seem to suppose that some modification of the standard economic model of self-interested behavior is needed in order to account for morality, though they are generally not very explicit about the specific kind of modification they have in mind. There seem to exist two views which, though they are not totally different, can be separated. The one view conceptualizes morality as a motivating force which is beyond and outside of any interest calculation, or, in other terms, as a factor that is somehow outside a person's preference function

[13] As I mentioned earlier, the distinction between intrapersonal morality and external sanctions is relevant and meaningful, even if one acknowledges that an individual's (necessarily subjective) *expectation* of sanctions is the direct determinant of his choice.

[14] The dichotomy between morality and self-interest is reflected in statements such as that found in Stigler's "Friendly Economists," pp. 25f.: In "situations where self-interest and ethical values. . .are in conflict. . .most of the time. . .the self-interest theory. . .will win." And this dichotomy is equally implied when McPherson argues in "Limits on Self-Seeking," p. 72, that "people's willingness to be moral" raises difficult questions "for economists who are professionally predisposed to postulate a self-interested basis for action."

and is not captured by the economist's notion of rational choice.[15] The other view conceptualizes morality as part of a person's preference function; i.e., it considers human interests to be broader than "self-interest," as this notion is commonly understood, and to include something like an interest in being a moral person.[16]

The first view is problematic even at the conceptual level. If the principle of rational, self-interested behavior is meant to imply that individuals choose from among potential alternatives the one which they expect to serve their interests best, it is difficult to see how morality can be conceptualized as a factor that makes individuals choose something different from what they think is in their best interest. One would have to assume, it appears, that man has "two natures," an interest-oriented one and a moral one, and that sometimes the one and sometimes the other determines his behavioral choices. Without a more general theory which would integrate the two at some higher level and specify the conditions under which the one or the other of the "two natures" will prevail, explanations of observed behavior in terms of such a model would be totally arbitrary. If, on the other hand, such a higher order theory is specified, the relevant issue would be whether or not this theory is preferable to the self-interest model in terms of theoretical fruitfulness.[17] Until now, it seems, nobody has advanced a potential candidate for such a theory which could actually be compared with the model of rational, self-interested choice.

By contrast with the first view, the second interpretation of the dichotomy between morality and self-interest does not question the basic logic of the assumption that individuals will choose what they consider to be in their best interest. According to this interpretation, the problem with the self-interest model is that the concept of self-interest as it is commonly understood suggests too narrow an interpretation of human interests. The idea apparently is that a more appropriate model of

[15] This view is reflected, for example, in McKenzie's "Economic Dimensions," p. 9: "Exploration of the Christian or Kantian motives for individual action, which makes the 'goodness' of the act itself an end distinct from the gain to the actor, may yield useful predictions about behavior which can complement the economist's insights founded on 'economic' rationality."

[16] This view is expressed, for example, in Terry L. Anderson and Peter J. Hill, "Constitutions and Ideology: Substitutes or Complements?" (Bozeman, MT: Montana State University, 1985, Mimeographed), pp. 7ff. The authors propose a modification of the standard utility function in order to incorporate an "ethical component." The modified approach explains individual choices in terms of a "behavioral objective function" which includes a normal preference function along with an "ethical function," allowing for substitutions between "welfare" and "ethics."

[17] In this sense, what is actually at stake in controversies about the status of the "model of self-interested behavior" is not whether we allow for more or less "self-interest" in our explanations, but whether we preserve a unitary behavioral principle or allow for dual or triple human "natures." Suggestions for changes in the self-interest model should be stated in terms of alternative unitary models of behavior which can be compared in their explanatory power to the self-interest model.

VII. Alternative Approaches: Supergames and Gauthier's Rational Morality

It may be useful to pause here for a look at two approaches to a rational explanation of morality which are similar but not identical to the perspective outlined above.

The one approach is characteristic of all those conceptions which, in one way or another, argue along the lines of the game-theoretical notion of repeated Prisoner's Dilemma games, in particular the notion of a cooperative solution to infinite-horizon supergames.[50] The other approach is the one elaborated by David Gauthier in his attempt to show "why an individual, reasoning from non-moral premises, would accept the constraints of morality on his choices."[51]

The repeated PD game approach and the present approach have in common that they both stress the relevance of the "shadow of the future" in people's choices. The central argument is that in settings where people are likely to meet again, they have good reasons to consider the more indirect and long-term consequences of their choices, and not just the immediate and short-term effects. Properly taking into account those indirect and long-term consequences, however, will let many moral choices appear as rational which otherwise may seem to be "self-sacrificing."

The iterated PD game approach is different from the approach advocated here in that it does not draw the distinction between the "constitutional" and the "situational" level of choice which is so central to the present argument. The iterated game argument remains at the level of particular, situational choices. It aims to show that in particular choice situations in which—if they are analyzed as one-shot games—the defective choice would be dominant, the cooperative choice is the rational one if due consideration is paid to long-term future consequences. In terms of the distinction explained above, it can be said that the iterated PD game approach is about *moral behavior*, not about *morality*. It explains why it may be rational to act morally, but it does not explain why it may be rational to adopt a moral routine.

[50] See, e.g., James W. Friedman, *Oligopoly and the Theory of Games* (Amsterdam: North Holland, 1977); Andrew Schotter, *The Economic Theory of Social Institutions* (Cambridge: Cambridge University Press, 1981).

[51] Gauthier, *Morals by Agreement*, p. 5.

high.[49] And, typically, a person will have the more reason to care about his reputation the more stable and the less anonymous the relevant social environment, and the more his own orientation in this environment is of a long-term character.

[49] To be sure, the potential benefits from "moral reputation" may invite mimicry and attempts to pretend morality and to realize gains from moral reputation without actually bearing the costs of genuine morality. The advantages from genuine rule-following behavior, however, by their very nature, cannot easily be realized by mimicry. To be unmasked as a hypocrite is the most damaging thing that can happen to one's moral reputation, while on the other hand, it is most beneficial to one's reputation if one is observed acting morally in situations where one could not reasonably expect to be observed.

human behavior ought to allow for *other* interests—such as moral preferences—which people pursue along with their more narrow self-interests. That is, self-interest is viewed as a specific subcategory of all interests which an individual might pursue, a subcategory which exists along with other, presumably more "noble" interests.

The interpretation of morality as an element in an extended preference function cannot be accused of the conceptual inconsistencies which characterize the first view. But there are other reasons why economists would typically be reluctant to embrace such an interpretation, reasons that make them generally prefer explanations in terms of constraints over those in terms of differences and changes in preferences. To explain intra- and interpersonal variations in moral behavior in terms of variable moral preferences is, it would typically be argued, not a more satisfactory account than explaining variations in people's fitness activities in terms of their different preferences for exercising. Any explanation of particular behavioral patterns in terms of specific preferences—whether moral or other—is, according to this view, redundant or ad hoc so long as no evidence for those preferences can be provided which is independent of the behavior that is to be explained. And the supposition is that once one allows for such intrapersonal factors as explanatory variables, there will be no limit to inventing special preferences or motives whenever observed patterns of behavior resist standard explanations.

The attempt to incorporate "morality" into economic theory seems thus to lead into a dilemma. One seems to be forced into the choice either to deny that something like morality as an internal, dispositional determinant of behavior exists, or to set aside the methodological objections against explanations in terms of preferences. Before choosing one of the two horns of this dilemma, it may be useful to examine whether there might not be some other way of incorporating morality into economic theory, a way which is more consistent with the standard methodological orientation of economics. The most challenging approach to examine for this purpose is probably Gary S. Becker's reformulation of the theory of consumer choice, in which the economists' argument for relying as much as possible on constraints and as little as possible on preferences for explanations of differences and changes in behavior has been carried to its logical conclusion. Becker's approach is based on the premise that no interpersonal differences or intrapersonal changes in preferences exist and, therefore, differences and changes in behavior are to be explained exclusively in terms of (income and price) constraints. If an intrapersonal, dispositional variable *morality* can be shown to find a consistent place within a Beckerian framework, this should prove most effectively that such a variable can be systematically incorporated into economic explanations.

The remainder of this essay will discuss how a Beckerian perspective may indeed be used to arrive at an interpretation of morality which can be more fruitfully incorporated into economic theory than alternative accounts. Following a discussion of Becker's approach (Section IV), sections V and VI will offer an interpretation of morality which provides an alternative to the dichotomy between morality and self-interest. Instead of viewing morality as a motivational factor, outside or inside a person's preference function, the conception to be advanced here will discuss the dispositional variable *morality* in the context of a person's *production function.*

understood as a general label for all inalienable sources of a person's productivity. And deliberately forgoing potential gains from rule violation may be considered an investment in the human capital of a good reputation. Whether or not the reputation of being a moral person actually is an asset, and whether or not it is worthwhile to invest in it, obviously depends on the character of the relevant environment and on one's own position in it. Like all investments, investments in reputation mean current sacrifices for expected future benefits. Such sacrifices can be rational only to the extent that there is an "expected future." And, where there is a future, the expected payoffs from such investments may vary significantly for persons in different positions. For instance, for a person with an already established "negative reputation" the potential returns from such investments may be quite small compared, e.g., to those of a newcomer who still has to establish a reputation, or to those of a person with an established positive reputation.

Depending on his experiences in certain environments, a person may have more or less reason to expect that "honesty pays." It should be noted that for moral learning, as for human learning in general, *vicarious learning*—learning from indirect (observed, reported, etc.) experience—may play a significant role in shaping people's moral attitudes.[48] And one has to take into account the fact that theories about the payoffs from investment in reputation are typically about the *long run* and therefore not as readily testable as theories about short-run gains and losses from moral behavior. People may well learn systematically to distinguish between environments where they consider an "honesty pays" theory appropriate and other environments where considerations of reputation are considered irrelevant, a phenomenon to which notions like "honor among thieves" refer. The relevant point in the present context is that the more reason a person has to care about his reputation, the more rational it may be for him to adopt the rule always to act morally, because it may simply be too costly to try to identify those situations in which rule violation would not be detrimental and/or because the risk of misidentifying such a situation may appear to be too

[48] See John Finley Scott, *Internalization of Norms—A Sociological Theory of Moral Commitment* (Englewood Cliffs, NJ: Prentice-Hall, 1971), pp. 108f.: "What amounts to an implicit belief that sanctions will follow deviant action has to be reinforced by sufficiently persuasive and present evidence that they do in fact follow. . . . So rewarding does a social animal find a cognitive orientation toward his fellows that attention to what happens to these others. . .is probably impossible to suppress completely. . . . If he observes that others are punished for deviation, his own learned commitment is thereby reinforced; but if he sees that others are not punished for deviation, or are rewarded for it, his moral commitment is weakened. . . . The curiosity of the moral about the immoral is notorious. The degree of moral behavior. . .is a function not only of the sanctions applied to a single person, but also of the sanctions applied to the group of persons who may learn vicariously from each other. Sanction one and you influence all: the good man is kept good not only because of his own rewards, but also because he sees the punishment dealt to the evil."

why morality may be an efficient behavioral routine, since formal enforcement nowhere is—and, for reasons of cost, never can be—tight enough as not to permit a broad range of situations where rule violations would remain the more advantageous alternative. Taking informal enforcement into account will allow some narrowing of the range of such "opportunities to cheat," but it is common opinion that there will always remain a more or less significant set of situations in which formal and informal external sanctions alone typically cannot be assumed effectively to eliminate the temptation "to cheat." The question of how moral behavior in this kind of situation can be explained, and in particular whether recourse to a dispositional variable *morality* can be an appropriate and fruitful explanatory approach, is exactly the issue raised at the beginning of this essay.

Can adopting a rule "always to act morally" be rational for an actor, even if he expects that there will be situations in which nonmoral behavior would be more advantageous? *It is rational, if he perceives the costs of identifying the relevant subset of situations or the risk of misidentifying a situation as high enough.* Whether this is the case or not will crucially depend on the actual structure of the social environment in which a person is acting, on his own "position" within this environment, and on his relevant theories about "the world." If an actor's relevant theories imply, for instance, that the essential remuneration for moral behavior is transcendental, then to him morality can be a rational strategy whatever the actual structure of his social environment, although the probability that an individual will adopt and maintain such a theory will not itself be independent of his social environment. Theoretically more challenging, however, is the case in which the theories on which the rationality of morality depends are at least in principle testable. In this case, if one rules out systematic and permanent misperceptions as a relevant source for differences in people's theories, such differences would have to be explained in terms of differences in the environments which condition people's theories.

Apart from immediate disadvantages resulting from formal or informal sanctions, a main interest that may motivate an actor to act morally is a concern about his *reputation.* To have the reputation of being a moral person, i.e., a "man of principle," may considerably increase an individual's productive efficiency, i.e., his ability to produce the ultimate objects of choice. The reputation of being a moral person can be viewed in this sense as a kind of human capital, where "human capital" is

p. 9: "As rational persons understanding the structure of their interaction, they recognize a place for mutual constraint, and so for a moral dimension in their affairs. . . . Agreed mutual constraint is the rational response to these structures." Gauthier's approach to a "rational explanation of morality," which is similar but not identical to the approach taken here, will be examined in more detail later in this essay.

IV. Morality and Choice: Theories as Subjective Constraints

Becker's proposal for a reformulation of the theory of choice is motivated by precisely the problem discussed above: that explanations of differences or changes in people's behavior in terms of preferences are bound to be arbitrary since, as Becker supposes, "no useful theory of the formation of tastes" exists in the social sciences on which one could rely in "choosing the appropriate taste proxies. . .or in formulating predictions about the effects of these variables on behavior."[18] The remedy Becker suggests is to "treat tastes as stable over time and similar among people"[19] and to explain differences or changes in observed behavior exclusively in terms of differences in price and income constraints, considering that the relevant prices and incomes may take quite subtle forms and include more components than money prices and money income. What allows Becker to shift all of the explanatory burden to price and income constraints is essentially a narrower definition of preferences and the reinterpretation of the household (consumer) as a *producing unit.* The "preferences" which Becker assumes to be "stable and identical" for all individuals are not specific tastes for particular goods, as is suggested by the common use of the term. Rather, people just as human beings are assumed to share certain fundamental preferences or "basic pleasures."[20] The consumer is viewed as producing "commodities," i.e., ultimate goods (such as health, status, etc.) from which he directly obtains utility. In producing these "final objects of choice,"[21] the consumer–producer is using as inputs his own time and resources, along with goods and services purchased in the market. An essential implication of this conception is that the demand for market goods is to be modeled as *derived demand*, comparable to a firm's demand for factors of production.[22]

The consumer's (household's) utility (U) is a function of the commodities (Z), the "ultimate objects of choice":

[18] Becker, *Economic Approach*, p. 133.
[19] Stigler and Becker, "De Gustibus," p. 76.
[20] Becker, *Economic Approach*, p. 145.
[21] *ibid.*, p. 146.
[22] *ibid.*, p. 134.

$$U = u(Z_1, Z_2, \ldots, Z_n).$$

Since, by assumption, interpersonal differences or intrapersonal changes in the utility function are ruled out, systematic differences in people's behavior which cannot be accounted for in terms of conventional income and price constraints are to be explained in terms of differences or changes in people's production methods or, in Becker's terms, in their productive efficiency with regard to the Z-commodities.[23] The consumer's (household's) production function for the commodity Z_i is:

$$Z_i = z_i(x_i, t_i; E)$$

where x_i refers to market inputs, t_i to the consumer's own time input, and E to "*environmental variables*" which, in particular, are supposed to reflect the production *technology* employed by the consumer–producer.

In the present context, the question about Becker's reformulation of the theory of (consumer) choice that is of interest is whether and, if so, how a variable *morality* can be incorporated into a Beckerian framework.[24] If morality is considered to be a *variable* which may account for systematic differences in people's behavior, it obviously cannot be a preference in the Beckerian sense. And since, for the reasons I mentioned earlier, it cannot be interpreted in terms of conventional income and price constraints, the only category under which it can be subsumed, if it can at all, is the environmental variable E. It is therefore the specification of the E-variable which is crucial for our present purposes.

Becker's explanations of E (or its equivalent in the slightly different version of the household production function in the work of Stigler and Becker) are not very extensive and systematic. It is obvious, however, that although they are called "environmental," the E-variables are not strictly *external* variables which the observer–analyst could discover in the "environment in which the production takes place."[25] Actually, the notion that the environmental variables reflect "the level of technology of the production process"[26] is ambiguous in this respect. In conventional production theory for the firm, technology is typically viewed as an external variable which is *given* and is the same for all relevant firms. Technology, therefore, cannot account for systematic differences in the behavior of firms. Such a view is justified since technology is, in

[23] *ibid.*, p. 145: "One might argue that indeed all households have precisely the *same* utility function. . .that they all derive that utility from the same 'basic pleasures' or preference function, and differ only in their ability to produce these 'pleasures'."

[24] The relevance of Becker's approach is also discussed, though with different conclusions from those which are drawn here, in Alexander Rosenberg, "Prospects for the Elimination of Tastes from Economics and Ethics," *Social Philosophy & Policy*, vol. 2 (1985), pp. 48-68.

[25] Becker, *Economic Approach*, p. 135.

[26] *ibid.*

expect a pattern of rule-constrained choices to serve their long-term interests better than case-by-case choices made under "temptations" of the immediate choice situation.[45]

That, for the reasons mentioned, rule-governed behavior may be preferable to a strategy of case-by-case "optimization" does not imply, of course, that just any rule may be adopted. There are many conceivable behavioral rules which obviously would produce a pattern of outcomes clearly inferior to what could be realized by deciding each case on its own. The preceding arguments, therefore, do not by themselves imply that it must be rational for an individual to adopt *moral rules*. An individual's direct "moral interest" is not in his own obedience to rules but, instead, in the obedience of others. His own obedience directly benefits others, and an individual's own immediate interests would be best served if he were allowed to ignore rules while others were bound by them. Therefore, to explain why morality can be an efficient behavioral technology requires more than simply showing why adopting behavioral rules may be rational. What has to be shown is why it can be rational for an individual to adopt a rule which appears to imply self-sacrificial demands.

It seems obvious that morality can be considered an efficient behavioral technology by an actor only if he perceives the relevant *typical* choice situation not to have Prisoner's Dilemma characteristics.[46] If an actor does assume that in typical choice situations respect for moral rules is not in his best interest, it simply cannot be rational for him to adopt the rule always to act morally. That is, if morality is to be a rational strategy, the actor must perceive the relevant typical choice situation to be different from the setting that has been described above. Our analytical attention has to be focused, therefore, on the reasons why an actor may have such a different perception.

The obvious answer is that in ordered social communities, we will typically find some kind of enforcement mechanisms which are designed to "correct" the expected payoffs from rule violation and rule obedience so as to make the latter the rational choice. After all, the standard argument for how a group of individuals may escape recurrent social dilemma situations that take the form of a Prisoner's Dilemma is that they may agree jointly to submit to constraints that eliminate the temptation to "cheat."[47] But this certainly is not sufficient to explain

[45] See, e.g., J. Elster, *Ulysses and the Sirens* (Cambridge: Cambridge University Press, 1979); Thomas C. Schelling, *Choice and Consequence* (Cambridge and London: Harvard University Press, 1984).

[46] As James Buchanan argues in "Ethical Rules," the "moral dilemma" referred to here can be more precisely discussed as a "large-number dilemma."

[47] With reference to those kinds of social situations, Gauthier argues in *Morals by Agreement*,

different rule), the relevant comparison is, of course, not between specific single outcomes but between patterns of outcomes over a sequence of choices. That is, it is the overall performance of alternative strategies which counts, where "to decide each case on its own merits" can be considered one strategy. The potential benefits from adopting a rule for making certain kinds of decisions can be seen to result from avoiding the disadvantages of case-by-case decisions, for which there seem to exist at least three potential sources:[41] first, the decision-making costs such as the costs of collecting and evaluating the information needed in order to decide each case; second, the risk of mistakes resulting, for instance, from inadequate information; third, the tendency of case-to-case decisions to be biased in favor of short-term interests at the expense of long-term interests.

There exist a number of theoretical contributions on the "reason of rules"[42] issue which stress the one or the other of these sources. The first source figures prominently in A. Gehlen's theory of institutions as cultural substitutes for instincts, which allow for routines and relieve human actors from the burden of being constantly required to make decisions.[43] The second source is stressed in R. Heiner's pertinent and informative essay on the "Origin of Predictable Behavior." Heiner reviews a broad selection of relevant theoretical sources which all support—in one way or another—the general notion that "rule-governed" behavior, such as instincts, habits, customs, and norms, can be viewed as an adaptive response to environmental complexities and uncertainties which make it difficult for an individual in recurrent choice situations to select the most preferred alternative; such complexities and uncertainties make rule-governed behavior in recurrent situations an overall useful strategy. To adopt behavioral rules means to adopt "patterns of behavior for which deviations exist that are preferred under certain conditions, but which are nevertheless ignored because of uncertainty in reliably interpreting information about when to deviate."[44] The third source, finally, is the central theme in the so-called "theory of self-management, or self-control" which emphasizes that individuals may rationally choose to adopt a rule constraining their choices if they

[41] It should be kept in mind that what I discuss here are the potential advantages to an actor of adopting a rule for himself, that is, advantages he realizes by *his own* rule obedience. This is to be distinguished from the totally different issue of what are the potential advantages to an individual of *other people's* rule obedience, that is, the advantages from living in a social community where certain rules are practiced.

[42] Geoffrey Brennan and James M. Buchanan, *The Reason of Rules: Constitutional Political Economy* (Cambridge: Cambridge University Press, 1985).

[43] Arnold Gehlen, *Studien zur Anthropologie und Soziologie* (Neuwied a.R. und Berlin: Luchterhand, 1963).

[44] Ronald A. Heiner, "The Origin of Predictable Behavior," *The American Economic Review*, vol. 73 (1983), p. 572.

fact, "out there" as "objective knowledge," in the sense of Karl Popper's theory of a "third world" of objective ideas.[27] It can be assumed that firms will systematically tend to adopt efficient technologies once they are "out there," if they are competing in markets in which the user of less efficient technologies would not be able to survive.

Whatever one might assume about competition among consumers (households), they certainly do not compete as "producing units" in the same sense that firms do. They produce for *themselves*, not for a market. And different technologies can be expected to have a much greater chance to survive in such a context. In fact, there seems to be little reason to expect a simple and uniform connection to exist between the technologies "out there" and the technologies actually adopted by different consumer–producers. Rather, it seems to be perfectly reasonable to assume that differences in adopted technologies may indeed account for systematic differences in consumer–producer behavior.

Quite apart from this, the variable "production technology" *has* to be interpreted in a way which allows for systematic differences between consumer–producers if it is to serve its explanatory purposes in Becker's approach. And Becker, in fact, specifies that the technology variable includes such things as consumer knowledge ("whether real or fancied"), skills, and human capital in general,[28] things which obviously vary from person to person. Clearly, however, with such variables as knowledge and human capital construed as constraints, the *preferences vs. constraints* dichotomy can no longer be assumed to correspond to the *intrapersonal, subjective vs. external, objective* distinction. Knowledge and human capital are genuinely *intrapersonal, dispositional* variables, and in this respect, they are more like preferences than like constraints in the conventional sense.

Becker's reformulated theory of choice does not substitute an "objective" constraint variable for a "subjective" (preference) variable. Rather it shifts the burden of explanation from one kind of intrapersonal, dispositional variable to some other kind. Whether this is only a matter of semantics and camouflage or an actual analytical improvement[29] obviously depends on whether it can be assumed that the relevant dispositional variable in the production function context is not—or, at least, is less—subject to the deficiencies that make explanations in terms of preferences so unattractive to economists. In fact, Becker apparently assumes that for production function variables like knowledge and human capital, we are in a much better position to identify appropriate

[27] Karl R. Popper, *Objective Knowledge—An Evolutionary Approach* (Oxford: Clarendon Press, 1972).

[28] Becker, *Economic Approach*, p. 145; Stigler and Becker, "De Gustibus," pp. 77ff., 82, 84.

[29] Becker, *Economic Approach*, p. 144.

proxies and to formulate predictions about their effects on behavior than we are in the case of preferences.[30]

The human capital component in the consumer production function can be said to reflect a person's inalienable resources which, to the extent that they are not genetically inherited, have been acquired or learned over the person's lifetime. If one ignores the possibility of relevant inborn differences, systematic differences in people's human capital endowments must reflect systematic differences in their "history," previous experiences, investment decisions, and so forth. By their very nature, human capital variables are, as a rule, not directly observable. One typically has to rely on indirect evidence, either in the form of observable indicators or in the form of information on relevant aspects of a person's history (e.g., on a person's education), together with some general assumptions about how various aspects of a person's history translate into his current human capital.

It should be noted that, once one is willing to allow for "subjective constraints" and to rely on dispositional variables for which only indirect evidence can be obtained, the economist's standard methodological objections against explanations in terms of preferences are no longer conclusive. Methodologically, the problems faced by an explanation in terms of subjective constraints (a person's theories) are not different, in principle, from those faced by an explanation in terms of preferences. What are needed in either case are additional theoretical assumptions which specify independent evidence for supposed differences and changes in the intrapersonal variable. To say that people demand certain market goods because they perceive them as being instrumental in obtaining ultimate Z-goods focuses attention on people's theories about the relevant characteristics of goods and on their theory-learning in explaining those variations in revealed demand which cannot be readily explained in terms of conventional income and price constraints. It would seem to be theoretically equivalent to this to argue that people's preferences for market goods are *derived* preferences, derived from their invariable, ultimate preferences. These derived preferences can be assumed to vary, depending on what individuals learn—by direct or indirect experience—about the relation or association between market goods and Z-goods. Whether stated in terms of theory-learning or in terms of preference-learning, the substantive theory that seems to be needed should not differ very much between the two approaches.

Though Becker and others assert that there is no useful theory of the formation of preferences,[31] there exists, at least, a behaviorist learning theory developed by B.F. Skinner and others; by referring, like the pre-

[30] *ibid.*; Stigler and Becker, "De Gustibus," p. 89.
[31] Becker, *Economic Approach*, p. 133.

VI. The Calculus of Morality

Can it be rational to be *always moral* if it is *not always rational* to be moral? This, stated in a seemingly paradoxical way, is the issue to be discussed in this section.

An analysis of morality as it is understood here (i.e., as a general disposition to act in accordance with moral rules) has to draw a careful distinction between a person's situational or particular choice (whether deliberate or not) in a particular situation to respect or not to respect moral rules, and the "constitutional" choice to act morally in general.[38] It is trivial to suspect that in particular situations it may be expedient for an individual to act in accordance with moral rules. And it is equally trivial to suspect that, considering each situation individually, conformity with moral rules will typically not always be rational in terms of expected payoffs.[39] The interesting issue is whether it can be rational for an individual to choose (deliberately or habitually) always to obey moral rules, knowing that there will be situations in which his payoffs would be greater from nonmoral rather than from moral behavior. The following discussion of this issue will proceed in two steps. It will, first, address the general issue of why it can be rational for an individual to adopt *rules* rather than to decide each case individually, trying to choose the "optimal" alternative. And it will, second, specify the reasons why this may also be true for the particular category of rules that is of concern in the present context: moral rules.[40]

If the benefits from following a rule are to be compared to the benefits from deciding each case individually (or to the benefits from adopting a

[38] The notion of an individual's "constitutional" choice is used here in analogy to the familiar notion of constitutional *collective* choice. In the latter context the term "constitutional" refers to the rules for making collective decisions. In the same sense, an individual's constitutional choice is his choice of rules for making certain kinds of individual decisions.
[39] This may be different for a person who firmly believes in transcendental remuneration for moral behavior.
[40] As I will discuss in more detail later, the issue addressed here is different from the one to which game-theoretical arguments on iterated PD games are supposed to give an answer. Those arguments show why in recurrent game settings the "shadow of the future" may make it rational to choose the cooperative strategy in a particular move. The issue discussed in this essay is why it can be rational for an actor to adopt a moral or cooperative *routine* instead of calculating the expected benefits and costs in each and every choice situation. Another way of stating the difference is to say that the repetitive-game arguments are about situational choices, while the argument to be elaborated here is about constitutional choices.

what are "efficient technologies" to reflect their direct and indirect experiences with potential alternative technologies.[36] And their theories can be assumed to be the more homogeneous the more similar their past experiences and vice versa.[37] In particular, with respect to the issue of morality, it should be noted that one relevant dimension along which people's theories of "efficient technologies" may vary is their testability. There are certainly technology theories which are more readily testable than others, and therefore prone to be eliminated if they are incorrect. For instance, theories about short-term payoffs from morality will be typically more easily testable than those about ultimate, long-term payoffs, and theories about what will produce more material wealth are certainly more readily testable than theories about what produces greater happiness. And it can be assumed that the less testable the relevant theories are, the easier it will be for interpersonal differences to persist (which is not the same as saying that less testable theories as such are more likely to persist than more readily testable ones). Theories about whether morality "pays" may even be inherently immune to falsification—for instance, if the relevant remuneration for moral conduct is expected from some transcendental entity.

The interpretation of morality as a theory component in a Beckerian consumer production function explains the choice of the title for this essay. Whether one understands the Latin idiom "De gustibus non est disputandum" as saying that differences in preferences cannot be rationally disputed, or, in the Stigler-Becker interpretation, as implying that such differences do not exist, with morality things are different in both respects. *De moribus est disputandum*: People obviously differ in the extent to which they exhibit the behavioral disposition called "morality," and these differences can be rationally analyzed as reflecting systematic differences in their (explicit or implicit) theories about what are efficient ways to produce the ultimate objects of their preferences.

[36] Taking into account the impact of people's past experiences on their present theories has implications for the concept of rational choice: people's choices can be assumed to be rational, *given* their theories. People's theories can be assumed to be "rational" in the sense that they reflect the character of the environment in which they have been formed. This allows, however, for people's theories to be different and not equally adequate within the same kind of *current* choice situation.

[37] Conventional applications of economic analysis typically focus on sufficiently "homogeneous" populations or define the issue in a sufficiently limited way (e.g., "what happens if the 'price' of activity A increases?") that relevant cultural or other differences in people's theories can be ruled out. This is an essential condition for the relatively satisfactory explanations these standard applications can provide by relying exclusively on "external" constraints.

ceding argument, to a person's learning history, it provides some systematic information on the formation of "secondary reinforcers," the behaviorist's term for the kind of phenomena which in economics are traditionally called "preferences."[32] Certainly, one may consider this kind of psychological theory to be unsatisfactory for the explanatory purposes of economics. But it might not be so easy to demonstrate that these theories are in principle inferior to the "implicit learning theory" on which Stigler and Becker necessarily rely for their own assumptions about how a person's previous consumption of "music appreciation" or heroin translate into his current "productive efficiency."[33] On the other hand, adopting insights from behavioral psychology may provide economics with a new grip on some of its notoriously puzzling theoretical issues, a new grip which is—as some of the following arguments are supposed to indicate—perfectly in line with the core thrust of the economic paradigm.[34]

Where human behavior is to be explained, we cannot avoid coping with subjective determinants of choice. Simply to rule out, by definition, the possibility that preferences are variable does not eliminate the subjective dimension. Its only consequence may be that it requires us—as is obviously the case with Becker's approach—to define the constraint-variables in a way that makes them include genuinely subjective, intrapersonal variables like people's *theories* about the world. There is no a priori reason, however, to assume that explanations in terms of such subjective constraint-variables are somehow methodologically superior to those in terms of preferences. Whether the one or the other is preferable is a matter of fact, not of principle, and it is easily conceivable that for some purposes the one, and for other purposes the other, approach may prove to be more fruitful.

However, although I have criticized the economist's standard view that explanations in terms of constraints are *in principle* superior to those in terms of preferences, the following discussion will proceed upon the presumption that, with respect to the issue of morality, an interpretation in terms of subjective constraints—or in terms of people's

[32] Behavioral psychology makes a worthwhile distinction that is quite compatible with Becker's notion of stable, fundamental preferences: the distinction between primary reinforcers which are essentially the same for all human beings because of common genetic characteristics, and secondary reinforcers which vary between persons and change over a person's lifetime because of differences and changes in learning experiences.

[33] Stigler and Becker, "De Gustibus," pp. 79ff.

[34] In "Rational Egoism versus Adaptive Egoism as Fundamental Postulate for a Descriptive Theory of Human Behavior," his presidential address to the 1986 Public Choice Society Meeting, Dennis C. Mueller pleaded for modeling "economic man" more along the lines suggested by behavioral psychology. In *Die zwei Soziologien—Individualismus und Kollektivismus in der Sozialtheorie* (Tuebingen: Siebeck (Mohr), 1975), I have discussed in some detail the theoretical affinities between a behaviorist and an economic, rational choice approach in social science.

theories—does, in fact, provide a more fruitful theoretical perspective than one in terms of preferences. Asking what affects people's explicit and implicit theories about the payoffs from morality seems to lead into a more promising avenue of inquiry than asking what shapes people's preferences for being moral.

V. *De Moribus Est Disputandum*

Morality has been defined above as a general disposition or tendency of a person to act in accordance with moral rules, relatively independently of the particular circumstances the person faces in various choice situations. Morality is considered to be a dispositional variable which can account for differences or changes in observable behavior which appear to be unexplainable in terms of conventional "external" constraints. As I discussed above, what is needed in order for such a dispositional, intrapersonal variable to allow for more than merely arbitrary ad hoc explanations is some kind of connecting theory which establishes a systematic link between assumed states of the dispositional variable and some kind of evidence which is independent of the behavioral phenomena under explanation.

To argue that morality is a determinant of behavior that is distinct from external constraints is not the same as assuming that morality is independent of any external influence. A person's *future* morality will be affected (strengthened or weakened) by current constraints the person experiences, just as his *current* morality has been shaped by previous experiences. A person's morality can be said to reflect some generalized—explicit or implicit—expectation or theory about the consequences which may result from rule obedience or rule violation.[35] In the consumer production function context these expectations can be viewed as "technology" components or, more specifically, as a person's knowledge or theories—whether explicit or implicit—about what kind of behavioral patterns are conducive to producing the ultimate objects of choice.

That the theory components of the consumer production technology reflect the (necessarily *subjective*) theories of a person does not imply that they allow for unlimited idiosyncracy. Although, as I mentioned above, the competitive pressure that tends to homogenize firm production technologies in a market context does not constrain consumer production, choice of technology in the latter context is certainly not arbitrary. If we presume that people are informed about their relevant, more immediate environment, we should expect people's theories about

[35] The phrase "explicit or implicit" is used here in order to indicate that the relevant "expectations" or "theories" may be habitual and that the actor may not be able to articulate them.