EXPRESSIVE LAW AND ECONOMICS

*ROBERT COOTER**

ABSTRACT

This article develops an economic theory of expressive law. By expressing social values, law can tip a system of social norms into a new equilibrium. This process can create or destroy a social norm without changing individual values. In addition, law can change the individual values of rational people. Internalizing a social norm is a moral commitment that attaches a psychological penalty to a forbidden act. A rational person internalizes a norm when commitment conveys an advantage relative to the original preferences and the changed preferences. I call such a commitment a "Pareto self-improvement." By creating opportunities for Pareto self-improvements, law induces rational people to change their preferences. Inducing change in this way respects individual preferences, rather favoring a particular moral theory.

THE imperative theory of law defines a law as an obligation backed by a sanction.[1] Economic analysis has enjoyed great success by analyzing a legal sanction as if it were a market price.[2] Viewing it as a price, the actor sees a sanction as an external constraint. Alternatively, the actor can view an obligation as an internal value.[3] When many people in a community internalize an obligation, it becomes a social norm. People who internalize obligations express their commitment in various ways. Economic analysis of

* Professor of Law, University of California, Berkeley. This paper was first presented at the conference "Social Norms, Social Meaning, and the Economic Analysis of Law," University of Chicago Law School, Chicago, April 19, 1997.

[1] Raz reviews the imperative theory of law in Joseph Raz, The Concept of a Legal System (2d ed. 1980).

[2] I have explained this success (Robert Cooter, Laws and Prices: How Economics Contributed to Law by Misunderstanding Morality, in 3 Iuris: Quaderns de Political Juridica (Generalitat de Catalunya, Departament de Justicia, Direccio General de Dret i d'Entitats Juridiques i Formacio Especialitzada) 35–56 (1994)) and criticized the treatment of sanctions as prices (Robert Cooter, Law from Order, in A Not-So-Dismal Science: A Broader, Brighter Approach to Economies and Societies (J. Mancur Olson & S. Kahkonen eds. 1984)).

[3] H. L. A. Hart (The Concept of Law (1961)) has an especially influential discussion of the internal point of view toward law.

[*Journal of Legal Studies*, vol. XXVII (June 1998)]
 0047-2530/98/2702-0015$01.50

law, which has recently turned to the study of social norms,[4] has said little about their internalization and expression.[5] This article attempts to build the foundations for an economic theory of expressive law. According to the expressive theory of law, the expression of social values is an important function of the courts[6] or, possibly, the most important function of the courts.[7]

A system of social norms typically has multiple equilibria.[8] Law can create a focal point by expressing values. A focal point can tip the system into a new equilibrium. The process of changing the equilibrium can create or destroy a social norm without changing individual values. Creating focal points is the first expressive use of law.

In addition, law can change the individual values of rational people. Internalizing a social norm is a moral commitment that attaches a psychological penalty to a forbidden act. A rational person internalizes a norm when commitment conveys an advantage relative to the original preferences and the changed preferences. I call such a change a "Pareto self-improvement."[9] By creating opportunities for Pareto self-improvements, law induces rational people to change their preferences. I analyze how law can tip aggregate behavior and change individual preferences by expressing values. Changing individual values is the second expressive use of law.

Social Norms

I begin by explicating some conventions that I follow in discussing social norms.[10] Social scientists sometimes use the term "norm" to mean "aver-

[4] For examples, see this issue or Symposium on Social Norms and the Law, 144 U. Pa. L. Rev. (1996).

[5] Recent discussions relating expressive law to economic reasoning are found in Lawrence Lessig, Social Meaning and Social Norms, 144 U. Pa. L. Rev. 2181 (1996); and Cass Sunstein, On the Expressive Function of Law, 144 U. Pa. L. Rev. 2021 (1996).

[6] Hart (H. L. A. Hart, Punishment and Responsibility (1968)) argues that expressing social judgments is one of the uses of criminal law.

[7] This was apparently Durkheim's view, as analyzed in David Garland, Punishment in Modern Society: A Study in Social Theory (1990). Note that I draw no connection between the emotive theory of law, which belongs to jurisprudence, and the emotive theory of the meaning of value, which belongs to epistemology.

[8] For pioneering work, see Jack Hirshleifer, Economic Behaviour in Adversity (1987), ch. 9 on Evolutionary Models in Economics and Law: Cooperation versus Conflict Strategies; Jack Hirshleifer & Juan Carlos Martinez Coll, What Strategies Can Support the Evolutionary Emergence of Cooperation? 32 J. Conflict Resol. 367 (1988).

[9] I introduced the phrase "Pareto self-improvement" and the underlying idea in Robert Cooter, Self-Control and Self-Improvement for the "Bad Man" of Holmes, B. U. L. Rev. (1998), in press.

[10] I also explained these conventions and adopted them in Robert Cooter, Normative Failure Theory of Law, 82 Cornell L. Rev. 947 (1997); Robert Cooter, Decentralized Law for a Complex Economy: The Structural Approach to Adjudicating the New Law Merchant, 144 U. Pa. L. Rev. 1643 (1996).

age behavior." For example, statisticians talk about the "normal distribution," and sociologists sometimes use "norm" to mean what people normally do, as opposed to what deviants do. According to this usage, a norm is a regularity. In contrast, philosophers often use "norm" to refer to what people *ought* to do. According to this usage, a norm is an obligation.[11] To illustrate the difference, men regularly take their hats off in a boiler room from inclination, and men take their hats off in church from obligation.

Many economists apparently believe that a behavioral theory can dispense with the distinction between regularities and obligations. This view is mistaken. I explain later in detail that obligations, which restrict people from acting on their inclinations, affect behavior in distinctive ways.[12]

Since this article focuses on obligations, my use of "norm" conforms to philosophical usage and contradicts statistical usage. Furthermore, I mostly discuss social norms. I place "social" before "norm" to indicate a consensus in a community concerning what people ought to do. By this convention, agreement about what people ought to do indicates a possible social norm, whereas disagreement indicates a struggle to create a social norm. Consensus over an obligation, however, is not enough for the existence of a social norm. Following the positive theory of law, I also require a social norm to affect what people do, not just what they say. In brief, I use "social norm" in this article to mean an *effective consensus obligation.* By this definition, a norm exists when almost everyone in a community agrees that they ought to behave in a particular way in specific circumstances, and this agreement affects what people actually do.

In a noncooperative setting, moral restraint is a disadvantage, rather like fighting with one hand tied behind your back. In cooperation ventures, however, moral restraint can increase productivity, so people with good character may enjoy an advantage over people with bad character. For example, agents who faithfully serve their principals increase the productivity of principal-agent relationships by reducing monitoring costs.[13] As another example, sellers who disclose the truth about their products promote commerce by providing buyers with valuable information at low cost. In general, the obligation to be faithful, truthful, fair, and reasonable lubricates cooperation.

Social norms can subordinate one group of people to another, as with India's caste system or segregation in the American South. This article,

[11] Georg Henrik von Wright, Norm and Action (1963).

[12] Amartya Sen, Rational Fools: A Critique of the Behavioural Foundations of Economic Theory, 6 Phil. & Pub. Aff. 317 (1997); Amartya Sen, Maximization and the Act of Choice, 65 Econometrica 745 (1997).

[13] Cooter, *supra* note 9.

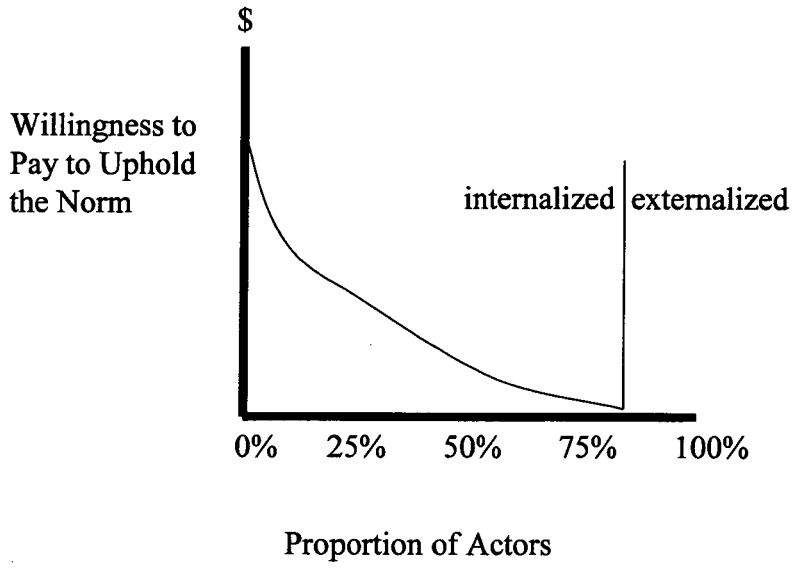

FIGURE 1.—Upholding a norm

however, does not discuss such norms. Many social norms contribute to productivity by increasing cooperation. In this article I only model social norms that contribute to productivity by increasing cooperation. I will show how such social norms, by their nature, create multiple equilibria, thus providing an opportunity for lawmakers to create focal points.

UPHOLDING A NORM

Upholding a social norm takes various forms, such as proclaiming commitment to an obligation, enforcing the obligation on others, or sacrificing in order to conform to the obligation. Upholding a social norm may cost money, time, or effort. In addition to the cost, upholding a social norm can yield advantages to the actor, such as deterring future injuries, undermining a competitor, or enhancing a reputation for honesty. The "net price" refers to the price paid by the actor minus the advantage he gains. According to the definitions used in this article, a person will pay a net price to uphold an *internal* obligation, whereas a person will *not* pay a net price to uphold an *external* obligation.

Internalizing a norm makes a person willing to pay a net price to uphold it. Figure 1 depicts willingness to pay to uphold a norm. The vertical axis indicates the net price the actor must pay to uphold a social norm. The horizontal axis indicates the quantity of actors, expressed as a percentage, who

are willing to pay the net price to uphold the norm. A few actors are willing to pay a lot to uphold the norm, many actors are willing to pay something, and some actors who externalize the norm are not willing to pay anything. The curve holds "tastes" constant in the sense of holding constant the strength of individual commitment to the norm. Internalization puts morality into preferences, not external constraints.

The curve in Figure 1 resembles final demand for an ordinary commodity. Like demand for an ordinary commodity, behavioral tests can measure willingness to pay to uphold a norm. For example, an experiment can present a subject with a choice between committing a wrong and receiving a payoff, or not committing the wrong and not receiving a payoff. Or an experiment can present a subject in a cooperative game with a choice between not sanctioning a wrongdoer or paying a price to sanction a wrongdoer. I assume that such an experiment would yield a distribution resembling the curve in Figure 1.

INTERIOR EQUILIBRIUM

A social norm imposes an obligation that partitions the set of possible actions into permitted and forbidden zones. People conform to a norm by staying in the permitted zone, and people violate a norm by entering the forbidden zone.

When an actor adopts the pure strategy of doing right or the pure strategy of doing wrong, the resulting payoff depends on the strategy pursued by others. As mentioned, this article only considers social norms that contribute to productivity by increasing cooperation. Under this assumption, an increase in the proportion of wrongdoers decreases the economy's productivity, which reduces the payoffs to wrongdoers and rightdoers. Although everyone's payoffs fall, the reduction need not be the same for rightdoers and wrongdoers. I will consider several possibilities with important consequences for equilibria.

I first discuss a unique, stable, interior equilibrium. In evolutionary equilibrium, all behavior that persists yields the same objective payoff.[14] Corner equilibria occur if one strategy yields the highest payoff to each actor when everyone follows it. Interior equilibria, in contrast, occur because when different people follow different strategies, all of them yield the same expected payoff.

In one common pattern, as the proportion of wrongdoers increases, the payoffs to wrongdoers fall faster than the payoffs to rightdoers. This possi-

[14] Abhijit Bannerjee & Jorgen W. Weibull, Evolution and Rationality: Some Recent Game-Theoretic Results, in Economics in a Changing World (B. Allen ed. 1996).

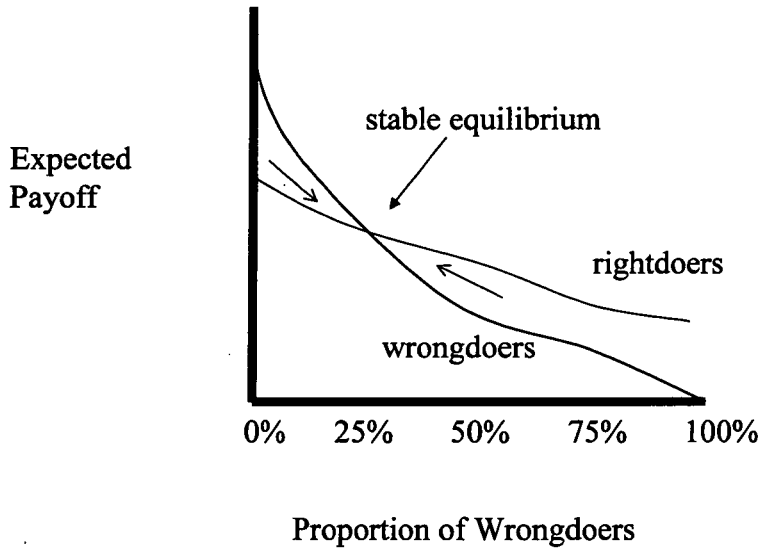

FIGURE 2.—Stable equilibrium

bility results in a stable interior equilibrium. Figure 2 depicts this possibility. The vertical axis represents the payoff, and the horizontal axis represents the proportion of wrongdoers. One curve represents payoffs for conforming to the norm, and the other curve represents payoffs for violating the norm. The intersection in the curves depicts an interior equilibrium in which rightdoers and wrongdoers receive the same expected payoff.

Now consider why the equilibrium is stable. If the proportion of wrongdoers increases beyond the equilibrium level, the payoff to rightdoers rises above the payoff to wrongdoers. With rightdoers receiving higher payoffs, some wrongdoers change their behavior. The number of wrongdoers declines until equilibrium is restored. Conversely, if the proportion of wrongdoers decreases below the equilibrium level, the payoff to wrongdoers rises above the payoff to rightdoers, so wrongdoers increase in number until equilibrium is restored.

Having used a graph to describe a stable interior equilibrium in a system of social norms, I provide some possible examples. First, consider the agency relationship. As more agents become disloyal, principals devote more resources to monitoring agents. Diversion of resources into monitoring reduces expected payoffs below the level achieved with fewer disloyal agents. Everyone's payoffs fall, but not equally. More monitoring might allow principals to reward loyalty and punish disloyalty more often. Conse-

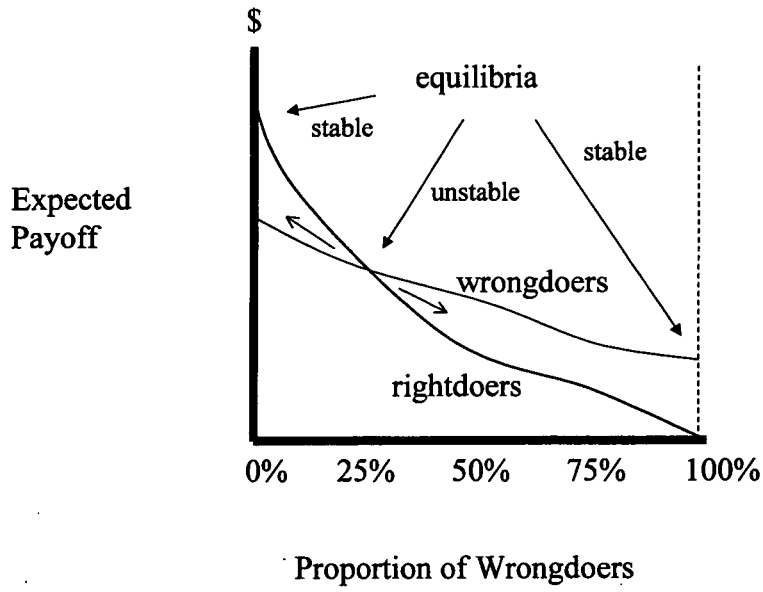

FIGURE 3.—Unstable equilibrium

quently, the loss in productivity from more monitoring probably harms disloyal agents more than loyal agents.

As another illustration, dishonest sellers often extract high profits per sale from low sales volume, whereas honest sellers extract low profits per sale from high sales volume. As the proportion of dishonest sellers increases, buyers buy less of the good, which harms all sellers, but not equally. More dishonest sellers cause buyers to increase their loyalty to honest sellers, thus leaving more dishonest sellers to compete for fewer buyers. So the loss in profits caused by more dishonest sellers might harm dishonest sellers more than honest sellers.

CORNER EQUILIBRIUM

As explained, stability results from assuming that an increase in the proportion of wrongdoers harms wrongdoers *more* than rightdoers. Now change the assumptions and assume that an increase in the proportion of wrongdoers harms wrongdoers *less* than rightdoers. Figure 3 depicts the situation. The intersection of the curves represents an interior equilibrium. The interior equilibrium, however, is unstable. Beginning from the interior equilibrium, an increase in wrongdoers causes the payoffs of wrongdoers to rise above the payoffs to rightdoers, so the number of wrongdoers contin-

ues to increase. The process ends at the stable equilibrium at the lower corner where everyone does wrong.

Conversely, beginning from the unstable interior equilibrium, a decrease in wrongdoers causes the payoffs of wrongdoers to fall below the payoffs to rightdoers, so the number of rightdoers increases. The process ends at the stable equilibrium at the upper corner where everyone does right.

I have explained that the system in Figure 3 stabilizes when everyone does right or everyone does wrong. The system goes to a corner because an increase in the proportion of wrongdoers harms wrongdoers *less* than rightdoers. Several possible causes could explain this possibility. Upholding a norm often involves confrontation. As fewer people uphold a norm, doing so becomes more risky. For example, the risk of confrontation from criticizing a smoker in a public building increases as more people in the building smoke. Similarly, the risk of retaliation from dismissing a disloyal agent presumably increases as more agents become disloyal. Finally, the risk of boycotting a dishonest seller presumably increases as more sellers become dishonest. When upholding a norm involves confrontation, the system may resemble Figure 3, in which case the system settles at a high level of conformity to the norm, or at a low level of conformity, but not at a level in between.

Racial discrimination in the American South provides a possible example of changing from one equilibrium to the other in Figure 3. During the period of segregation, social norms punished people for refusing to discriminate. Consequently, no individual or small group could abolish the discriminatory social norms. After the law imposed desegregation, new social norms developed to punish discrimination. Consequently, no individual or small group could engage in discrimination without paying a price. Thus the system arguably jumped from a high level of discrimination to a low level of discrimination.

Mixed Equilibria

For norms of cooperation, the curves expressing payoffs to rightdoers and wrongdoers slope down to express the loss in productivity as the proportion of wrongdoers increases. Theory, however, does not prescribe the gradient of the curves. The curves might intersect more than once. Figure 4 depicts this possibility. Figure 4 has a stable interior equilibrium with few wrongdoers, an unstable interior equilibrium with many wrongdoers, and a stable corner equilibrium with all wrongdoers. In Figure 4 the system stabilizes when few people (25 percent) do wrong or when everyone (100 percent) does wrong.

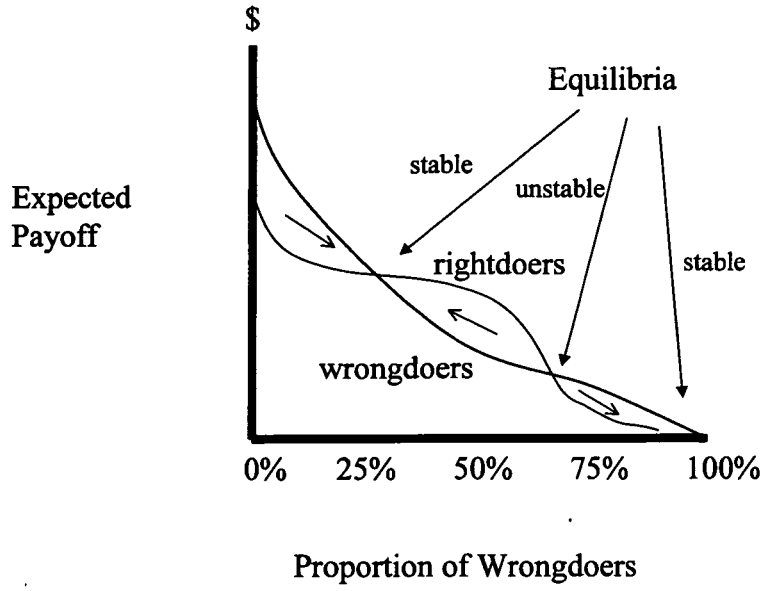

FIGURE 4.—Multiple equilibria

EXPRESSION AND FOCAL POINTS

The effects of enacting a law without enforcing it depend on the underlying characteristics of the system of social norms. If social norms form a stable equilibrium as depicted in Figure 2, enacting the law without enforcement causes modest benefits at best. The stable equilibrium in Figure 2 leaves no scope for an enduring change in behavior caused by changing expectations. Recall that the payoff curves in the figures are constructed assuming constant tastes. Enacting a law in Figure 2 without enforcing it will have no effect unless enactment causes tastes to change. "Tastes" in this context refers to the strength of the norm's internalization.

To illustrate how enacting a law might change tastes, assume that many people respect the law. When a new law is enacted, some people respond by devoting more resources to upholding it. Like state sanctions, informal sanctions can deter wrongdoing. If enacting the law induces more people to punish wrongdoing by informal means, then the payoff curve for wrongdoers shifts down in Figure 2, which moves the equilibrium a short distance to the left. The change in the equilibrium modestly increases everyone's payoffs.

Alternatively, enacting the law without enforcing it might have no effect.

To illustrate, assume that no one respects the law. When a new law is enacted, no one will pay a net price to uphold it. The payoff curves in Figure 1 do not shift, so the equilibrium remains unchanged.

I have shown that enacting a law without enforcing it can cause modest benefits provided that people respect law. Now I turn to more dramatic benefits from enacting a law without enforcing it. Unlike the modest benefits that I have been discussing, these dramatic benefits do not depend on enactment of the law causing tastes to change.

In Figure 3 and Figure 4, I showed how interdependent payoffs can cause multiple equilibria in a system of social norms. Given multiple equilibria, history and chance determine where the system settles. In the case of social norms, however, law can influence where the system settles by coordinating expectations.

Assume that a game has N players. In a first-order Nash equilibrium, no one can increase his payoff by changing his strategy so long as the other $N - 1$ players continue following their current strategies. If two or more players acted together, however, they might be able to increase their individual payoffs. Generalizing, in an n-order Nash equilibrium, no group of n actors can increase their individual payoffs by changing their strategies so long as the other $N - n$ players continue following their current strategies. If $n + 1$ or more players acted together, however, they might be able to increase their individual payoffs.

In a game with multiple Nash equilibria, some first-order equilibria may be Pareto inferior to others. If the system settles in a Pareto-inferior first-order equilibrium, no player acting on his own can improve his individual payoff. By acting together, however, a group of players usually has the power to change the equilibrium. Assume that the first-order Pareto-inferior equilibrium is an n-order disequilibrium. Thus, n people acting together can improve the payoffs to some players without harming anyone. Making the change requires n players to coordinate their behavior and change strategies together.

I will apply this reasoning to Figure 3 and Figure 4. As mentioned above, the imperative theory of law regards state sanctions as law's essence. This view comes from understanding law as a deterrent. Instead, think of law as solving a problem of collective action. Specifically, imagine a system of social norms stuck in a first-order Nash equilibrium that is Pareto inferior. To move to a Pareto-superior equilibrium, a group of actors must coordinate their behavior and change strategies. In an effective democracy, citizens respect the law and feel obligated to obey it. Lawmaking is a collective decision that could induce the coordination required to change to a Pareto-superior equilibrium.

To illustrate using Figure 3, assume that a system of social norms is stuck

at the lower corner where 100 percent of the actors do wrong. Another stable equilibrium exists at the upper corner where 100 percent of the actors do right. Everyone's payoff would increase if the system could move from the lower equilibrium to the upper equilibrium. Notice that the unstable, interior equilibrium occurs at the point where 25 percent of the actors do wrong and 75 percent do right. In order to move from the lower equilibrium to the higher equilibrium, at least 76 percent of the actors must change strategies and do right. Once 76 percent of the actors do right, the system will move to the upper equilibrium where 100 percent of the actors do right.

Perhaps enacting a law forbidding wrongdoing, without enforcing the law, can induce 76 percent of the actors to do right. If most citizens obey the law from respect, enacting the law without enforcing it can probably achieve the desired result. I have suggested that prohibiting smoking in American airports and requiring dog owners to clean up after their animals ("pooper-scooper" laws) work this way. Most people began to obey these laws as soon as they became aware of them. For the small recalcitrant group of lawbreakers, rude remarks by citizens and other informal punishments deter without state coercion.

According to the conventions adopted in this article, an obligation must affect behavior in order to count as a social norm. In my discussion of Figure 3 and Figure 4, I explained that enacting a law might change the equilibrium and cause most people to switch behavior from wrong to right. By making an obligation effective, the law can create a social norm. Behavior switches in this example while tastes remain constant. Thus enacting a law can change social values without changing individual values.

The expressive theory of law holds that eliciting voluntary obedience from most citizens makes law effective, and the effects may be greater than applying state sanctions to a few recalcitrant wrongdoers. In reality, a combination of expression and coercion accounts for the effectiveness of many laws. To illustrate using Figure 4, assume that a system of social norms is stuck at the lower corner where 100 percent of the actors do wrong. A stable interior equilibrium exists where only 25 percent of the actors do wrong. Everyone's payoff would increase if the system could move from the corner equilibrium to the stable interior equilibrium.

Notice that the unstable, interior equilibrium in Figure 4 occurs where 70 percent of the actors do wrong and 30 percent do right. In order to move to the stable interior equilibrium, at least 31 percent of the actors must change strategies and do right. Once 31 percent of the actors do right, rightdoers will continue to increase until 75 percent of the actors do right. In Figure 4 the law must induce only 31 percent of the citizens to change strategies in order to get a dramatically better result.

Assume that enacting a law without enforcing it induces at least 31 per-

cent of the citizens to change, so the system in Figure 4 moves to the stable interior equilibrium. Although improvement is dramatic, 25 percent of the actors continue doing wrong. Further reductions in wrongdoing would require state coercion. Supplementing informal sanctions with state coercion shifts down the curve representing expected payoffs to wrongdoers in Figure 4, thus reducing the equilibrium number of wrongdoers. The combination of expression and coercion brings wrongdoing down to a low level.

Information and Expressive Law

I have discussed examples in which enacting a law without enforcing it produces a dramatic improvement. Sometimes, however, enacting a law without enforcing it has no effect. Reinterpreting Figure 4 explains failures of expressive law. Expressive law succeeds in Figure 4 when enactment induces at least 31 percent of the actors to change. If, however, the law induces less than 31 percent to change, the system will eventually fall back to the original equilibrium. To be concrete, if 25 percent of the citizens in Figure 4 change their behavior and do right, with time everyone will lapse back into doing wrong.

Using law to create focal points requires information to make accurate predictions. With multiple equilibria, accurate predictions require knowledge of most or all of the payoff curves, not just knowledge of their slopes at the initial point. In other words, accurate predictions require nonmarginal information. To illustrate, the lawmakers in Figure 4 begin with a situation where everyone does wrong, yet the lawmakers need to know that an unstable equilibrium occurs where 30 percent of the citizens do right. In addition, the lawmakers need to know that at least 30 percent of the citizens will change their behavior in response to the law's enactment. So an effective use of expressive law demands a lot of information.

Scholars disagree about the extent to which courts can cause social change.[15] I believe that law breeds respect by tracking morality. To succeed in creating focal points, legal expression must enlist the natural sense of justice among citizens. Conversely, law breeds disrespect by imposing irrelevant or immoral obligations and asking more of citizens than they can ac-

[15] Consider the ability of courts to influence racial discrimination. For a pessimistic view, see Gerald Rosenberg (The Hollow Hope: Can Courts Bring about Social Change? (1993)), who argues that *Brown v. Board of Education* failed to integrate southern schools. For an optimistic view, see Lauren Edelman, The Endogeneity of Law: Constituting Law and Society in Organizations and Courts (paper presented at Univ. California, Berkeley, Law School seminar, Berkeley 1995). She argues that laws prohibiting discrimination get filtered through the structure and culture of organizations, where the modes of compliance symbolize conformity to law and become evidence for it. For example, to handle complaints of discrimination among workers, the corporation implements personnel procedures that mimic courts.

complish. Since lawmakers seldom possess nonmarginal information, attempts to create focal points by law can often produce cynicism. In special circumstances, instead of strengthening morality, law can crowd it out.[16] Lawmakers should proceed cautiously and skeptically with proposals for self-enforcing laws.

ENDOGENOUS PREFERENCE

Now I turn to analyzing how law changes individual values. Theories of endogenous preferences, which go back at least to Adam Smith,[17] have not flourished in economics.[18] Modern microeconomics trivializes moral commitment by treating it as an exogenous taste.[19] The renaissance in legal scholarship on social norms, although vigorous, suffers from the inability of economics to comprehend normative commitment. I will develop a theory of endogenous preferences and apply it to moral commitment and law.

First I extend the familiar concept of Pareto efficiency to explain why an actor satisfying economic standards of rationality might want to change his preferences. Figure 5 represents two public goods on its axes. Assume an initial allocation of resources that produces x_1 of the first public good and y_1 of the second public good. This allocation enables person 1 to achieve

[16] Crowding out of morality by law is a special concern of Bruno Frey. For example, see Bruno S. Frey, Felix Oberholzer-Gee, & Reiner Eichenberger, The Old Lady Visits Your Backyard: A Tale of Morals and Markets, 104 J. Pol. Econ. 1297 (1996); Bruno S. Frey, A Constitution for Knaves Crowds Out Civic Virtues, 107 Econ. J. 1043 (1997); Bruno S. Frey, Not Just for the Money (1997). Also note that competitive markets can reduce the reward for virtue by reducing the need for enduring relationships, whereas small, imperfect markets promote virtue by increasing the need for enduring relationships. In Brennan's attractive phrase, competition "economizes on virtue" (Geoffrey Brennan & Alan Hamlin, Economizing on Virtue, 6 Const. Pol. Econ. 35 (1995)).

[17] Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations (Random House 1937) (1776).

[18] Examples of endogenous preferences in economic theories include Gary S. Becker, Accounting for Tastes (1996); S. M. Goldman, Adaptive Behavior, Demand and Preferences, 1 Econ. & Phil. 189 (1985); Michael Hechter, The Role of Values in Rational Choice Theory, 6 Rationality & Soc'y 318 (1994); Albert O. Hirschman, Against Parsimony: Three Ways of Complicating Some Categories of Economic Discourse, in Rival Views of Market Society and Other Recent Essays (1986); R. A. Pollak, Habit Formation and Longrun Utility Functions, 13 J. Econ. Theory 271 (1976); R. Thaler & H. Shefrin, An Economic Theory of Self-Control, 89 J. Pol. Econ. 392 (1981); C. C. von Weizsacker, Notes on Endogenous Changes of Tastes, 3 J. Econ. Theory 345 (1971); M. Yaari, Endogenous Changes in Tastes: A Philosophical Discussion, in Decision Theory and Social Ethics (H. W. Gottinger & W. Leinfellner eds. 1977).

[19] The significance of the difference between morality as a preference and a constraint is explored in Matthew Rabin, Moral Preferences, Moral Constraints, and Self-Serving Biases (Univ. California, Berkeley, Dep't Economics, 1995).

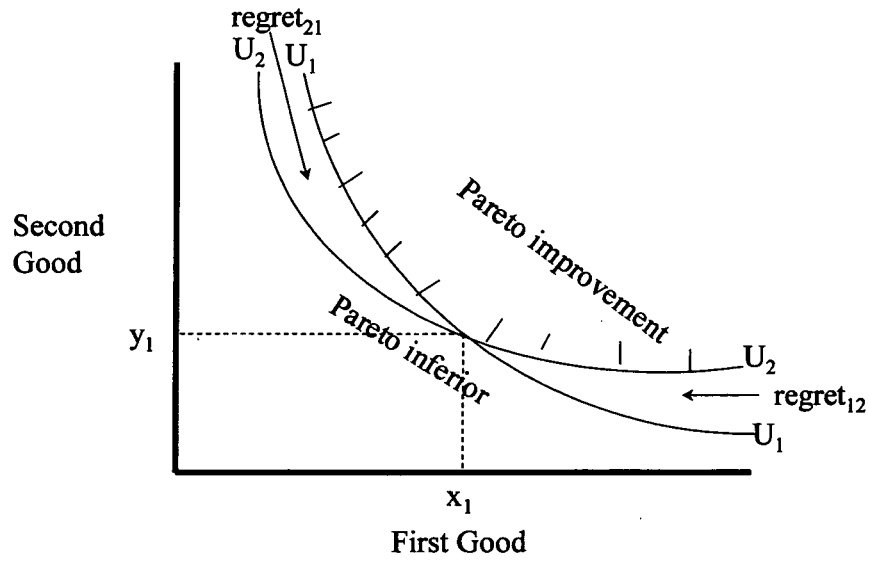

FIGURE 5.—Pareto improvement

utility U_1 and person 2 to achieve utility U_2. Hatch marks indicate the set of Pareto improvements relative to point (x_1, y_1).[20]

I will exploit the analogy between different people at the same time and the same person at different times. Reinterpret Figure 5 as depicting a single person with different preferences at different times. At time 1 the actor in Figure 5 enjoys the allocation of goods (x_1, y_1) that yields utility U_1. At time 2 the actor's preferences change to U_2. The hatch marks now represent Pareto improvements relative to point (x_1, y_1) for the same individual with different tastes. (With this reinterpretation, the goods represented on the two axes can be public goods or private goods.)

PARETO SELF-IMPROVEMENT

Now I use the concept of a Pareto improvement to explain why a person might deliberately change his preferences. Good character increases a person's value in cooperative activities. Participants in cooperative activities often get paid according to their value. So good character can convey an advantage in cooperative activities. For example, a person with more self-

[20] To generalize to many preferences and many goods, pick a starting point in n-space and draw m indifference curves through it. The upper envelope forms the boundary of the Pareto set.

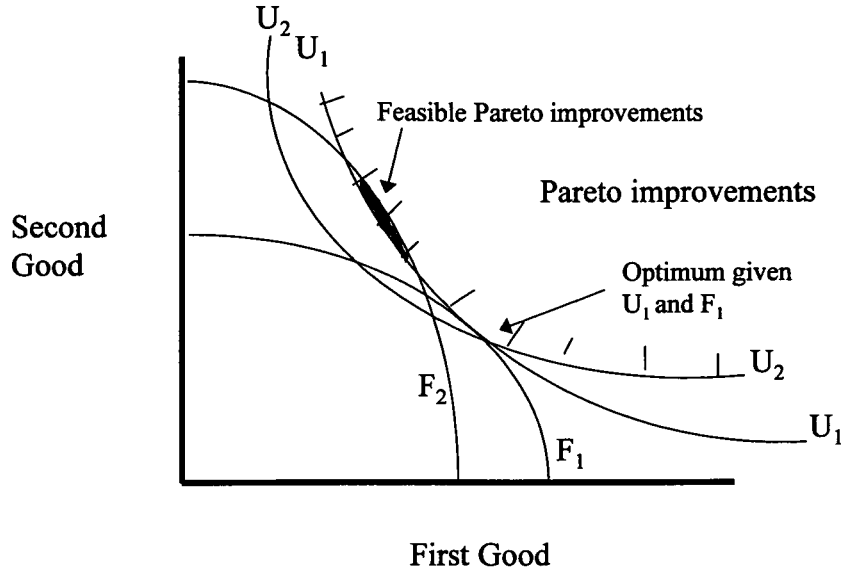

FIGURE 6.—Pareto self-improvement

control may have more opportunity to work in jobs that demand reliability. Similarly, a more honest person may have more opportunities to manage valuable assets.

To focus on the pure logic of choosing character, I will make two simplifying assumptions. To reward good character, people must observe it. One person can observe another's character imperfectly. Character is translucent, not transparent or opaque. At this stage in developing my theory, however, I want to avoid problems of information. Consequently, I will not analyze explicitly the problem of observing character.

Exactly how people develop their character remains murky. Presumably adults improve their character by the same means that parents apply to their children, such as cultivating good habits, choosing model associates, and acquiring moral or religious education. At this stage in developing my theory, however, I want to avoid specifying the technology for transforming character. Consequently, I will not analyze explicitly the problem of developing character.

I assume that the actor can choose his character and that the choice of character influences the opportunities available him. Figure 6 depicts a feasible set F_1 indicating the opportunities available to an actor with preferences U_1. Figure 6 also depicts a feasible set F_2 indicating the opportunities available to an actor with preferences U_2. Assume that an actor with prefer-

ences U_1 can choose to retain the same preferences U_1 and opportunities F_1 or acquire new preferences U_2 and opportunities F_2.

Would a rational actor make the change? The standard of Pareto improvements provides a compelling answer. Given preferences U_1 and feasible set F_1, the actor's initial optimum occurs at the point where F_1 is tangent to U_1, as indicated in Figure 6. As in Figure 5, the hatch marks in Figure 6 indicate the set of Pareto improvements relative to the initial optimum. Some of the Pareto improvements are feasible with opportunities F_2. Specifically, the shaded lozenge contains the *feasible* Pareto improvements. Thus the actor who changes preferences from U_1 to U_2 creates the opportunity for a better payoff as measured by original preferences and final preferences. In general, I used the phrase "Pareto self-improvement" to mean a change made by the actor in his preferences that makes feasible an allocation preferred by original preferences and final preferences.[21]

To illustrate, I apply my model to the work ethic that Weber attributed to Protestantism.[22] Assume that a worker can choose whether or not to join a religious sect and internalize a work ethic that values production and devalues leisure. To fit these assumptions, reinterpret the horizontal axis in Figure 6 as leisure and the vertical axis as income. Thus a person with preferences U_1 likes leisure, whereas the person with preferences U_2 internalizes the work ethic and likes income. An employer rationally expects a convert to such a sect to work more and relax less, so internalizing this ethic will improve the worker's opportunities to earn income and possibly reduce his opportunities to enjoy leisure. In Figure 6, F_1 indicates the worker's initial opportunities with preferences U_1, and F_2 indicates his opportunities after internalizing the work ethic and acquiring preferences U_2. Internalizing the work ethic is a Pareto self-improvement.

The concept of a Pareto self-improvement is apparently novel,[23] although a related idea has been discussed in the economics of advertising.[24] Some

[21] A stronger criterion would require that the allocation actually chosen with the new preference be preferred by the old preferences. This article relies on the concept of hypothetical Pareto self-improvements (an actual Pareto improvement is feasible), not the concept of actual Pareto self-improvements (a Pareto improvement is actually made). While the difference could be significant for some kinds of moral problems, I do not consider them in this article.

[22] Max Weber, The Protestant Ethic and the Spirit of Capitalism (Talcott Parsons trans. 1958).

[23] While developing this idea, I recalled the saying, "Law school is the hubcap where the loose nuts of social theory rattle around."

[24] Avinash Dixit & Victor Norman, Advertising and Welfare, 9 Bell J. Econ. 1 (1978); Avinash Dixit & Victor Norman, Advertising and Welfare: Reply, 10 Bell J. Econ. 728 (1979); Avinash Dixit & Victor Norman, Advertising and Welfare: Another Reply, 11 Bell J. Econ. 753 (1980). Dixit and Norman observe that advertising changes preferences, so they evaluate the consequences of advertising from the viewpoint of initial preferences and final

parallels can be found in the philosophy, especially where consequentialists defend morality as rational. For example, a prominent philosopher recently argued that the advantage a person gains from making a commitment provides a reason for carrying through later, even though the person subsequently can gain an advantage by not following through.[25] Nonconsequentialist philosophy often treats morality as rational, but rationality in nonconsequentialist philosophy hardly resembles economic rationality.[26]

WHY MAKE PARETO SELF-IMPROVEMENTS?

Applying the Pareto standard does not require the actor to compare one set of preferences to another. The individual who lacks a deep ethical theory can still make intrapersonal choices based on Pareto improvements. For example, the individual in Figure 6 has a reason to act without knowing whether preferences U_2 are inherently better or worse than preferences U_1. Nor does the individual have to know how much he would be willing to pay to change his preferences. The individual does not need a deep ethical theory to make intrapersonal choices causing Pareto improvements.

In contrast, intrapersonal choice among Pareto-efficient points requires a deep ethical theory and much information. To illustrate, assume that a dishonest seller can earn higher profits in a certain line of business than an honest seller can earn. To decide what to do, a seller must have an ethical theory that compares the value of honesty to its cost. Specifically, the ethical theory must say whether the intrinsic value of honesty exceeds its material disadvantage. Many people cannot decide such questions without soul-searching or agony.

With changing preferences, regret occurs when a choice produces a better result from the viewpoint of the initial preferences and a worse result from the viewpoint of final preferences. Since Pareto improvements are better from the viewpoint of the initial preferences and final preferences, the actor cannot regret a Pareto improvement. To illustrate, consider possible changes from the initial point (x_1, y_1) in Figure 5. The wedge between the utility curves, labeled "regret_{12}" in Figure 5, indicates points the actor would prefer with preferences U_1 and regret with preferences U_2. If prefer-

preferences. This approach resembles my own in this article, except I consider the individual as choosing whether or not to change his preferences.

[25] David Gauthier, Morals by Agreement (1985).

[26] Systematic Western philosophy is often traced to Plato, whose Republic inquires into the rational basis of justice. The more recent magisterial book by Rawls continues that inquiry (John Rawls, A Theory of Justice (1971)). Theories of rational morality that reject utilitarian reasoning often draw on Kant. For example, see Thomas Nagel, The Possibility of Altruism (1970).

ences change from U_1 at time 1 to U_2 at time 2, then a decision by the actor at time 1 to choose a point in the set regret_{12} would cause regret at time 2. The set of points indicated by hatched lines and labeled "Pareto improvement" in Figure 5 does not intersect the set of points labeled "regret_{12}" or "regret_{21}."

An actor cannot regret a Pareto self-improvement, but after exhausting the opportunities for Pareto self-improvements, further changes in character can cause regret. A rational person might be uncertain about how he will feel after changing his preferences. Uncertainty over possible regret might create psychological resistance to making the change.

To illustrate, a dishonest person might strive to become honest in the hope that the change will make him feel better about himself. After changing himself, however, instead of feeling better about himself, he might feel like a chump. If honesty makes him feel like a chump, then he will regret having become more honest. Instead, he might wish that he were a more effective liar. Recognizing the difficulty in predicting how he will feel after changing his preferences, a rational person feels more confident about Pareto self-improvements than changing himself in other ways. People with opportunities for Pareto self-improvements will tend to change their preferences, and, after exhausting the opportunities for such changes, people will encounter psychological resistance to further changes.

In addition to this positive reason, a normative reason commends using the Pareto criterion. When preferences change, some ethical theories favor the original preferences, and some ethical theories favor the final preferences. This fact creates a dilemma for evaluating public policies that change preferences. The Paretian standard avoids this dilemma. Policies that create opportunities for Pareto self-improvements respect the judgments of individuals about their preferences, rather than imposing a judgment on them about the superiority of some preferences to others.

The concept of Pareto self-improvements might help to revitalize cooperative game theory. The theory of cooperative games, which requires normative commitments from players, languishes while the theory of noncooperative games flourishes.[27] Excluding cooperation from game theory favors purity over reality. Experimental evidence indicates the pervasiveness of cooperation in spite of the requirements of narrow self-interest.[28] Players

[27] To illustrate, the classic textbook on game theory devotes a chapter to cooperative games (R. Duncan Luce & Howard Raiffa, Games and Decisions: Introduction and Critical Survey (1967)), whereas one of the best modern books omits it (Drew Fudenberg & Jean Tirole, Game Theory (1991)).

[28] Elizabeth Hoffman, Kevin McCabe, Keith Shachat, & Vernon Smith, Preferences, Property Rights and Anonymity in Bargaining Games, 7 Games & Econ. Behav. 346 (1994); Elizabeth Hoffman & Matthew L. Spitzer, Entitlements, Rights, and Fairness: An Experimental

TABLE 1

IMMEDIATE AND FUTURE MONEY PAYOFF TO HONESTY AND DISHONESTY

	Social Sanctions for Dishonesty		Social and Legal Sanctions for Dishonesty	
Honest	w_1, w_2	Low	w_1', w_2'	Highest
Dishonest	$w_1 + b, \quad w_2 - c$	Moderate	$w_1' + b', \quad w_2' - c$	High

who "irrationally" cooperate often gain an advantage in competition with narrowly instrumental players, thus straining the definition of rationality.[29] In experimental economics, the initial discovery of the resilience of moral commitment has yielded to progressive refinements that explain what people are committed to.[30]

In order to command allegiance, social norms require justification. The requirement of justification restricts the behaviors that can become obligatory. To illustrate, accepted standards of morality cannot justify the proposition, "Everyone but me should tell the truth," so this proposition cannot become a social norm. By eliminating strategies that cannot sustain social norms, the theory of cooperative games could alleviate the problem of too many equilibria that plagues game theory.[31]

LEGAL INCENTIVES FOR SELF-IMPROVEMENT

Now I explain some ways that law can change preferences. Some legal theorists believe that the essence of law is an obligation backed by a coercive state sanction. I begin by showing how coercive sanctions attached to acts can change preferences. Specifically, I extend Figure 6 to show how contract law creates opportunities for Pareto self-improvements.

Assume that the state chooses whether or not to enforce contracts, and the actor chooses whether to be honest or dishonest. Table 1 indicates the money payoffs from these four possibilities. Without contract law, honesty

Examination of Subjects' Concepts of Distributive Justice, 14 J. Legal Stud. 259 (1985); Iris Bohnet, Fairness, Inequality and the Identifiable Victim Effect: A Behavioral Institutional Analysis (paper read at the Seminar on Law, Economics, and Organizations, Univ. California, Berkeley, April 1998).

[29] R. Axelrod, The Evolution of Cooperation (1984); Robert Frank, If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience? 77 Am. Econ. Rev. 593 (1987); Robert Frank, Passions within Reason: The Strategic Role of the Emotions (1988).

[30] Hoffman et al., *supra* note 28; Hoffman & Spitzer, *supra* note 28.

[31] The Folk Theorem formulates the problem of multiple equilibria. See Drew Fudenberg & Eric Maskin, The Folk Theorem in Repeated Games with Discounting or with Incomplete Information, 54 Econometrica 533 (1986).

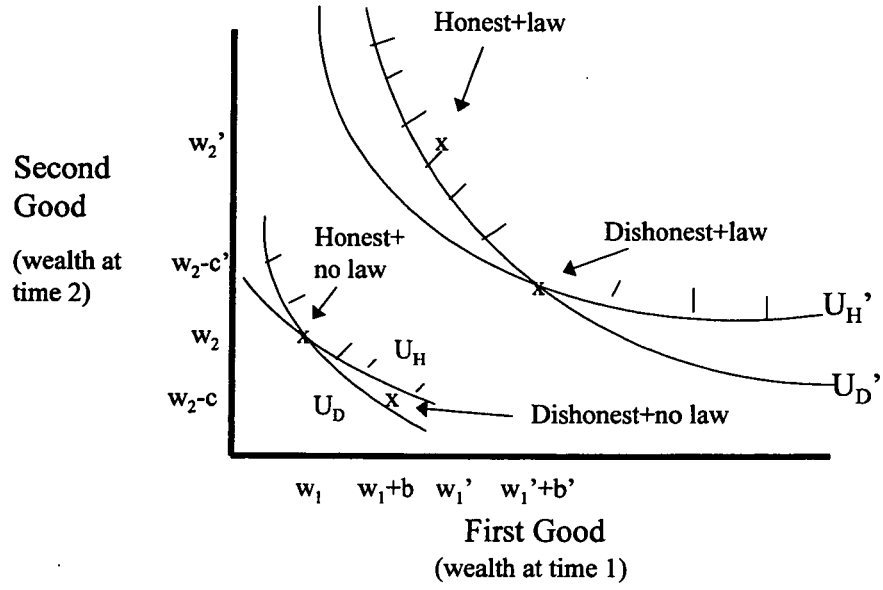

FIGURE 7.—Promise keeping with and without contract law

yields an immediate payoff of w_1 and a future payoff of w_2. Without contract law, dishonesty yields an immediate payoff of $w_1 + b$ and a future payoff of $w_2 - c$. According to Table 1, social sanctions for dishonesty are not very effective with respect to the promises under consideration. Consequently, given social sanctions and no legal sanctions, the immediate benefit from dishonesty outweighs the future cost. The gain b from dishonesty is larger than the modest social sanction c, so dishonesty pays better than honesty in the absence of contract law.

Without enforceable contracts, people have difficulty cooperating with each other, so productivity is relatively low. With enforceable contracts, people cooperate more, so productivity is relatively high. Consequently, the honest and the dishonest people enjoy a larger payoff with contract law than without it. According to Table 1, the payoff to honesty with contract law increases to w_1' in the first period and w_2' in the second period. Given enforceable contracts, dishonesty yields the immediate payoff $w_1' + b'$ in the first period and $w_2' - c'$ in the second period. With contract law, the sanction for dishonesty is social and legal. According to Table 1, legal sanctions for dishonesty are effective with respect to the promises under consideration, so honesty yields a higher overall payoff than dishonesty.

Now I evaluate the four possible outcomes described in Table 1 according to the preferences of honest and dishonest actors. The horizontal

axis in Figure 7 indicates payoffs in the first period, and the vertical axis indicates payoffs in the second period. An honest person has different preferences from a dishonest person. To keep the representation simple, I assume that an honest person applies a low discount rate to future payoffs, whereas a dishonest person applies a high discount rate. In Figure 7, the dishonest preferences indicated by U_D and U'_D and give relatively more weight to wealth in time 1 and less weight to wealth in time 2. Conversely, the honest preferences indicated by U_H and U'_H give relatively less weight to wealth in time 1 and more weight to wealth in time 2.

Consider the point in Figure 7 labeled "Honest + no law," which indicates the payoff to being honest without contract law. Compare this point to the point labeled "Dishonest + no law," which indicates the payoff to being dishonest without contract law. The dishonest person prefers the high present payoff and the low future payoff from dishonest behavior, rather than the low present payoff and the high future payoff from honest behavior. The honest person, however, has the opposite preference. The hatch marks on the utility curves indicate the Pareto improvements relative to the preferences of an honest person and a dishonest person. In the absence of contract law, a Pareto self-improvement is impossible, so a dishonest person and an honest person prefer to remain as they are, rather than changing their preferences.

Contract law, however, produces a different result. State sanctions make dishonesty less attractive. The point in Figure 7 labeled "Dishonest + law" indicates the payoff to being dishonest with contract law, whereas the point labeled "Honest + law" indicates the payoff to being honest with contract law. In Figure 7, the dishonest person prefers the payoff received by the honest person rather than his own payoff. The honest person in Figure 7 also prefers his payoff to the payoff received by the dishonest person. So contract law creates a situation in which a person who changes from being dishonest to being honest makes himself better off relative to his initial and final preferences. Thus contract law creates the opportunity for a Pareto self-improvement where none existed without contract law. In general, *the law prompts improvement in character whenever a legal sanction creates an opportunity for a Pareto self-improvement.*

In Figure 7, the increase in productivity caused by contract law is so great that everyone is better off relative to their initial preferences and their improved preferences. Recognizing these facts, cynics who place no intrinsic value on keeping promises and moralists who place high intrinsic value on keeping promises might agree that the state should enforce contracts.

I have shown how coercive state sanctions can cause rational people to change their character. The same argument might extend to a more manipulative state policy to enhance promise keeping. To illustrate, assume that, instead of liability, the state could shame people who break contracts by

publicizing their misdeeds.[32] Furthermore, assume that shaming is more effective than liability for changing peoples' character, so shaming induces more promise keeping at less cost than liability for certain kinds of contracts. Replacing liability with shaming for this class of contracts might make some people better off relative to their initial and final preferences, and the policy makes no one worse off. Under these assumptions, everyone, including cynics, might agree to replace liability with shaming as the sanction for breaching certain types of contracts.

COMMITMENT AND EMOTION

Something has meaning that conveys information by symbols. To illustrate, graffiti on a wall has meaning, whereas the marks on the wall from weathering have no meaning. Some symbolic acts express the actor's commitment to internalized values. Expressing commitment is one way to uphold a norm. Thus Figure 1 can be interpreted as depicting willingness to pay to express commitment to a norm. According to this interpretation, more people will express their commitment to norms when doing so costs less.

This proposition figures prominently in the economic analysis of the state. Economists are familiar with designing institutions to align self-interest and the public interest. Another strategy severs the relationship between them so that the actor can express his views about right and wrong at no personal cost. To illustrate, constitutions often strive to make judges independent and disinterested. When this goal is achieved, the decisions of a judge do not influence his power or wealth, so the material costs are negligible for the judge to express his views about right and wrong. These facts have lead theorists to propose that the motive of some judges is to express their political and moral vision.[33] Judging from the financial sacrifice, some lawyers will pay a lot to become judges and express their political and moral vision. Similarly, given a secret ballot and a large electorate, the way an individual votes does not influence his wealth or power. Under these circumstances, the voter, like the judge, may want to express his political and moral vision.[34]

I have been discussing the cost of expressing moral commitment. Sometimes, however, the expression of moral commitment yields a net benefit instead of a cost. As explained, the internalization of morality can convey

[32] Dan M. Kahan, Social Influence, Social Meaning, and Deterrence, 83 Va. L. Rev. 349 (1997).

[33] Richard Posner, What Do Judges Maximize? (The Same Thing Everybody Else Does), 30 Sup. Ct. Econ. Rev. 1 (1993).

[34] Geoffrey Brennan & Loren Lomasky, The Pure Theory of Electoral Preference (1993).

a competitive advantage in cooperative activities. To do so, people must observe the actor's commitment. I cannot survey the means of signaling commitment in this article, but I will discuss briefly the role of emotion.

Moral commitment can be fake or genuine. Genuine moral commitment has an emotional aspect. People *feel* committed to internalized values. The connection between commitment and emotion has a useful function. Telling a cool lie is easier for many people than faking emotion. For example, children who can tell a cool lie are often incompetent at faking emotion. Aspiring actors devote much time and effort to perfecting the art of faking emotion. Thus the emotion attached to the expression of moral commitment helps to authenticate it. According to one theory, emotions evolved among people partly to provide the means to signal commitment.[35]

Whereas economic rationality seems relatively cool, discussion in politics and law seems relatively hot. The heat comes from the connection between emotion and expression. To illustrate, people often contest the symbolic values in laws concerning issues such as abortion,[36] discrimination,[37] or even closing the range to cattle.[38] The presence of emotion in law and politics suggests the prominent place of expressing internalized values.

CONCLUSION

Some people obey most laws from fear, and all people obey some laws from fear. The economic analysis of deterrence explains this behavior. Social psychologists have accumulated impressive evidence, however, that most people obey most laws from internalized respect.[39] I try to explain this behavior by developing an economic analysis of expressive law.

Expressing commitment to internalized norms has intrinsic value. This article, however, mostly concerns the extrinsic value of expression. In a system of social norms with multiple equilibria, expressing commitment can change the equilibrium by providing a focal point, even without invoking coercive force. The change in the equilibrium can change social norms without necessarily changing individual values. Law provides an instrument for changing social norms by expressing commitments.

Moralists have long understood that sanctions for wrongdoing create incentives for improving oneself, but this idea has eluded economic models. My formulation of Pareto self-improvement should bring this idea under

[35] Frank, Passions within Reason, *supra* note 29.

[36] Kristin Luker, Abortion and the Politics of Motherhood (1984).

[37] Edelman, *supra* note 15.

[38] Robert C. Ellickson, Order without Law: How Neighbors Settle Disputes (1991).

[39] Tom R. Tyler, Why People Obey the Law (1990).

the analytical power of economic models. More generally, the concept of Pareto self-improvement extends economic reasoning to endogenous preferences and the internalization of norms. Coercive state sanctions can induce people to internalize norms by creating opportunities for Pareto self-improvements. Internalization of social norms decentralizes law and increases production through cooperation. By reducing the need for state coercion, voluntary obedience makes liberal government possible.