Morality, Maximization, and Economic Behavior

Alan G. Isaac
American University
Washington, D.C., USA

15 Aug 1995

This is the working paper version of "Morality, Maximization, and Economic Behavior", published in the Southern Economic Journal 63(3), January 1997. Please cite the published paper.

Abstract

This paper clarifies the relationship between morality and self-interest, defends extended-utility representations, and explores two problems for the neoclassical paradigm: morality may generate inconsistencies in behavior due to a lack of coherence in the goals or the rules that govern behavior, and economists may "observe" inconsistencies due to interpretational inadequacies of their paradigm. The first problem is a potentially radical empirical challenge. Extended utility representations defeat important aspects of the second problem, thereby allowing empirically oriented economists to pursue a unified treatment of motivations, but they elevate the role of interpretation in the economic modeling of human behavior.

Morality, Maximization, and Economic Behavior*
Alan G. Isaac
American University
Washington, D.C.

1 INTRODUCTION: MORALITY AND ECONOMIC BEHAVIOR

Morality influences economic behavior. It has been invoked to explain worker solidarity, statesmanship, voting patterns, cooperative behavior, bargaining outcomes, the production of public goods and externalities, worker performance, and the degree of competition in labor markets [1, 10, 14, 19, 21, 26]. So economists ought to consider the economic consequences of prevailing moral norms. In some areas such considerations will be distinctly secondary (e.g., the study of arbitrage pricing relationships). But in many areas that economists traditionally consider to be in their purview—including the study of labor markets, organizational design, income distribution, long run growth, and the provision of public goods—economic models can be improved by the accommodation of moral behavior.[1]

However, the ability of the neoclassical paradigm of rational economic behavior to accommodate moral behavior remains controversial. Prima facie, moral motivations can generate behavior that is not profit maximizing or utility maximizing. A primary goal of this paper is to clarify the relationship between moral motivations and the neoclassical paradigm. I trace some of the controversy in this arena to incorrect or misleading formulations of key issues, as in the debates over the necessity of multiple utility representations. Nevertheless, the neoclassical paradigm encounters two major challenges in accommodating moral behavior: actual inconsistencies in behavior due to a lack of coherence in the goals or the rules that govern behavior, and apparent inconsistencies due to the interpretational inadequacies of the neoclassical paradigm.[2] A second goal of this paper is therefore to explore the nature of these challenges and describe some responses to them.

2 RATIONALITY, RULES, AND THE NEOCLASSICAL PARADIGM

Moral commitments are often embodied in behavioral rules, such as “keep promises” or “respect property rights”. Some writers claim that the neoclassical paradigm of rational choice cannot accommodate rule-governed behavior in general and morality-governed behavior in particular. Of course, rule-governed behavior is not *necessarily* an empirical problem for the neoclassical paradigm, even if we eliminate rules such as “be rational” or “optimize”. For example, an optimal consumption plan—eminently compatible with the neoclassical paradigm—might be embodied in a rule tying expenditure patterns to the current economic environment. A critique based on the prevalence of rule-governed behavior therefore must specify a domain of application. This section offers a specification that provides qualified support for the critique.

Define 'rules of conduct' to be (tacit or explicit) behavioral guidelines that one might readily imagine replacing with (possibly inferior or superior) alternatives. I refer to actions selected (tacitly or explicitly) for their accord with rules of conduct as 'rule-governed'. The adherence to rules may or may not be rigid; the value of adherence to rules may be lexically prior to other values or may be entirely derivative.[3] I will call an action 'strongly rule-governed' when it is influenced by a rule of conduct independently of aspects of the choice situation that might justify or vitiate a particular application of the rule.[4] Alternatively, rule-governed actions might consistently manifest an integrated preference for rule adherence (e.g., to rules embodying social or moral norms); I call such actions 'weakly rule-governed'.[5]

Let 'behavior' denote the observed manifestation of action. Weakly rule-governed behavior may respond to the full particularities of a choice situation, while by definition strongly rule-governed behavior does not. Many of the economic rules actually followed by individuals appear to be strongly rule-governed. For example, consumers often establish fairly rigid budget categories to constrain their spending, and many firms set prices according to markup rules. Such behavior poses obvious explanatory challenges for the neoclassical paradigm.

Let 'the rules critique' denote arguments that the neoclassical paradigm of economic rationality is incompatible with rule adherence.[6] The rules critique is quite plausible when applied to strongly rule-governed behavior. The critique fails, however, when applied to weakly rule-governed behavior. In addition, the critique may fail empirically if "irrational", strongly rule-governed behavior satisfies the abstract consistency requirements that constitute the neoclassical paradigm's most fundamental characterization of rational choice.

The neoclassical paradigm has emphasized that the efficient pursuit of stable, well-defined, context-independent goals should generate certain consistencies among choices as the feasible actions change. This consistency has been characterized as "internal" insofar as it refers only to feasible sets and choices, not to "external" criteria such as underlying goals or motivations [25]. For example, consumption choices deriving from maximization of a stable utility function must satisfy the generalized axiom of revealed preference, which refers explicitly only to feasible sets (budget sets) and choices (consumption bundles). It is well known that choice behavior satisfying internal consistency of choice can be given a maximizing interpretation.

Internal consistency of choice does not imply rationality in any broad sense: many types of patterned behavior, such as rigid or habitual behavior, may produce internal consistency of choice.[7] This insufficiency is seldom seen as a defect of the neoclassical paradigm, however. For example, the revealed preference literature originally aspired to discard even the "vestigial" mental categories of utility and preference [15, 18, 22]. While such behaviorist aspirations were foredoomed (see section 3), the fundamental prediction of the neoclassical paradigm of rational choice remains this behavioral consistency.[8]

Given the emphasis on behavioral consistency, logic alone does not imply that strongly rule-governed behavior is problematic for empirical applications of the neoclassical paradigm. If observations of stable, consistent behavior result, then we can generate a maximizing interpretive framework *ex post*. Our framework will sacrifice understanding and descriptive accuracy, but—as these goals are secondary for most neoclassical economists—an argument for radical change in the neoclassical paradigm must include evidence of foregone predictive power or policy control.

However, we can extend the rules critique by focusing on the interpretive framework that undergirds the neoclassical paradigm. Even consistent strongly rule-governed behavior creates two problems for this interpretive framework. First, despite the paradigm's behaviorist tendencies, the open acknowledgment that behavior is not actually motivated by utility maximization and profit maximization generates internal tension.[9] Second, such rules are likely to generate behavior inconsistent with this interpretive framework. For example, it is difficult to reconcile markup pricing rules with profit maximizing behavior, and moral rules remain problematic for interpretations of human action as rationally self-interested.

In contrast to the initial critique, the modified rules critique does apply to weakly rule-governed behavior. Behavior may deviate from the consistency predicted by the neoclassical paradigm due to interpretive failures on the part of the observer. That is, the source of inconsistency may be the observer rather than the actor. As a result, even rules that select efficient means to stable goals can present empirical challenges to the neoclassical paradigm. This is the basis of an argument that economists must attend to moral motivation, which I present in the next section.

Given an individual's rules of conduct, we have seen that weakly rule-governed behavior admits a formal reconciliation of rule-governed behavior with the core neoclassical notions of rationality, whereas strongly rule-governed behavior creates a difficulty for these notions. Less obviously, we found that both weakly and strongly rule-governed behavior can be problematic for the neoclassical interpretive framework. In sum, while there is no logically necessary conflict between rule-governed behavior and rational choice modeling—given the restricted criteria of rationality informing the neoclassical paradigm—reconciliation of actual rule-governed behavior with rational choice modeling appears problematic. Since moral commitment is often embodied in behavioral rules, these arguments identify one potential conflict between morality and rationality. Setting aside concerns about the fundamental conflicts inherent in strongly rule-governed behavior, we find that morality can still create problems for the interpretive framework of neoclassical economics. For example, neglect of the value of rule adherence can cause predictive failures. Problems with the notion of internal consistency of choice rule out a behaviorist response to this difficulty: economists wishing to accommodate moral behavior must grapple with actual moral motivations.

3 MOTIVATION AND ECONOMIC MODELS

The neoclassical paradigm involves more than abstract maximization: it also involves particular specifications of the maximanda. A crucial ingredient of the neoclassical paradigm is its restriction of admissible interpretations. For example, the theory of the firm assumes a profit maximization objective. Similarly, the theory of the consumer effectively assumes utility maximization, with many restrictions on what provides utility.[10] The present section argues that the maximization hypothesis derives its empirical content from these restrictions, that economists therefore must make substantive motivational assumptions, and that they should include moral motivations in their analyses.

Many authors characterize the core of neoclassical economics in terms of abstract maximization or its equivalent. At this level of abstraction, it appears that the neoclassical economist has no commitment to particular motivations. Robbins [20, 95] gave what is probably the most famous statement of this view: "So far as we are concerned, our economic subjects can be pure egoists, pure altruists, pure ascetics, pure sensualists or—what is much more likely—mixed bundles of all these impulses." In their recent survey of the implications of ethics for economics, Hausman and McPherson [14] agree. However they include a caveat: "[T]he *objects* over which preferences are defined may have to be quite different from the familiar range of economic commodities if the preferences are to rationalize moral choices." For example, a simple list of feasible commodity bundles may no longer capture the essence of a consumer choice situation, which may be influenced by a perceived moral dimension (which will in turn depend on social norms and personal history). Boycotts, so popular in the U.S., illustrate this—beer, fruit, and newspaper demands have shifted in response to news about labor practices.

Problems accompany the recognition of a moral dimension to consumer choice, since utility theory—if it is to be useful empirically—must incorporate an intersubjective specification of the objects of preference. For example, applied economists must agree on observable proxies for moral variables. The theory of the firm is equally vulnerable to difficulties with moral entrepreneurs. Finding intersubjectively verifiable correlates to prevailing norms is a challenging task. It requires economists to include historically specific institutional and normative detail in their models of economic action, and it requires economists to understand the motivations of the actors they are modeling. That is, economists will be forced to embrace areas of inquiry that they have been inclined to repudiate.[11]

The most explicitly behaviorist strand in the neoclassical economics literature was the early research on revealed preference. More recently, research on expected utility theory similarly attempted to avoid inquiry into motivation. But both approaches completely failed to discard concepts of belief and preference: the very notion of behavioral consistency implicitly refers to stable beliefs and preferences [24]. In addition, attempts to dodge issues of motivation generated a neglect of the empirical problems arising from the necessarily subjective characterization of any choice situation.

These problems are of varying degrees of importance to the empirically oriented economist. The first and most fundamental problem is epistemological: how can the agent and the observer achieve a common understanding of the choice situation? The import of this question derives from a necessary condition for utility theory to generate observational implications: it must be supplemented with specific claims about motivations. Peeling the motivational assumptions off the (implicit) core of abstract maximization yields a "theory" that is without empirical content. When stated so abstractly, this consideration may appear ludicrously irrelevant to pragmatic empirical work. Yet if taken as an admonition that striving for such common understanding is crucial to good empirical work (at least in certain areas), the epistemological critique motivates a concern with the role of morality in economic behavior.

Finally, moral (and other) norms may be a source of context dependent preferences, which pose additional difficulties for the neoclassical paradigm. For example, Sen [25] posits a person who, when offered cake, always chooses the second largest piece available; this satisfies the social convention "never pick the largest piece". Sen's point is that the consistency in this behavior—which adequately achieves the individual's integrated goals—will not appear to an observer seeking the internal consistency among choices predicted by neoclassical theory. If the observer characterizes the available choices in terms of the sizes of the pieces, for example, whether a piece is second largest depends on the context of the rest of the available choices.

A more telling example might be the maxim "maximize profits within the spirit of the law", which will generate different behavior in different institutional settings. For example, legal definitions and interpretations of unfair restraint of trade vary widely across time and across countries. When the norm of law-abiding behavior guides individual actions (either strongly or weakly), such institutional variation will influence behavior in ways not captured by traditional cost-benefit calculus. Behavioral variation will have an institutional correlate that will be missing from a neoclassical model.

Not only is utility maximization not a testable theory about human behavior, it is not even a guide to the interpretation of human behavior until it is supplemented with particular interpretational constraints. The alternatives in a choice set cannot be characterized without interpretation—interpretation that implicitly involves the goals of the chooser. Without an understanding of the normative environment in which the chooser operates, for example, an observer may completely mischaracterize the the choice situation. Aside from the loss of descriptive accuracy and understanding, such mischaracterization can compromise such goals as prediction and control. Even economists who otherwise downplay the role of understanding should therefore be interested in modeling the influence of normative rules. As the cake eater and boycott examples illustrate, normative behavior is often context dependent in a way that is problematic for the empirical content of the neoclassical paradigm.[12]

These examples show that concern with motivation based on moral commitment is just part of a *necessary* concern with motivation. Even a strict methodological adherence to the self-interested rationality embodied in the neoclassical paradigm *requires* a characterization of motivation. Morality is highlighted only contingently, i.e., because of its salience in human affairs. Since action often reflects moral considerations, behavioral predictions should draw on our understanding of the moral significance, for the actor, of the different feasible actions.

Of course we can always presume that any behavior, no matter how non-maximizing it appears, is a maximizing action. This may be a productive research strategy, *if* it leads us to a motivation that rationalizes the behavior under study. It seems we could on the same pragmatic grounds presume that any behavior, no matter how moral it appears, is self-interested. However, the examples in this paper—along with many others—suggest the limitations of that neoclassical project. So even economists who accept utility maximization on *a priori* grounds will have an interest in moral motivation—perhaps in the form of moral tastes—when morality affects the behavior under study.

In sum, economists cannot escape specifying the sources of motivation in their models of economic behavior. Further, a brief consideration of normatively guided economic behavior highlights some interesting empirical problems. Since moral norms are an important source of motivation, economists should explore formal characterizations of moral motivation that are sensitive to these empirical problems. The next two sections reconsider some possible difficulties for the neoclassical paradigm when moral as well as self-interested motivations are recognized. Since economists must attribute motivations to economic actors, they must also decide which motivations to attribute, how to model individuals with multiple motivations, and whether conflicting motivations can be reconciled.

4 MORALITY AND SELF-INTEREST

Whether moral action can deviate from rational action is an interesting and complex question. The question this section briefly explores is much narrower: do moral motivations vitiate the neoclassical model of human action? If the primacy of self-interest is viewed as core to the neoclassical paradigm, then any proposal of alternative motivation is an attack on the paradigm. (The next section argues that co-optation is a natural defense against such attacks.) The present section considers a different potential problem: self-interest may recommend actions that conflict with morality. A consistent resolution of such conflicts is crucial for the neoclassical paradigm, whether or not moral behavior is rule-governed.

In discussions of motivation, to speak of self-interest is to draw a distinction. It is to deny, or at least question, the univocality of human motivation. For the neoclassical paradigm, motivational plurality is potentially problematic. It suggests the possibility of motivational conflict—conflict that might not be resolved in consistent behavior.

Among the motivations with the potential for conflict, self-interest and morality are salient. This is independent of the nature of moral theory. For example, whether individuals adhere to moral theories that evaluate acts teleologically or deontologically is irrelevant. The potential problem for economics arises at a different locus, that of the reconciliation of moral commitments—however acquired—and other behavioral motivations. For example, moral commitments derived from simple utilitarian considerations may act as a constraint on the prudential maximization of personal welfare.

Whether conflicts between self-interest and morality generate behavior inconsistent with the neoclassical paradigm is an empirical question. Arguments that the nature of morality necessitates behavioral inconsistencies are therefore most plausibly construed as empirical claims.[13] Of course there are many possible sources of in behavioral inconsistency. Conflicts between self-interest and morality are salient because they are generally recognized, they inform most ethical debate, and they derive from no obvious deficiency.[14]

Thus for the neoclassical paradigm, the abandonment of the presumption that human action is self-interested has an evident cost: it focuses attention on a possible lack of integration among human motivations. But there is also a more general methodological cost: previously unexplained behavior can always be attributed to some some new source of utility. (Or, we might add, to some new *kind* of utility.) We seem to let the camel's nose into the tent: what is to prevent motivations from multiplying as entelechies, instincts, or drives once multiplied? This is one reason to sympathize with Edgeworth's famous claim that "[t]he first principle of economics is that every agent is actuated only by self-interest" [8, 16]. Given a narrow enough background understanding of what constitutes self-interest, admitting only self-interested motivations provides *some* restraint on admissible interpretation.[15] Unfortunately, given the evidence that moral motivations influence economic behavior, that particular restraint is too binding.

The possible conflicts between morality and self-interest are part of a broader issue: the implications of manifold motivations for maximizing behavior. (By 'maximizing behavior' I mean observed behavior that can be usefully *interpreted* as maximizing.) In particular, the recognition that internalized norms (other than norms of rationality) influence behavior raises the question whether the utility maximization model of the neoclassical paradigm is fatally deficient as a model of human behavior. This section has explored one difficulty: the potential conflict between the behavioral consistency predicted by the neoclassical paradigm and the actual, normatively influenced behavioral inconsistencies that may arise when morality and self-interest conflict. Of course, any competing model of human behavior faces a difficulty at the same nexus: explaining the behavioral outcome of conflicting motives. And the actual importance of such incoherence remains an empirical question.

5 IS MORAL MOTIVATION SPECIAL?

The previous sections uncover two primary challenges to the neoclassical paradigm: actual deviations from neoclassical rationality, and observed deviations deriving from interpretational inadequacies of the paradigm. More specifically, we found that the neoclassical paradigm may face an empirical challenge but not a logical challenge from rule-governed moral behavior, that the reconciliation of moral motivation with self-interest poses a challenge of the same nature, and that—since economists must attend to motivation—there is a strong case that moral motivations deserve incorporation in economic models.

None of these results appear to carve out a special place for moral motivation: on these criteria morality is similar to other sources of value. However, many argue that moral motivation is special at least in the sense that it defies analogy to self-interested tastes. This special nature of morality has prompted proposals to introduce morality as a separate source of value, or separate "utility", not subsumable under and not commensurable with self-interested tastes. I refer to such approaches as 'multiple-utility representations'.

This section considers whether multiple-utility representations should attract those empirically oriented economists who are ready to acknowledge the importance of moral motivations in economic behavior. Multiple-utility representations appear to be a radical departure from the neoclassical paradigm, and they raise questions about the possibility and even meaning of maximization. If useful representations of moral motivations do in fact require multiple utilities, then moral motivation is indeed special: in some sense, moral motivations and self-interested tastes are strictly incommensurable. Combined with the importance of moral considerations in economic behavior, this would threaten neoclassical utility theory, which clearly assumes a formal commensurability of available choices.[16]

Recall however that if diverse motivations are coherently reconciled in behavior, they lend themselves to a maximizing formulation. Opponents of multiple-utility representations develop this observation into arguments that multiple utilities offer no gains over natural extensions of neoclassical utility theory in the representation of moral motivations [4]. I will supplement these contentions by arguing narrowly that multiple-utility representations have little attraction for empirically oriented social scientists.

The addition of moral motivations to the neoclassical paradigm proceeds most naturally by the addition of objects of preference; I will refer to such approaches as 'extended-utility representations'. In extended-utility representations, the formal treatment of moral motivations and self-interested motivations is identical. Opponents of extended-utility representations have argued that this sheds no light on motivation, negates the important conceptual distinction between moral and self-interested behavior, repudiates motivational conflict, is static (while the acquisition of norms is dynamic, historical, and socially dependent), ignores the role of moral rules in justifying behavior, and denies the fusion of means and ends in moral action [9, 10]. Most of these criticisms are equally problematic for any theory of choice, whether formulated in terms of extended utility or multiple utilities. I will focus on the two criticisms that appear most plausibly to give multiple-utility representations an advantage over extended-utility representations: motivational conflicts vanish, and the distinction between moral and self-interested behavior appears negated.[17]

First consider the complaint that extended-utility representations, with their formal commensurability of motivations, mask the conflict between moral and other motives. Here we may grant multiple-utility representations a rhetorical advantage: insistence on a multiple-utility framework tends to highlight possible motivational clashes. However, conflict is not inherent in the multiple-utility framework. To both the extended-utility and multiple-utility conceptions, we may add a discussion of the human experience of struggle between moral and self-interested motivations, or for that matter between conflicting moral motivations. If we think it useful, we can even introduce the struggles between conflicting self-interested tastes—consider the choice between a relaxing and an exhilarating vacation, explored by Taylor [29]. Neither framework clarifies what is involved in hard decisions, whether or not these are moral decisions.

Further, I suggest that the formal commensurability of motivations on which this complaint is based is a logical requirement of any interesting theory of *choice*. A theory that predicts human choices by referring to the operative motivations in a specified environment must characterize the reconciliation of any conflicting motivations. But a consistent reconciliation is all that is involved in the formal commensurability of extended-utility representations. Thus the proponents of multiple-utility representations must accept formal commensurability whenever they develop a theory of choice.

For example, when Etzioni [10, 67] argues "that people seek a *balance* between their moral commitments and their pleasures ... rather than seeking to 'maximize' either", any neoclassically trained economist must suspect that this balance implies a formal commensurability adequate for an encompassing maximizing description.[18] To compete with the neoclassical paradigm of rational choice, proponents of multiple utility representations must show how conflicting motivations are resolved into outcomes. But as soon as this step is taken, they are involved in all of the "commensurability" of extended utility representations.

The other complaint considered in this section is that the addition of "moral tastes" to the neoclassical framework fails to draw the important distinction between self-interested and other motivations. For example, Etzioni [10, 63] objects "to the reductionism of neoclassical economists, to the notion that people act morally only as long as it makes sense in economic terms." Butler's [7] refutation of psychological egoism should lay such fears to rest: the distinction between my self-interested desires and my other desires does not disappear just because they are both *my* desires. Rather, as concisely argued by Schmidtz [23, 250], "The fact that we get psychic rewards from helping others proves we are directly concerned with the welfare of others."

The second complaint is therefore misdirected. It should be aimed at an *interpretation* of extended-utility maximization, in which moral behavior appears only as enlightened self-interest. Perhaps some of the mental discomfort from including moral "tastes" in a utility maximization problem is semantic; it subsides if we simply recast choice as balancing moral motives and other motives, rather than as maximizing utility, or replace the utility function with a balancing function. Perhaps an even simpler resolution resides in a return to an earlier usage of 'utility', where the utility of an object of preference was not conceived as some kind of experienced satisfaction but rather as the capacity of the object to fulfill preferences [5]. If some preferences are for morally satisfactory outcomes, this perspective severs the link between utility maximization and any notion of maximizing personal satisfaction. In any case, no implications for formal modeling follow logically from such interpretive shifts.

The proponents of multiple-utility representations have not adequately clarified the difference between adding a "utility" and adding a functionally similar taste (or collection of tastes). Worries about losing the representation of conflict apply equally to the extended- and multiple-utility frameworks, and the distinction between moral and self-interested motivations can be made in both frameworks. It is still true that, if we credit arguments that norms generate irrational behavior, something is wrong with the consistent reconciliation of motivations that is inherent in the extended-utility framework. Yet in the absence of a theory of the reconciliation of multiple utilities in observed behavior, this defect in the extended-utility framework cannot be treated as a substantive advantage of the multiple-utility framework. Indeed, it is natural to presume that any theory of reconciliation will have the same shortcoming. Rather than turn to irreconcilable multiple utilities in seeking to model norm induced "irrationality", a theory of choice will do better to allow, for example, normative considerations and other context specific factors to condition the reconciliation.

In sum, morality neither requires nor deserves special treatment in economic models. First of all, morality and other norms raise similar issues and can therefore be given a similar treatment. Secondly, multiple-utility representations have no clear advantages over extended-utility representations as underpinnings of a theory of normatively guided human behavior.

6 CONCLUSION

The relationship between morality and the neoclassical paradigm has been a source of considerable controversy. Much of this paper explores the implications of moral motivation for the behavioral consistency predicted by the neoclassical model of rational choice. Moral motivations may generate inconsistencies in observed behavior for at least two general reasons: the goals or rules that govern behavior may lack coherence, and apparent inconsistencies may derive from the interpretational inadequacies of the neoclassical paradigm. These challenges emerged from our discussion of the relationship between morality and the neoclassical paradigm, but they arise equally in any discussion of social norms. These challenges are fundamentally empirical and cannot be resolved a priori.

Morality is an important cause of economic behavior, and empirically oriented economists will want to incorporate moral motivation in their models. Much of the controversy surrounding the treatment of moral motivation implicitly or explicitly assumes that moral motivations deserve or require a different treatment than other motivations. In contrast, this paper argues that economists can generally be satisfied to treat moral motivations similarly to other motivations. Despite some rhetorical advantages, the challenge to the extended-utility framework from multiple-utility advocates remains inchoate and unpersuasive. For now, at least, economists wedded to the neoclassical paradigm can proceed with a conservative strategy: they can retain their formal apparatus by developing extended-utility representations of moral motivation. This requires new attention to the interpretation of human behavior, which may cut against the grain for many economists—especially those influenced by the behaviorist tendencies within the neoclassical paradigm. Yet we have seen that interpretation is fundamental to economics: observed choices are never "brute data". Once this is acknowledged, the new attention to moral motivation finds its place as one constituent of a commitment to accommodate empirically important causes of human behavior.

References

[1] Baker, George P., Michael C. Jensen, and Kevin J. Murphy, "Compensation and Incentives: Practice vs. Theory." Journal of Finance 43(30), July 1988, 593–613.

[2] Becker, G. S., "Irrational Behavior and Economic Theory." Journal of Political Economy 60(1), February 1962, 1–13.

[3] Binmore, Ken. *Game Theory and the Social Contract*, volume 1. Cambridge, MA: MIT Press, 1994.

[4] Brennan, Timothy J., "A Methodological Assessment of Multiple Utility Frameworks." *Economics and Philosophy* 5(2), October 1989, 189–208.

[5] Broome, John, "Utility." *Economics and Philosophy* 7(1), April 1991, 1–12.

[6] —, "Deontology and Economics." *Economics and Philosophy* 8(2), October 1992, 269–282.

[7] Butler, Joseph. *Fifteen Sermons upon Human Nature*. Charlottesville, VA: Ibis Publishing, 1987(1726).

[8] Edgeworth, Francis Ysidro. *Mathematical Psychics*. London: Kegan Paul, 1881.

[9] Etzioni, Amitai, "The Case for a Multiple Utility Conception." *Economics and Philosophy* 2(2), October 1986, 159–184.

[10] —. *The Moral Dimension: Toward a New Economics*. NY: The Free Press, 1988.

[11] Friedman, Milton. *Essays in Positive Economics*. Chicago: University of Chicago Press, 1953.

[12] Gauthier, David. *Morals By Agreement*. Oxford: Oxford University Press, 1986.

[13] Gibbard, Allan, "Constructing Justice." *Philosophy and Public Affairs* 20(3), Summer 1991, 264–279.

[14] Hausman, Daniel M., and Michael S. McPherson, "Taking Ethics Seriously: Economics and Contemporary Moral Philosophy." *Journal of Economic Literature* 31, June 1993, 671-731.

[15] Hicks, John R. *A Revision of Demand Theory*. Oxford: Clarendon Press, 1956.

[16] Kreps, David M. *Notes on the Theory of Choice*. Boulder, CO: Westview Press, 1988.

[17] Lancaster, Kevin, "A New Approach to Consumer Theory." Journal of Political Economy 74, April 1966, 133–145.

[18] Little, I. M. D., "A Reformulation of the Theory of Consumer's Behaviour." Oxford Economic Papers 1, January 1949, 90–99.

[19] Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics." American Economic Review 83(5), December 1993, 1281–1302.

[20] Robbins, Lionel. *An Essay on the Scope and Nature of Economic Science*, 2nd Edition. London: Macmillan, 1935.

[21] Roth, Alvin, Michael Malouf, and J. Keith Murnighan, "Sociological versus Strategic Factors in Bargaining." Journal of Economic Behavior and Organization 2(2), June 1981, 153–77.

[22] Samuelson, Paul, "A Note on the Pure Theory of Consumer's Behavior." *Economica* N.S. 5, February 1938, 61–71.

[23] Schmidtz, David, "Choosing Ends." *Ethics* 104(2), January 1994, 226-251.

[24] Sen, Amartya, "Behavior and the Concept of Preference." *Economica* 40(158), May 1973, 241–259.

[25] —, "Internal Consistency of Choice." Econometrica 61(3), May 1993, 495–522.

[26] Solow, Robert. *The Labor Market as a Social Institution*. Oxford: Basil Blackwell, 1990.

[27] Stigler, George J. "Economics and Ethics?" in *Tanner Lectures on Human Values*, edited by S. McMurrin. Cambridge: Cambridge University Press, 1981, pp. 143–191.

[28] —, and Gary Becker, "De Gustibus non est Disputandum." American Economic Review 67(2), March 1977, 76–90.

[29] Taylor, Charles. "What is Human Agency?" in *The Self*, edited by T. Mischel. Oxford: Blackwell, 1977, pp.103-35.

[30] Vanberg, Viktor J. *Rules and Choice in Economics*. New York: Routledge, 1994.

Notes

* I would like to thank John Broome, Robert Lerman, Jeffrey Reiman, and Phillip Scribner for their comments and suggestions.

[1]Let ‘morality’ denote a collection of actual (internalized) norms and values specifying the nature of good behavior toward other individuals, and let ‘moral action’ denote actions motivated by a morality. By ‘moral behavior’ I mean behavior that is interpreted by observers as moral action. Thus moral action produces morality-governed behavior, but it might not produce moral behavior. This usage is not universal: see Vanberg [30, 43] for a recent contrast.

[2]Like other acquired motivations, morality is also problematic for welfare economics. Even without such complications, the arguments of this paper apply.

[3]The notion of “morality as constraint” generally involves inappropriate reification but is partially accommodated here by allowing for lexical valuation (the relevance of which I doubt). Reputational payoff or avoided social sanctions provide examples of entirely derivative value of rule adherence.

[4]For example, we may keep promises or shop specific stores habitually, without fully reassessing the changing circumstances in which we follow these rules. (However if such rules were optimizing responses to a stochastic environment given costs of reassessment—e.g., search costs—the behavior would not be strongly rule-governed.)

[5]Thus weakly rule-governed actions may not even be “genuinely rule-following behavior” in the sense of Vanberg [30]. However Vanberg’s otherwise excellent analysis underplays the inherent (as opposed to derivative) value of rule adherence. When the value of rule following is inherent, some weakly rule-governed actions should be considered “genuinely rule-following” in Vanberg’s sense, despite full responsiveness to the particularities of each choice situation. (Of course, the interesting question still remains of when rule adherence becomes inherently valuable, and Vanberg’s “behavioralist” analysis can be helpful in formulating answers.)

[6]For example, Etzioni [10, 173–176] argues that rule-governed behavior will select inefficient means for the achievement of given ends for a variety of reasons: rules are too simple and therefore do not respond to important information in a choice situation, they are too vague to determine a specific choice, collections of rules are not generally integrated and will therefore conflict, and conflicts between rules tend to be resolved on non-rational grounds.

[7]Becker [2] argues that even random behavior, on average, can exhibit the consistency predicted by neoclassical theory. Thus the neoclassical characterization of rational behavior does not even imply deliberate intent, which is generally understood to be an important feature of rationality. This has been emphasized in recent years by the inclusion of work with laboratory animals in the area of experimental economics.

Furthermore, internal consistency is not necessary for choice to be rational. Most obviously, rational individuals may reassess their beliefs and values. Finally, *observed* internal consistency of choice cannot be an unqualified implication of rational choice, even if we treat beliefs and preferences as exogenous, since the relationships between observed choices are determined in part by the interpretive framework of the observer.

[8]Consider the claim in Binmore [3, 51]: “A rational individual is only said to behave *as though* he were satisfying preferences or maximizing a utility function and nothing is claimed at all about the internal mental processes that may have led him to do so. A utility function, in the modern sense, is nothing more than a mathematically tractable means of expressing the fact that an individual’s choice behavior is *consistent*.” Note that in addition to focusing on behavioral consistency, Binmore takes a behaviorist line that Sen has shown to be incoherent (see the section 3).

[9] As an illustration, consider the extensive controversy evoked by the apparently instrumentalist proposals of Friedman [11].

[10] This does not preclude anything from being an object of preference in a utility function. Rather some inclusions move you away from the neoclassical paradigm. More critically, for empirical applications one needs a small set of objects of preference. (Note that the paradigm does allow the inclusion of specific commodities in the utility function, a practice that Lancaster [17] and others have noted is quite peculiar.)

[11] For example, Stigler [27, 176] allows that "the self-interest hypothesis" will occasionally fail, but he still vigorously insists on the primacy of self-interest in the economic explanations of human behavior.

[12] Consider any theory that predicts which action $a \in A$ will be chosen in a given choice situation, which is characterized by the environment $e \in E$ and the feasible set of alternatives $A(e) \in \mathcal{A}$. Here the notation $A(e)$ indicates that the set of feasible alternatives generally depends on the environment. The neoclassical paradigm requires that the particular choices associated with elements of $\mathcal{A} \otimes E$ have two properties: (i) the conditionality on the environment is irrelevant (i.e., e matters only by affecting A, not directly) and (ii) variations in A generate consistent variations in a. (Consistency here means satisfaction of the Weak Axiom of Revealed Preference as presented, for example, in Kreps [16].) These two properties are necessary for choice to be characterized as the result of maximizing a utility function $U(a)$. If the consistency arises only after controlling for the environment, we might still hope for a utility representation of the form $U(a; e)$. (This will not be helpful if the set of feasible alternatives must be included in the environment, or if key environmental variables have no observable proxies.)

[13] For example, Etzioni [10] argues that morality is rule based and (therefore, as an empirical matter) does not have an internally consistent structure. He also offers a psychological claim that conflict between moral values and other values may reduce the ability of individuals to pursue either [10, 72–73].

[14] For Broome [6, 281], persons who do not achieve consistent resolution between their moral commitments and self-interest "are not acting on any rational principles at all." But suppose my ethical motivations derive from a considered and consistent ethical theory and that I am fully aware of how most efficiently to pursue my self-interest. I might still lack a principle for resolving conflicts between the two motivations. Whichever goal I pursue, I act on rational principles. It is only in choosing my goal that I have no rational principle. And this seems just right: what rational principle can determine the proper balance between morality and self-interest?

[15] Clearly there is also a danger in multiplying self-interested values. The key to maintaining empirical content, heuristically stated, is in the parsimony and scope of the motivational assumptions: good theories invoke a small number of basic tastes—or instincts, or kinds of utility—that can generate implications for an entire class of behaviors. (Of course, this fails to adequately distinguish taxonomy from explanation, which is a methodological challenge beyond the scope of this paper.)

Consider a well worn but concrete example: why do we tip strangers for good service at restaurants we will never again visit? Setting aside accounts portraying this as strategically prudential [12], neoclassical choice theory might rationalize such behavior with a "taste for tipping". However, as Stigler and Becker [28] argue, generating new tastes to account for specific behaviors is a poor explanatory strategy. If we posit instead that people often constrain their behavior out of feelings of fair exchange or reciprocity (Gibbard [13]), we get a parsimonious explanation of broader scope and greater predictive power. We might formalize this as a second "utility", or with equal generality we might add a "taste for fair exchange". (Either approach seems capable of admitting that socially acquired notions of fairness influence individual behavior; see the next section.)

[16] 'Formal commensurability' does not imply individuals calculate using "common units", but rather that their behavior offers the possibility of such an *ex post* representation.

[17] Problematically, even the strongest advocates of multiple utilities sometimes allow that subsumption of moral motivation in an extended-utility function may be a formal possibility [10, 254]. However, Hausman and McPherson [14, 688] opine that "Even though it may be possible formally to model the committed agent as maximizing utility, it seems enlightening not to do so." It is worth being specific about the nature of this enlightenment, which suggests moral motivation deserves or requires a different theoretical treatment than other sources of motivation.

[18]Etzioni [10, 42] also acknowledges “the fact that people are less likely to heed a moral commitment if the costs are high”. This implicitly suggests the kind of trade-off between moral and self-interested behavior that economists embody in the formal commensurability of utility theory.