1998

Modeling Morality: What Are the Limits to Self-Directed Preference Change?

Richard H. McAdams

MODELING MORALITY: WHAT ARE THE LIMITS TO SELF-DIRECTED PREFERENCE CHANGE?

RICHARD H. MCADAMS*

To enter the field of taste changes one ought to find danger exhilarating.[1]

INTRODUCTION

Robert Cooter pursues nothing less than the holy grail of behavioral theory, a theory of preference-formation, "arguably the most important unsolved problem in the social sciences."[2] For law and economics, the most important part of this puzzle is the effect that law has on preference formation. Cooter's article, *Models of Morality in Law and Economics: Self-Control and Self-Improvement for the "Bad Man" of Holmes*,[3] is an important step toward a theory of law and preference formation. His innovation is the idea of a "Pareto self-improvement," which he uses to show that law can *motivate* individuals to change their preferences. The general point is that opportunities motivate preference change because individuals sometimes need a certain kind of character to obtain economic gains. The specific point is that law can shape the opportunities that induce preference change. To the extent he succeeds, Cooter demonstrates that law is more than a command to the Holmesian bad man; it is a means of building the "vaguer sanction of conscience"[4] that influences the good.

Cooter joins other brave economists who attempt to make preferences endogenous to rational choice models and, thereby, to explain the creation of conscience and character.[5] I strongly endorse the effort, although we cannot

* Professor of Law, Boston University. I thank Robert Bone, Robert Cooter, and Stephen Marks for their insightful comments on an earlier draft.

[1] T.A. Marschak, *On the Study of Taste Changing Policies*, 68 AM. ECON. REV. PAP. & PROC. 386, 386 (1978).

[2] Jon Elster, *More Than Enough*, 64 U. CHI. L. REV. 749, 749 (1997) (review of GARY S. BECKER, ACCOUNTING FOR TASTES (1996)).

[3] *See* Cooter, *Models of Morality in Law and Economics: Self-Control and Self-Improvement for the "Bad Man" of Holmes*, 78 B.U. L. REV. 903 (1998).

[4] *See* Oliver Wendell Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 459 (1897), *reprinted in* 78 B.U. L. REV. 699, 701 (1998).

[5] *See, e.g.*, GARY S. BECKER, ACCOUNTING FOR TASTES (1996); ROBERT FRANK, PASSIONS WITHIN REASON 146-62 (1988) (chapter on "Becoming Moral"); Kenneth G. Dau-Schmidt,

know at this point whether it will succeed. I believe we will discover the answer only if, for a time, we accept the model and pursue its implications. Thus, my comment here adopts the pivotal and unfamiliar assumption that individuals can, to some degree, choose their preferences. And I will follow Cooter in attempting to discuss this choice without relying on the idea of "second order" preferences, or preferences about preferences.[6] Although the idea of Pareto self-improvement raises a host of interesting issues, I limit my comments to two. Cooter's model requires elaboration to explain (1) why individuals who have changed their preferences to gain an opportunity do not change their preferences *back* once the opportunity is realized; and (2) why individuals would choose the Protestant work ethic over a nonwork ethic that would allow them to extract greater pleasure out of leisure activities. Both points raise the same issue: What are the limits of self-directed preference change? My claim is that some constraints are needed to explain why we observe less preference change than would otherwise be predicted by the model. I begin, however, by describing the unique contribution of Cooter's article.

I. COOTER'S BREAKTHROUGH: PARETO SELF-IMPROVEMENT

Economics conventionally avoids making predictions or evaluations of preference change. Cooter attempts both.[7] His theory is predictive because it states conditions under which individuals will tend to change their preferences. As I explain further below, the conditions exist whenever an individual is "better off," under both his original and new preferences, with the outcome uniquely secured through his new preferences. Given these conditions of Pareto self-improvement, the theory also suggests a normative criteria. The individual makes the preference change because he expects to benefit from it; when he benefits without making anyone worse off, the change is also a social Pareto improvement. Indeed, Cooter seems to suggest that Pareto self-improvements often benefit others as well by making cooperation

An Economic Analysis of the Criminal Law as a Preference-Shaping Policy, 1990 DUKE L.J. 1; Robert Frank, *If Homo Economicus Could Choose His Own Utility Function, Would He Want One With a Conscience?*, 77 AM. ECON. REV. 593 (1987); Burton A. Weisbrod, *Comparing Utility Functions in Efficiency Terms or, What Kind of Utility Functions Do We Want?*, 67 AM. ECON. REV. 991 (1977).

[6] *See, e.g.*, John Christman, *Autonomy: A Defense of the Split-Level Self*, 25 S.J. PHIL. 281, 283-90 (1987); Harry G. Frankfurt, *Freedom of the Will and the Concept of a Person*, 68 J. PHIL. 5 (1971); Cass Sunstein, *Legal Interference with Private Preferences*, 53 U. CHI. L. REV. 1129, 1140-45 (1986). I note one exception. Arguably, Cooter does rely on one second-order preference—for happiness—to the extent that his analysis requires comparison of cardinal rather than ordinal utilities. *See infra* note 16. My comments occasionally invoke cardinal utility and, to that degree, rely on a second-order preference for happiness.

[7] *See* Cooter, *supra* note 3, at 922 ("Preferences influence opportunities, or, in plain speech, who we are influences what we get.").

with them more likely.

Although uncommon, other economists have ventured into the difficult subject of preference change. Most of them, however, offer evolutionary models, in which preferences that provide individuals with greater fitness are, in each generation, transmitted to more offspring.[8] Part of what makes Cooter's approach especially unusual is his decision to focus on preference change at the individual level and, specifically, on *self-directed* preference change.[9] To make this approach work, Cooter begins with a simple but essential insight: *A person's preferences affect his opportunities.*[10] An individual whose preferences include honesty, for example, may have opportunities for economic transactions that are denied to individuals without such a preference. This point depends on what Cooter terms "translucence": that an individual's character or the preferences that determine his character are observable by others, albeit only imperfectly.[11] Because other economic actors will seek to transact with individuals with an honest character, the preference for honesty will provide an individual with more opportunities. These new opportunities provide a motive for self-directed preference change.

To understand the importance of Cooter's insight, consider John Elster's recent criticism of Gary Becker's work on self-directed preference change. Like Cooter, Becker imagines that individuals can rationally choose their preferences, including time preferences—that is, the amount by which one discounts the future. Elster, however, rejects the contention that "[t]he preferences of [an agent] A can be traced to a rational choice by A for the very purpose of acquiring those preferences."[12] For example, "[w]e cannot expect people to take steps to reduce their rate of time discounting, because to want

---

[8] *See, e.g.*, FRANK, *supra* note 5; Steffan Huck, *Trust, Treason, and Trials: An Example of How the Evolution of Preferences Can Be Driven by Legal Institutions*, 14 J. L. ECON. & ORG. 44 (1998); Arthur J. Robson, *The Evolution of Attitudes to Risk: Lottery Tickets and Relative Wealth*, 14 GAMES & ECON. BEHAV. 190 (1996); Werner Guth, *An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives*, 24 INT'L J. GAME THEORY 323 (1995).

[9] Thus, Cooter's work also differs from prior attempts to compare preference sets of entire societies. For example, Burton Weisbrod proposed a "base-reversal" test: one set of preferences is "preferred" to another "if a person prefers to be in a society in which *everyone* has one particular type of utility function (say type II) rather than another (say type I)—prefers it in the sense that he prefers the consumption bundle he would expect to receive in society II to the bundle he would expect to receive in society I, and he has this preference ordering no matter which type of utility function he uses as the basis of evaluation." *See* Weisbrod, *supra* note 5, at 992 (emphasis added). Weisbrod's point does not permit any prediction about behavior because it does not identify any conditions under which one individual would want to change his preferences.

[10] *See* Cooter, *supra* note 3, at 922 ("Preferences influence opportunities, or, in plain speech, who we are influences what we get.").

[11] *See id.*

[12] Elster, *supra* note 2, at 753.

to be motivated by long-term concerns ipso facto *is* to be motivated by long-term concerns."[13] If an individual prefers a front-loaded consumption stream to a larger but back-loaded stream, then the individual will not be motivated to make himself prefer the back-loaded stream, especially if he must incur immediate costs to make the change.

Elster offers the following illustration: Suppose *A* prefers consumption stream *I*—$500 per term until retirement and then $100 per term thereafter (from welfare payments)—to consumption stream *II*—$450 per term before and after retirement, with $50 savings per term before retirement funding the post-retirement consumption. Now imagine that *A* is offered "a 'discounting pill' at a price of $100, which will reduce [his] rate of time discounting."[14] With the pill, *A* will have consumption stream *III*: $350 in the first term when he buys the pill, and $450 for every subsequent term. If *A* now prefers *I* to *II*, he will necessarily prefer *I* to *III*, which is worse than *II*.

Cooter's analysis, however, overcomes this objection.[15] Consumption streams *I* and *II* both assume a salary of $500 until retirement. But if preferences affect opportunities, then *A* may increase his salary by acquiring more desirable preferences. Cooter proposes, for example, that valuing the future more would make *A* a better employee and that this preference is (partly) observable by potential employers. We may therefore expect *A* to earn more after he changes his time preferences. Because a higher salary is better than a lower one whatever the time preference, the prospect of earning more may induce the preference change. Even if one has to bear the $100 cost of changing one's time preference in the term before receiving the higher salary, it is possible that *A* will, with his *initial* time preferences, value the new state of affairs more highly than the existing one. For example, suppose the new salary is $550. *A* could prefer consumption stream *IIIA* to all others: $400 in the first term (old salary minus change costs) and $550 thereafter (new salary with no savings; but once he makes the change, he might find himself using part of the new salary for savings).

The same is more obviously true of moral preferences. If *A* is honest and his honest character is partly observable, employers will seek his services over those who are not honest. The higher salary possible with honest preferences will motivate some individuals to make this change.[16] If so, then

[13] *Id.* at 754.

[14] *Id.*

[15] I do not have the space to evaluate the merits of Elster's criticism as applied to Becker's analysis.

[16] There is one important clarification to this analysis that Stephen Marks pointed out to me. The consistent ordinal ranking of the two opportunities, such as the different wealth two jobs provide, cannot provide a fully sufficient reason to change preferences. For example, even if one prefers job 1 to job 2 under both sets of preferences, it is also possible that, with the new preferences necessary to obtain job 1, one would derive less utility for happiness from job 1 than one derived from job 2 with the old preferences. Thus, Cooter's theory re-

Cooter's theory explains the idea of self-improvement using the relatively simple tools of rational choice theory.

Cooter's account of preference change is intentionally incomplete. He sets aside for present purposes the question of *how* preferences change and how an individual motivated to change his preferences would go about doing so. He simply assumes that individuals have some power to change their preferences.[17] He also assumes the "translucent" quality of character, meaning that one can imperfectly observe an individual's character or moral preferences.[18] Neither assumption detracts from the importance of Cooter's results, but I note them here because I will refer to them below.

II. THE DARK SIDE OF SELF-DIRECTED PREFERENCE CHANGE: MOTIVATED BACKSLIDING

If we assume that individuals have it within their power to change their preferences, then it is necessary to explain why individuals do not change their preferences even more often than they appear to do. As a first example, Cooter's model raises the question whether individuals ever choose to *abandon* moral preferences. Is there ever motivated *backsliding*? Let us focus on

quires that the individual rank job 1 higher than 2, under either set of preferences, in a cardinal, not merely ordinal, sense. Cardinal utility would seem to require at least one second-order preference—a metapreference for happiness—by which the individual can evaluate the desirability of his first-order preferences.

17 *See* Cooter, *supra* note 3, at 922-23. Cooter also implicitly assumes certain limitations on the means of preference change. His analysis assumes that preference change is neither too difficult nor too easy. I say this because Cooter's analysis is untenable if preference change is at either extreme. If preference change is very difficult and time-consuming, then there would be little change after parents instilled values or preferences in their children. By asserting the importance of Pareto *self*-improvement, Cooter assumes that preference change is sufficiently easy to be meaningful for adult behavior. On the other hand, if preference change were extremely easy and took very little time, morality would always be quite temporary and could exert very little restraint on behavior. Under this extreme condition, individuals would constantly change their preferences from day to day or moment to moment as necessary to exploit opportunities for honesty and dishonesty. There would be no special opportunities granted those whose character was—only for the moment—honest. Consequently, whatever the mechanism of preference change, it must change preferences between these extremes of time and difficulty. I discuss possible limitations to preference change in the remainder of this comment.

18 *See* Cooter, *supra* note 3, at 922. Economist Robert Frank has explored this possibility in detail, arguing that the biology of emotions makes it difficult, but not impossible, for individuals to fake their moral sentiments. *See* FRANK, *supra* note 5, at 114-33 (chapter on "Telltale Clues" to emotional states). For example, most individuals reveal their true emotional state through involuntary physical reactions such as facial expressions, perspiration, voice pitch, and blushing. *See id.* Other individuals may interpret these states to determine, with substantial but not perfect accuracy, the individual's emotional commitments on matters such as truth-telling or opportunism. *See id.* at 134-35.

the individual in Cooter's model who has recently chosen to improve his character to obtain an economic opportunity. Once the individual has obtained the opportunity, what, if anything, prevents him from reverting to his original preferences?

As a concrete example, suppose an employer seeks an honest employee because the job will give the employee an opportunity to steal from the employer without detection. The higher pay for jobs of this sort motivates *A* to acquire a preference for honesty, which, given the translucence of character, gains him the job. Cooter is correct that *A* prefers the higher pay both from the perspective of his old preferences and the new. More precisely, the preference change is a Pareto self-improvement because *A* at all times prefers consumption stream *II*—higher pay with preferences against stealing—to consumption stream *I*—lower pay with preferences that permit stealing. Once hired, however, we must ask: Will *A* prefer his existing situation—consumption stream *II*—to consumption stream *III*: higher pay *with* preferences that permit stealing? Once he acquires the opportunity that caused him to become honest, he might be better off if he were no longer honest.

There are several possible responses to this problem. One kind of response is to deny that backsliding is possible, to assert instead that the problem is illusory. I first consider and reject various arguments along these lines. A second sort of response accepts the possibility of backsliding and provides incentives within the model that limit this behavior. My point in raising the problem of backsliding is to illuminate the importance of these latter arguments.

A. *The Genuine Temptation of Backsliding*

The backsliding problem might seem superficial and easily resolved. Some may deny any parallel between the creation and abandonment of honest preferences. Perhaps those with an honest character recoil in horror at the prospect of losing that character, whereas dishonest individuals are content to view character change as merely one means among many of obtaining economic opportunity. Yet this account is hard to square with rational choice theory. We can easily imagine that *all* individuals find the idea of changing their character to be distasteful, but that would not explain why preference change toward morality was easier than backsliding. What lies behind the intuition that honest preferences are more stable than dishonest preferences is, I believe, just the moral intuition that honest preferences are *better* than dishonest preferences. But rational choice theory cannot easily accommodate the use of moral reasons as motivations. Certainly I do not read Cooter's model as relying on a theory of moral reasoning.[19]

[19] *See* Cooter, *supra* note 3, at 928 ("When preferences change, some ethical theories prefer the original preferences while other ethical theories prefer the final preferences. This fact creates a dilemma for evaluating public policies that change preferences. The Paretian approach partly resolves the dilemma by allowing the state to create opportunities for individu-

One might still reject the backsliding problem, however, with the claim that moral individuals do not perceive any gains from immoral behavior. Honest individuals, on this view, do not receive any utility from dishonesty, even if that dishonesty makes it possible for those individuals to satisfy some of their other preferences. Thus, immoral gains cannot make the individual better off viewed from the standpoint of his moral preferences. In sum, backsliding never pays.

This simple picture of morality may be correct. But the implications of this view seem excessively optimistic. To say that moral individuals do not subjectively perceive the gains of immorality as benefits implies that moral individuals are never *tempted* by those gains. If there is no temptation, this view further implies that moral individuals never violate their moral preferences for selfish reasons. Morality is dichotomous: One is either fully constrained or fully unconstrained by morality; one either obeys moral precepts or acts without regard to them.

Consider an alternative view consistent with Cooter's rational choice account. Suppose that moral preferences are like other preferences in that individuals weigh them against each other in making decisions.[20] On this view, having a moral character does not eliminate the perceived benefit of immoral behavior; morality instead works by creating an offsetting internal sanction—guilt—that the individual suffers if he violates his moral preferences.[21] The guilt counters the gains of wrongdoing and thereby induces honest behavior. But guilt may not deter all wrongdoing; sometimes the gain is so great that the individual acts immorally and lives with his guilt. The guilt model leaves room for moral individuals to behave immorally with guilt, which seems more realistic than the idea that moral individuals can do no wrong.[22]

If moral individuals perceive the benefits of immoral acts, then the backsliding question remains. Once hired, why doesn't *A* perceive the advantage of becoming the kind of person who would no longer feel guilt about stealing from his employer? The gains from theft provide an opportunity motivating

als to make Pareto self-improvements that do not harm anyone else.").

[20] I do not mean that moral preferences would have to work exactly like other preferences in all respects.

[21] As Cooter explains elsewhere: "[I]nternalization . . . can tip the individual's motivational balance. Economic models often view motivation as a calculus of psychological benefits and costs. From this perspective, internalization attaches a 'guilt penalty' to violating a norm, which can change the sign of the net psychological benefits." Robert Cooter, *Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant*, 144 U. PA. L. REV. 1643 (1996); *see also* FRANK, supra note 5, at 152-53 (creating an economic theory of moral sentiments based on the discussion of guilt and shame in JEROME KAGAN, THE NATURE OF THE CHILD (1984)).

[22] The guilt model implies that moral traits, like honesty, are continuous. Rather than being either honest or dishonest, individuals incur a range of guilt feelings from an act of dishonesty. Whether an honest individual resists a given level of temptation to act dishonestly depends on the degree of guilt he suffers from such acts.

the preference change. In a model without second-order preferences,[23] one cannot say that *A* has a preference to retain his preference for honesty. Thus, even with his existing preferences, if the reason he does not steal is the expected guilt, then he would presumably count himself better off if he could rid himself of this guilt inclination. Assuming he can keep the job, he will not regret backsliding once it is complete because he will gain the benefits of the job and of stealing, all without violating any of his preferences. Thus, once the individual acquires the opportunity, backsliding is apparently a Pareto self-improvement.

B. *Barriers to Backsliding*

Absent some barrier to backsliding, there will be no advantage to other market participants to dealing with moral individuals and hence no opportunities for those who (temporarily) have moral preferences. Absent such opportunities, there will be no Pareto self-improvements toward morality. For this reason, I believe it is necessary to complicate Cooter's model to account for the absence of universal backsliding. Defining the barriers to backsliding will clarify and extend the analysis. Here I speculate about two potential obstacles to backsliding and address some of their implications.

First, guilt may deter backsliding. The guilt sufficient to deter a given immoral act might also seem sufficient to deter backsliding to gain the benefits of that act. Suppose that the employer correctly perceives that *A* is sufficiently honest to resist the temptation to steal the $50 per week he could take without detection. Under the guilt model, *A* anticipates suffering more than $50 in guilt from one act of theft. Now when *A* contemplates backsliding, he might anticipate the same amount of guilt from acting—whatever acts are required by the mechanism of preference change—to become dishonest. Thus, we start off knowing that *A* prefers consumption stream *I*—no guilt each week—to consumption stream *II*—$50 minus "*g*" (guilt) each week. If acting to develop dishonest preferences incurs the same guilt as being dishonest and promises only the same benefit, then *A* may remain honest.

This analysis ignores, however, the effect of preference change. As *A*'s preference for honesty declines, so will the amount of guilt he suffers. Eventually it reaches zero. Depending on the speed of the preference change and his time preferences, *A* might prefer consumption stream *III*: $50 minus *g* each week, where *g* starts off as more than $50 and declines to zero. In this scenario, *A* is investing in the short run for a long term return, just as in the case of the initial move toward honesty.[24] Indeed, the more future-

[23] Or a model with only one second-order preference for happiness. *See supra* notes 6, 16.

[24] For this reason, I include the $50 in consumption stream *III* from the very first week when *A* decides to become dishonest. Even though the guilt outweighs the monetary reward at the beginning, an individual committed toward this preference change would immediately begin stealing (and incurring guilt) in order to decrease the short run net cost of making the

oriented *A* is, the more likely it is that he will backslide.

There is a second, much simpler barrier to backsliding: translucence. As long as character is translucent, then the gains of morality may be lost if one's good character is lost. If *A* backslides, then presumably *A*'s employer may observe his character change even if he cannot detect *A*'s acts of theft. If the opportunity given honest individuals is reversible, as with an employee who may be terminated, then *A* must maintain his character to keep the special opportunity he acquired by becoming honest.

These two barriers—guilt and translucence—are interesting because they limit backsliding without ruling it out. The explanations have interesting implications for when backsliding occurs. Translucence prevents backsliding only if there is a risk that the individual will lose the gains awarded to the moral when the new immoral preferences are detected. Thus, where the opportunity gained by honesty *is* irreversible, for example, a job with life tenure, then backsliding may occur. The analysis also implies an endgame problem: When the individual is for other reasons about to lose the gains of honesty, for example, when he is about to retire, then this barrier to backsliding no longer exists.

Guilt may nonetheless continue to deter backsliding when translucence fails. But the explanation given above implies that *unexpectedly* high temptation may still cause backsliding. If honesty is scarce, an employer will not search or pay for more honesty than he anticipates the job requiring. Suppose the employer hires on the assumption that the job gives the employee the opportunity to steal $50 per week. Suppose further, however, that the employer is quite wrong and the new employee—*A*—finds he can steal $500 per week without detection. Under the above analysis, we can imagine why *A* backslides. The employer correctly judged him honest enough to resist the temptation to steal $50 per week and to resist changing his preference for honesty given that temptation. But if that is all *A* can resist, a larger temptation will prompt *A* to make a Pareto self-improvement by abandoning his preferences for honesty.

All of this analysis suggests an interpretation of Lord Acton's famous claim that "[p]ower tends to corrupt and absolute power corrupts absolutely."[25] Positions of power are often reserved for those thought to be honest. These positions are therefore the kind of opportunities that may motivate Pareto self-improvements toward morality. But the acquisition of real political or economic power will often contain the conditions necessary for backsliding. First, to hold what is commonly thought of as *power* means the individual faces a temptation for self-dealing or mischief far beyond average

---

change. Indeed, once *A* decided to become dishonest, he might incur no additional guilt from acts of dishonesty. Further, the act of stealing might be a necessary part of the mechanism of abandoning a preference for honesty.

[25] John E.E.D. Acton, *Acton-Creighton Correspondence*, *in* ESSAYS ON FREEDOM AND POWER 335 (Gertrude Himmelfarb ed., 1972).

temptations. In this context, guilt may fail; most people may incur insufficient guilt to counter the pressure of extraordinarily high temptation. Second, political or economic power is often (nearly) permanent either by design or because one who acquires the position may use his power to preserve his tenure. When the position is permanent, its benefits no longer serve an inducement to remain honest. The translucence of character is then irrelevant. Thus, with some modification, Cooter's model predicts the corrupting influence of power.

In sum, backsliding is possible, but not inevitable. As Cooter demonstrates, individuals will seek to acquire moral preferences to gain certain opportunities. Market actors will create such opportunities because they value honesty and other moral traits and they predict the relative infrequency of backsliding. Nonetheless, under the right conditions, individuals will make a pareto self-improvement that consists of abandoning their moral preferences.[26]

III. SELF-DIRECTED HAPPINESS: WHY DOESN'T EVERYONE DEVELOP PREFERENCES THEY CAN SATISFY?

Cooter's model also raises the question why individuals don't more frequently change their preferences having nothing to do with morality. There seems to be much room for Pareto self-improvement if individuals diminish their desires so as to be happier with less. If one cannot afford tickets to more than one movie per month, why not *become* the kind of person who wants nothing more? More generally, if one earns a paltry salary by modern American standards, why not become the kind of person for whom this purchasing power is ample, as it would be for most people living in poorer nations or previous eras? Given a budget constraint or "feasible set," one would always be better off if one could gain more utility from less.[27] Once we say individuals can *will* themselves to have different preferences, we have to explain why they don't will themselves to have the kind of preferences they know they can satisfy, apparently the surefire way to happiness.[28]

---

[26] For lawyers, the latter fact may be as important as the former because it says something about the need to structure institutions to minimize the corrupting influence of power.

[27] Indeed, Jon Elster and Cass Sunstein argue that individuals commonly adapt their preferences to their opportunities (which makes it difficult to justify existing allocations by reference to preferences created by those allocations). *See* JOHN ELSTER, SOUR GRAPES (1983); Sunstein, *supra* note 6, at 1146-50. Thus, where Cooter says that our "preferences . . . influence[ ] what we get," Cooter, *supra* note 3, at 922, Sunstein observes that "what people want is sometimes a product of what they can get." Sunstein, *supra* note 6, at 1146.

[28] Some of these preferences have nothing to do with the material standard of living: If one lacks the talent to succeed as a comedian or chess player, why not give up the preference for such a life? If one cannot find a romantic partner, why not develop the character of one happy living alone?

The answer to this question probably has a lot to do with the mechanism of preference change. As I stated above, Cooter reasonably sets aside the question of how individuals change their preferences. No doubt there are some technical limits to what individuals change about themselves, and perhaps they simply cannot make themselves want less. If so, it will certainly be important to the overall project of explaining preference change to explain the constraints that limit or exclude this sort of strategy.

Even accepting some limit to preference change, however, I wish to explore the issue further because one of Cooter's examples of motivated character change is the Protestant work ethic. This ethic, Cooter says, "values production and devalues leisure."[29] If so, then I can restate my initial question in terms of leisure: Why don't we observe more people choosing to change their preferences so as to value leisure more? Some forms of leisure are very affordable. If one places a very high value on them, then one can extract more utility out of less material wealth. For example, if one cannot afford a late model sports car or an apartment with a view, why not *become* a person who values very highly sleeping until noon, listening to the radio, reading library books, taking long walks, and conversing endlessly while sipping a single coffee in a cafe? Even if one cannot want less, it will surely pay to want affordable things more.

Cooter's answer—job opportunities—is inadequate. No doubt he is right that individuals who internalize a heightened desire to work and suppress their desire for leisure are more attractive to employers and will receive better jobs. But that fact alone does not explain why the Protestant work ethic is a Pareto self-improvement. Those who value work less will wind up with less material wealth, but they will also place less value on the goods and services such wealth can buy. Individuals who extract great pleasure from low-cost leisure activities will benefit by trading off material wealth for greater leisure time. Thus, if an individual starts off without a strong work ethic, it is not clear why he would want to acquire one. Moreover, he may even want to abandon a work ethic. Suppose that parents transmit the Protestant work ethic by instilling an ambition for material wealth. Making children ambitious—desirous of high grades, prestigious jobs, and money—is a simple way of making them work hard. But some individuals will inevitably fail to achieve these ambitions. At some point, these individuals would seem to gain from discarding their work ethic.

The desirability of the Protestant work ethic might seem a bit tangential to the main thrust of Cooter's analysis—an example rather than a key concept. But the issue supports my general point: that, to generate determinate results, the model needs to include some limits to preference change. Absent such limits, we would observe far more Pareto self-improvements than exist. Again, I will speculate about what such barriers to preference change might be.

---

[29] Cooter, *supra* note 3, at 924.

First, decisions about preference change are probably plagued by information costs. To decide whether a preference change is in one's interests requires information about the opportunities created by different preferences, whether it is possible to acquire new preferences, and the costs of doing so. The latter two facts are particularly difficult to determine: We can often experiment with consumption, but—unless preference change is very easy[30]—we cannot often experiment with different preference sets. When faced with such informational problems, people often look to see what others are doing. For this reason, decisions about what character to seek or to instill in one's children will likely be heavily influenced by social norms. These decisions will then be subject to the problem of "herd behavior," where individuals with limited information presume that it is in their interest to follow the decisions being made by most other individuals.[31]

Thus, we should not too readily infer that prevailing choices about preferences are efficient. Imagine, for example, a story in which the Protestant work ethic has outlived its usefulness. Suppose that it was adaptive in an earlier century only because life expectancy depended upon working a large part of each day, but that the ethic is maladaptive with today's high level of worker productivity.[32] In other words, suppose that modern productivity makes it technically feasible for most people to work very little and still ensure their basic human needs,[33] and that, with perfect information, individuals would now prefer a leisure ethic to a work ethic. Nonetheless, given imperfect information, individuals might continue to embrace the work ethic because everyone else does the same, which they take as evidence of its superiority. Thus, information costs may impede desirable self-directed preference change. Information costs might be so pervasive that we can explain very little of the actual world of preference change without taking account of it.

A second possible barrier to preference change is a collective action problem. For example, suppose that American workers suffer from a "rat

[30] But if preference change were very easy, that would present more general problems to the theory. *See supra* note 17.

[31] *See* Abhijit V. Benerjee, *A Simple Model of Herd Behavior*, 107 Q.J. ECON. 797 (1992); Eric Posner, *Law, Economics, and Inefficient Norms*, 144 U. PA. L. REV. 1697, 1717-18 (1996).

[32] I do not mean that being future-oriented is maladaptive. To the contrary, planning for the future would seem to be crucial at any level of productivity. But one *could* care greatly about the future and still value cheap leisure activities very highly. Such a person would be highly motivated to work the minimum amount to ensure the future satisfaction of his basic needs, but would not work so much as to impede his enjoyment of substantial leisure time. Beyond basic needs, such a person places a low value on consumption goods.

[33] *Cf.* JULIET B. SCHOR, THE OVERWORKED AMERICAN: THE UNEXPECTED DECLINE OF LEISURE 1-2 (1991) ("Since 1948 . . . [t]he level of productivity of the U.S. worker has more than doubled . . . [but] we did not use any of the productivity dividend to reduce [work] hours.").

race" problem, in which everyone works harder just to stay even with each other.[34] This might occur if individuals derived utility not merely from the their absolute level of material wealth, but from how that level of wealth compares with the average.[35] If so, then everyone may struggle to obtain more wealth than others. Yet the result of this effort is not to expand relative wealth or to eliminate relative poverty but only to reallocate who gains what relative position. The resources invested in such a zero-sum activity—such as foregone leisure—are wasted.[36] Yet the best strategy for each individual may be to invest heavily in the effort to win the relative wealth and avoid relative poverty. Each may therefore seek to internalize a work ethic as a means of competing for a better job. They might pursue the work ethic even knowing that everyone would be better off without it.

Information costs and collective action problems may provide an essential explanation for why certain apparently useful preference changes do not occur. If so, they also challenge the efficiency of the preference changes we actually observe. *A* may make the preference change because he believes it will make him better off, but he may be mistaken. Or, he may be correct that he is better off making the change than avoiding it, but society may be worse off because everyone makes a similar change. None of these points contradict Cooter's explicit claims, but they suggest greater pessimism about preference change than he exhibits.

* * * * *

I have focused only on Cooter's general model of self-directed preference change. Cooter uses the theory to discuss how law may, by affecting opportunities, motivate preference change toward morality.[37] He recognizes that law can also have the opposite effect: without law, individuals may find that they productively interact with others only through trust; and they may gain

---

[34] *See* George A. Akerlof, *The Economics of Caste and of the Rat Race and Other Woeful Tales*, 90 Q.J. ECON. 599, 603-05 (1976). For critiques of economic growth on these grounds, see FRED HIRSCH, SOCIAL LIMITS TO GROWTH (1976); STAFFAN LINDER, THE HARRIED LEISURE CLASS (1970); TIBOR SCITOVSKY, THE JOYLESS ECONOMY 119-20 (1976); Trygve Haavelmo, *Some Observations on Welfare and Economic Growth*, *in* INDUCTION, GROWTH AND TRADE 65 (W.A. Eltis et al. eds., 1970).

[35] *See, e.g.*, ADAM SMITH, THE THEORY OF MORAL SENTIMENTS 112-13 (D.D. Raphael & A.L. Macfie eds. 1976) (1759) ("[I]t is chiefly from this regard to the sentiments of mankind, that we pursue riches and avoid poverty . . . . It is because mankind are disposed to sympathize more entirely with our joy than with our sorrow, that we make parade of our riches, and conceal our poverty."). *See also* ROBERT FRANK, CHOOSING THE RIGHT POND: HUMAN BEHAVIOR AND THE QUEST FOR STATUS (1985); THORSTEIN VEBLEN, THE THEORY OF THE LEISURE CLASS 26-34 (Random House 1934) (1899); Richard H. McAdams, *Relative Preferences*, 102 YALE L.J. 1 (1992).

[36] *See* sources cited *supra* notes 34-35. For a detailed example of how this problem could cause individuals to sacrifice leisure excessively, see McAdams, *supra* note 35, at 85-88.

[37] *See* Cooter, *supra* note 3, at 924-927.

trust only by becoming honest.[38] The net effect of law, however, will depend greatly on the basic model of preference change, which is why I have limited my remarks to that model. In general, I believe preference change remains a slippery subject. Cooter makes an important breakthrough. The problem for economic theory, however, is not just to explain what motivates change, but what limits it.

[38] *See id.* at 929.