CHAPTER 21

LEGAL THEORY AND
THE RATIONAL ACTOR

CLAIRE FINKELSTEIN

1. Introduction

The application of rational actor theory to the law has been dominated by the law and economics movement. Law and economics began as a methodology for approaching legal questions; it suggested using the tools provided by economic theory to solve problems in manifestly economic fields like antitrust and taxation. But it quickly moved beyond methodology to acquire both a normative and a descriptive thesis of its own. The normative thesis is that the legal system should serve the goal of maximizing social welfare. The descriptive thesis is that the common law has developed according to this implicit economic logic, even if judges and other legal actors have not consciously tried to fashion legal rules with economic theory in mind. Law and economics thus regards law as a vehicle for maximizing utility, and it maintains that much of law can be explained as serving that function already.1

In what follows, I shall leave law and economics's descriptive thesis about the nature of legal rules to one side and focus on its normative claim instead. On the normative side, law and economics shares a theory of value with act-utilitarianism in moral theory: Both posit some felicific notion like utility, well-being, happiness, or pleasure as the sumnum bonum for human beings. There are, however, several important differences between the two theories. First, legal economists have a
tendency to treat the notion of utility more generically than do ethical utilitarians. They include in it everything that is a source of value to an agent, whether or not it confers pleasure or happiness. Thus the legal economist might treat a person’s altruistic feelings, aesthetic sentiments, and moral commitments as part of his utility, since these are all things a person values. Arguably, this makes the notion of utility vacuous, since it is not clear what could falsify the claim that a rational actor seeks to maximize his utility. (The economist, however, actually embraces this result, since he constructs an agent’s utility function externally, namely from the choices he observes that agent making, rather than on the basis of the agent’s internal psychological state.) For purposes of argument, then, I shall assume a somewhat narrower conception of utility, namely one that roughly corresponds to the idea of personal well-being.2

A second difference between law and economics and act-utilitarianism has to do with the object of evaluation. While utilitarianism treats individual acts as the relevant units from which maximizing must take place, law and economics maximizes from the standpoint of legal rules. Law and economics thus proposes a two-tier theory: Legal rules should maximize social welfare, while individual acts falling under those rules need not be themselves maximizing. In other words, law and economics is a kind of rule utilitarianism, where the relevant rules are legal, rather than moral, in nature.

A third difference will provide the focus for our discussion. Law and economics is generally accompanied by a descriptive thesis about human nature, one that ethical utilitarians do not normally share. Legal economists tend to assume that the utilitarian theory of value to which they subscribe goes hand-in-hand with what we might call “rational actor psychology.” According to this view, human beings are rational maximizers who reason instrumentally toward the attainment of their ends. One can understand why the legal economist thinks the utilitarian theory of value presupposes a rational actor model of human agency: If society’s good consists in maximizing social welfare, it seems natural to suppose that individual good consists in maximizing personal welfare. But a moment’s reflection should make clear that the psychological theory is neither presupposed nor entailed by the theory of value, since we need not think of the subjects of a utilitarian regime as capable of reasoning about their own good. We could, for example, ask what the best life for a cow would be and seek to maximize its utility by providing it with grassy fields and plenty of water. But we need not think cows capable of utility maximization on their own behalf, still less of anything like instrumental reasoning. The only requirement that social utility theory does impose on the creatures to whom it applies is that they be capable of experiencing pleasure and pain, since without this we could not meaningfully speak of their having any utility or well-being to maximize.

Not only is rational actor psychology not inherently linked to the utilitarian theory of value, but the two are actually in some tension with one another. The tension stems from the fact that when we maximize social welfare, we do not necessarily maximize the welfare of each individual member of society. In most cases, maximizing social welfare will result in some people faring worse than they otherwise would, even though other people fare better. This fact produces a quick route to the central claim I wish to make: If human beings are assumed to be rational maximizers, there is no reason to suppose they will be social maximizers as well.4 And this suggests that a theory like law and economics that subscribes to a utilitarian theory of value at the same time that it adopts a rational actor psychology has some explaining to do. It must have a way of showing why the utilitarian theory of value it proposes should recommend itself to creatures with the particular constitution of rational maximizers. Indeed, I shall argue that rational actor psychology more naturally leads to a contractarian than to a utilitarian theory of legal rules.

We must now back up, however, and take the long route to my conclusion. For the legal economist thinks he has a ready answer to the above concern. Unlike act-utilitarianism, law and economics does not actually require individual agents to maximize social utility. It is only legal rules that must maximize social utility, while individuals living under those rules need think only of their own good. Good legal rules will bring the incentives of individual agents into line with the requirements of social welfare. Thus individuals living under correctly crafted legal rules will maximize society’s utility when they attempt to maximize their own utility. Utilitarianism and rational actor psychology can co-exist in law and economics, then, because it is possible to adopt social rules that equip rational, self-interested agents with incentives to maximize social welfare. In what follows, I shall nevertheless take issue with this way of reconciling rational actor psychology and the utilitarian theory of value. I shall suggest that the tension between the two persists as a result of two central features of legal rules. First, I argue that legal rules must be created and administered by individuals who are themselves rational maximizers, and thus they must see themselves as personally advantaged by the rules they institute. Second, I argue that if human beings are rational maximizers, we ought to be able to justify a legal principle like utility maximization by showing that rules adopted in accordance with that principle would be in their interest. Yet there is reason to think that rational agents would not select rules that maximize social welfare. Indeed, as I shall argue, there is reason to suppose they would select rules premised on some sort of contractarian norm of agreement instead.
2. Rational Actor Psychology and Legal Actors

The first problem with the legal economist's picture arises because rational actor psychology must be presumed to apply to what I shall call "legal actors," namely people who create and implement legal rules. The tension between rational actor psychology and the utilitarian theory of value will thus resurface, since there is no reason to suppose that legal actors such as judges, jurors, and legislators will maximize their utility by writing and implementing legal rules that maximize social utility. There would be no difficulty here if some other, perfectly utilitarian entity were to determine social and political arrangements for us. For in this case, like the crows about whose good we might reason, the motivations of the rule-makers need not mirror the motivations of the creatures to whom the rules would apply. If legal economists maintain that rational human beings would adopt and faithfully apply rules that maximize social utility, they should also be able to tell us why doing so should be in the interests of a rational legal actor.

Some legal economists have noticed the difficulty, and they have sought to bring their account of the incentives of legal actors into line with their general account of individual motivation. Consider, in particular, the attempts legal economists have made to explain judicial behavior. Robert Cooter, for example, argues that public judges are primarily driven by the desire for prestige, and this leads them to act "in such a way that they would be chosen to decide cases by litigants and their lawyers if choice were allowed" (1983, 107). They will thus maximize their personal utility if they faithfully apply the rules as they are written to the cases that come before them. And since those rules are ideally welfare-maximizing, judges will maximize social welfare when they maximize their own welfare.

Richard Posner gives a slightly different explanation. He argues that the judicial utility function is based on the pleasure judges experience in voting on cases, since it is by rendering decisions that they have a sense of their own power. (1995, part 1, sec. 3). He analogizes this pleasure to the satisfaction members of the ordinary public receive from voting in elections—the sense of one's own importance that comes from the fact that one is entitled to decide. Consistent with this, he argues, is the fact that judges do not particularly care about writing opinions. They see their function as residing in the decisions they render, rather than in the reasons they might articulate for their decisions. But given the job and salary security most judges enjoy, this does not explain why judges render decisions that adhere to legal rules, instead of deciding cases based on personal or idiosyncratic preferences. Why, that is, do judges care about the rule of law, when it does not maximize their personal welfare to do so?

Posner's answer is that adherence to the rules of judging is part of what makes the activity of judging pleasurable, much in the way that adherence to the rules of chess helps to make chess pleasurable (129). Judges adhere to the rule of law because they get more pleasure from following the rules, knowing it is this that defines their role, and hence their power, than they would if they decided cases for reasons of personal aggrandizement. He writes:

The reason is not that judges have different utility functions from those of other people; it is that the utility they derive from judging would be reduced by more than they would gain from giving way to [various] temptations... It is the same reason why many people do not cheat at games even when they are sure they could get away with cheating. The pleasure of judging is bound up with compliance with certain self-limiting rules that define the "game" of judging. ... It is by doing such things that you know you are playing the judge role, not some other role; and judges for the most part are people who want to be—judges. (131)

We could presumably tell a similar story about legislators. People who become legislators enjoy constructing rules that maximize social welfare. They are "public spirited," in the sense that their own utility increases to the extent they feel they are serving the public good. Indeed, carrying the parallel further, we might say that legislators derive their sense of their own power from their ability to create social legislation that incorporates their view of what the public good requires. For this reason, there need be no tension between the demands of individual welfare and those of social welfare. Presumably we could explain the behavior of jurors along similar lines.

There are several problems, however, with this attempt to reconcile rule of law values with the hypothesis that legal actors are rational utility maximizers. I shall focus on Posner's account of judging, but similar criticisms could be made of the extension of that account to the other legal actors I have suggested. First, like the expansive approach Kaplow and Shavell take to the notion of utility, the rational actor psychology Posner's analysis assumes makes the claim that human beings are self-interested maximizers trivially true. If, for example, an agent has a strong moral compunction not to lie, cheat, or steal, this account will purport to explain the phenomenon by saying that he derives more utility from adhering to moral rules than from breaking them, since only then can he think of himself as playing a moral role from which he derives pleasure. Once again, this suggests that there is no motivational state that would falsify the hypothesis that legal actors are rational maximizers. Of course moral agents probably do derive pleasure from behaving morally. But we cannot use that pleasure as itself explanatory of the person's behavior if we wish to make a nontrivial assertion that the behavior is motivated by self-interest.

Second, Posner's account maintains that adhering to legal rules has instrumental value for a rational judge. But no account of this form is likely to succeed,
for it fails to capture the experience of judges in relation to legal rules. Judges do not usually approach legal rules instrumentally. Instead, they regard themselves as bound by those rules, and do not feel free to ignore them if their personal utility would be enhanced by doing so. On Posner’s story, in other words, we cannot explain the possibility of true internal commitment to the rule of law, since we cannot tell an instrumental story about such commitment.

Third, if Posner were right, judges would systematically violate legal rules when, in a given case, they would gain more utility from breaking the rules than from abiding by them. The widespread adherence to legal rules among judges, even in cases in which individuals would have much to gain from violating the rule, belies this suggestion. It is possible, of course, that judges receive such positive utility from following legal rules that it is nearly impossible for them to receive more from violating those rules. But, as I have suggested, this hypothesis is either trivial or highly implausible: It is trivial if we have no independent measure of a judge’s utility than what we discover he happens to choose, and it is implausible if the notion of utility is taken seriously as an independent variable, and the claim is that following legal rules confers such great personal pleasure that no bribe or other extra-legal inducement could outweigh it.

I thus return to my suggestion that insofar as legal rules and institutions must be created and implemented by rational agents, there is a tension between the utilitarian theory of value the legal economist espouses and the rational actor psychology to which he subscribes. For the theory to work, legal actors must be shown to be personally benefited by adherence to legal rules, and the few attempts that have been made to show this seem unpromising.

3. Rational Actors and the Social Contract

The second problem with the legal economist’s proposed reconciliation of rational actor psychology and the utilitarian theory of value requires us to look back to an earlier point at which general principles for the establishment of legal institutions would be selected. What we must ask is what principle or set of principles putative members of a legal regime would settle on as the basis for constructing the legal institutions under which they must live. That is, what kind of legal institution would individual maximizers see as most in their interest to adopt? Broadly speaking, there are three possibilities: They could adopt institutions that follow the principle of utility maximization; they could establish institutions based on specific deontological commitments; or they could adopt legal institutions premised on the principle of mutual benefit. In other words, rational individuals might choose to make their legal institutions utilitarian, deontological, or contractual. In what follows, I shall attempt to show why I think rational actors would not select the principle of utility as the theory of value for legal rules. I shall claim that they would also be unlikely to select deontological norms to structure legal rules. I conclude that rational agents selecting basic legal institutions are likely to select substantive norms of contractual agreement. I shall attempt to demonstrate these claims by considering two examples of possible legal institutions.

First consider a somewhat simplified version of an accident law regime proposed by Kaplow and Shavell (2002, III. C). Suppose rational actors are given the choice between two different regimes: allow people to drive and assign liability for resulting accidents according to fault, or allow people to drive but let losses lie where they fall. (A third option would be to disallow driving entirely, but for present purposes I will assume this option would not be adopted.) In the first regime, a victim could sue an injurer and recover damages for his losses if the injurer was negligent in causing the victim’s injuries. In the second regime, no victim could sue any injurer. Which regime would rational actors choose? Suppose that each person could expect to be both an injurer and a victim exactly once in his life, so that the risks imposed between injurers and victims were “reciprocal.” In this case, Kaplow and Shavell argue, the parties would be indifferent between being able to sue for damages and leaving everyone to cover his own accident expenses. Now suppose there are high administrative costs involved in running a liability scheme, so that it is expensive for society to adjudicate and administer liability suits brought by accident victims. In this case, rational agents would choose to let losses lie where they fall, since the savings in administrative costs constitute a surplus that can be divided among the parties. The no-liability regime would leave at least one, and possibly all of the parties, better off than they would be under the liability regime—that is, the no-liability regime would be pareto superior to the liability regime. Thus, they conclude, the claim that a tort liability scheme is a required part of corrective justice should be rejected. Rational agents would not choose to equip themselves with a fairness-based remedy when a different legal scheme would leave at least one person better off and no one worse off.

Now consider the more realistic “non-reciprocal” case, the case in which the parties do not know when or whether they will be injurers or victims. In the absence of more specific knowledge of their situation, the parties should be indifferent between this case and the case in which the risk-taking is reciprocal. If there are high administrative costs to running a liability regime, Kaplow and Shavell once again argue, rational agents would choose the regime that leaves losses where they fall, since that regime would again be pareto superior to the
more expensive liability regime. As a general matter, Kaplow and Shavell assert that rational agents who do not know their ultimate positions under the regime they select will maximize their expected utility by selecting the regime with the highest social return. They would never choose a legal regime on grounds of "fairness." Any such regime could leave everyone worse off, since even the worst off group in an unfair regime might have a greater absolute share than they would in a regime where returns are distributed fairly.

For this reason, Kaplow and Shavell may be correct that rational agents would not choose to equip themselves with a legal regime predicated on norms of fairness, and so to dispense with the suggestion that rational agents would select deontological norms as such to govern specific legal institutions. It hardly follows, however, that they would adopt the accident regime dictated by utilitarian norms. For the fact that a rational agent has a higher expected utility from a given legal regime does not necessarily mean that he would choose that regime, even if he is choosing strictly on the basis of self-interest. To see this, let us make use of an assumption that is common among legal economists, namely that a person who engages in risky behavior in some way benefits himself at the expense of others. The no-liability regime is thus one in which injurers improve their own position without having to pay for it, and victims are harmed without receiving compensation. Individuals living under such a regime are enrolled in a lottery, in which the injurers are the winners and the victims are the losers. By contrast, the liability regime would be one in which the parties' post-accident welfare would be restored to its pre-accident state through court-ordered damage awards. In such a regime, there would be no winners and losers once compensation was paid. Now the question we might ask is: Would rational agents prefer the regime in which they gamble, hoping they will be non-liable injurers rather than uncompensated victims? Or would they prefer the regime in which they are assured of their pre-accident, baseline welfare, but in which they do not particularly stand to gain if they are not injured?

In answering this question, let us assume that society would not allow people to drive without adopting one or the other of these liability regimes. That is, the "pre-agreement baseline" is the third option I mentioned above, namely that people do not drive at all. Arguably rational agents would prefer not to take a gamble that would potentially give them a lower payoff than they would receive from their ex ante baseline welfare. Indeed, their preference for protecting their baseline might be sufficiently strong that it would overcome even a higher expected utility from the regime involving a gamble. Rational agents, in other words, might be particularly disinclined to gamble with their baseline welfare, and be willing to forego the potential for even greater gains to protect that baseline. Why might this be the case? Consider how an agent contemplating an agreement for a future accident regime might reason. He might evaluate the rationality of the agreement by asking himself whether he would later be glad he entered that agreement or whether he would regret having done so. It is plausible to suppose that he would think himself well-served by the agreement if he is better off under the terms of the agreement than he would be had he never entered into the agreement in the first place. That is, even if an agreement would require him to perform individual actions that leave him worse off, the agreement as a whole should leave him better off than he would be had he never entered into it. Let us call this condition the "benefit principle."

The requirement that an agent must receive a net benefit is a plausible threshold criterion for rational agents to apply in assessing options regarding the basic institutions of their society. The idea is that rational agents would not endorse basic institutions that they expect would leave them worse off than they would be in the absence of that institution. If rational agents did in fact apply such a test, they would not simply be concerned to maximize their utility in selecting basic social institutions. Their interest in maximizing utility in this context would be subject to a constraint, namely that no matter how much expected utility an institution conferred ex ante, each agent should be reasonably confident that he would fare at least as well under that institution as he would in its absence. The benefit principle should not be treated as a general test for the rationality of all agreements, plans or courses of action rational agents might adopt. For as a general condition of rationality, the principle would be much too strong: it would appear to rule out the rationality of homeowner's insurance, gambling (no matter how favorable the odds), and stock market investment. But I am suggesting that such a strong condition would be rational with regard to agreements that govern the basic structure of society. Since rational contractors do not currently know the life choices they will be making, nor the kinds of preferences they are likely to have, they cannot count on ordinary calculations of expected utility to adequately protect their interests.

Would the agreement to establish the no-liability regime satisfy the benefit principle in the non-reciprocal case? Under the no-liability regime, if a person happened not to become injured in an accident, he would regard himself as better off under this regime than he would be had he never agreed to that regime, since he would be able to enrich himself by engaging in risky behavior without paying the "price" for it. But he might turn out to be a victim without ever being an injurer. In such a case the no-liability regime would leave him worse off than he would be if he had never agreed to this regime, since his welfare would then be substantially below his pre-agreement baseline, and he would receive no compensation. Since the rational agent does not know which of these scenarios will occur, he cannot be sure that the agreement to establish the no-liability regime is one he will be glad to have entered. The no-liability regime thus appears to violate the benefit principle.

What about the liability regime? No matter what position he happens to find himself in, an agent will regard himself as better off under the agreement to
establish the liability regime than if he had never entered into that agreement in the first place. For if he is an injurer, he will have to pay for the injuries he inflicts, but he will regard himself as better off for the opportunity to drive negligently, paying the price, than if he were not able to do so. (He presumably values the risky activity more than the discounted amount he must pay in compensation, otherwise he would not engage in the activity.) And if he is a victim, he is restored to his pre-accident baseline welfare, at the same time that he has the opportunity to engage in risky activities as a potential injurer. There is thus some reason to think the rational agent would prefer the liability regime. Indeed, he might prefer it sufficiently that he would be willing to accept it even if his expected utility were higher under the no-liability regime, as Kaplow and Shavell suggest. And if this is so, then we cannot assume that rational agents would systematically select the utility-maximizing solution.

In saying that rational agents would select a legal regime that satisfies the benefit principle, I have not uniquely identified the legal regimes that rational agents would adopt. In theory, there might be many possible legal regimes that satisfy that principle. My claim is only that rational contracting agents would reject any legal regime that failed the benefit test, and so we have reason to think they would not consistently choose the legal regime that maximized social welfare. In a fuller account, one would need to specify some further principle of selection that would allow the parties to choose from among the various eligible regimes. Presumably that principle would be consistent with contractarian assumptions. My suggestion, then, is simply that rational agents would select the benefit principle as a minimum necessary condition for judging the acceptability of any agreement regarding legal regimes into which they might enter.

Let us now apply the framework we developed in considering the accident law regime to a second question of legal-policy, namely the death penalty. Assume that the death penalty has deterrent value over and above life in prison without parole. Would rational agents choose to include it in their punishment regime? If potential criminals are rational actors, the same level of deterrence can be accomplished either by combining a high probability of detection with low penalties, or by combining a lower probability of detection with more severe penalties. The legal economist therefore recommends that penalties be made as severe as possible, in order to minimize the use of costly resources in detecting and preventing crime. In general, the severity of punishment can be increased at less added cost than can the chances of detection. Thus economic analysis seems to point in the direction of using death as a penalty, if, as its proponents claim, the death penalty has greater deterrent value than life in prison without parole.

But would rational agents setting up an institution of punishment select the form of that scheme by applying the principle of utility maximization? There are two reasons to think they might not. First, notice that the legal economist's argument places no upper boundary on the level of acceptable punishment. Suppose, for example, that torturing a person prior to executing him would increase the deterrent efficacy of the death penalty significantly. By the legal economist's lights, we should then opt for torture, since it would enable us to lower the amount spent on detection. It is probable, however, that the parties to an original social contract would treat certain kinds of very harsh penalties as off-limits, no matter what the utilitarian gains. The legal economist, by contrast, famously has difficulty explaining why certain morally impermissible punishments are unacceptable. Second, there is no reason to suppose that the person executed must be guilty of a crime in order for the death penalty to have deterrent efficacy. The legal economist's argument seems to sanction executing innocent agents. It is true that others will not be deterred from committing crimes unless they see at least a significant number of those who have committed those crimes receive a stiff enough sentence that they would rather abandon their plans than risk the associated punishment. But there is no need for the person punished actually to have committed the crime, strictly speaking. All that is really necessary is for the public to believe that he has done so.

Now it might seem that the only way parties to a social contract would be able to eliminate such implications is if they were able to agree directly on a set of deontological principles that would rule out torture and sacrifice of the innocent. But it is unlikely that putative members of society could reach agreement on substantive moral principles in this way. Thus once again I suspect that Kaplow and Shavell are correct to think that rational agents would not choose to govern their institutions according to specific deontological norms. The benefit of seeking to establish the basic structure by agreement among rational agents is to accommodate the fact that individuals have different views of the good and so cannot reach substantive agreement. How, then, would the parties to a social contract determine their principles of punishment?

To determine whether rational agents would opt for the death penalty, let us once more apply the benefit principle, and ask whether a rational agent would regard himself as better off once he is subject to the death penalty, having lived with the benefits of a regime in which the higher deterrent effects of the death penalty are available, than he would have been in its absence. Does he prefer his life with the death penalty, now that he is subject to it, or the life he would have lived without that particular penalty? Presumably he would almost always prefer his life without the death penalty than with it, since he will not regard himself as better off once he is executed than he would have been had he never lived in a regime with the death penalty. That is, the agreement to institute the death penalty would have left him worse off, on balance, than he would have been in the absence of the agreement. The death penalty thus seems to violate the benefit principle and rational agents would therefore reject it.

In this regard, the death penalty is like a kidney lottery society, in which each person has the option of entering into the following agreement: Should he suffer
dual kidney failure and require a kidney to survive, a member of the society will be chosen at random to supply a kidney in order that the victim of kidney failure might live. In such a case, the kidney lottery clearly passes the benefit test, since the person receiving the kidney is better off than he would have been in the absence of the lottery. But suppose he never does suffer kidney failure. Instead, someone else suffers kidney failure and our agent is chosen by lot to supply him with the needed kidney. The kidney lottery agreement would clearly fail the benefit test in that case, since the person chosen to supply the kidney will then wish he had never entered the kidney lottery. In this case, the lottery has turned out to be all cost and no gain. It is true that he might still benefit from the agreement, because should his remaining kidney now fail, he will be the beneficiary of a lottery in which someone else must supply a kidney. But this is not a case of net benefit, since in the end he will be left with one kidney, which is what he would have had if he had never entered into the agreement in the first place.

Arguably we should think of the death penalty as a kidney lottery. An agent who must be put to death by the state is like the person who suffers kidney failure under the lottery agreement: while he has undoubtedly benefited from the agreement to have the death penalty thus far, the agreement has not conferred a net benefit on him, since the fact that he will be put to death will now more than counter-balance any benefits he has received from the agreement. But how do we know this? Perhaps in the absence of the death penalty, everyone would die of violent death at someone else’s hands at a very young age. So being able to live longer, and then be put to death, does not necessarily leave one worse off than one would have been in the absence of the death penalty. But in all likelihood, the availability of other severe punishments, such as life in prison without parole, would provide most of the deterrent benefits rational agents would require. The marginal deterrent benefits death provides over other available punishments would almost certainly be insufficient to outweigh the loss of one’s own life.

One might object that the death penalty is distinctly unlike the kidney lottery, in that the decision to commit a crime is subject to choice, whereas suffering kidney failure is not. But once we allow that the state makes occasional mistakes in the administration of the death penalty, and that indeed it is quite impossible to run a death penalty without occasional loss of innocent life, the death penalty becomes precisely like the kidney lottery, since for an innocent person put to death, the loss of life is as much under his control as kidney failure. Moreover, even if we could ensure that no innocent people were put to death, there is reason to see even one’s later decision to commit a crime as beyond one’s control. This is because, from the ex ante perspective, a rational agent knows nothing of his future motivations. The need or desire to commit a crime may be so compelling that his future self has no choice but to comply. Under either scenario, a rational agent is no more likely to enter a death penalty lottery than he would be to enter the aforementioned kidney lottery.

These examples are intended to suggest that utility maximization is not the only or even the most appropriate theory of value for a legal system designed by and for rational agents. Rational agents have reason to prefer legal institutions organized around the principle of mutual benefit, since each agent can be sure he will benefit in such a scheme. Thus if we take rational actor psychology seriously as a starting-point for legal theory, there is reason to question the claim that legal institutions would be organized around economic principles. It is at least as likely that they would be organized around contractarian principles instead.

4. DOES WEALTH MAXIMIZATION SOLVE THE PROBLEM?

In an article in the Journal of Legal Studies (1979), Richard Posner tries to distance law and economics from utilitarianism for precisely the reason discussed in the preceding section, namely that utilitarianism endorses the sacrifice of individual welfare for the sake of the greater good. He does this by suggesting that law and economics need not be premised on utility maximization after all. Instead, the legal economist suggests that legal institutions should seek to maximize wealth.

Posner argues that wealth maximization better captures our intuitions about appropriate legal institutions. For example, a poor man who steals an expensive necklace he could not afford to buy may increase the total amount of utility in the world, since his wife, to whom he gives the necklace, may enjoy it more than the wealthy person from whom it was stolen. But the thief nevertheless decreases society’s wealth, assuming he could not afford to purchase the necklace, since in an economic sense that means he does not value it. Thus arguably we can make sense of the law of theft in a wealth maximization system, but not in a system that seeks to maximize utility. It is worth revisiting the criticism I made of law and economics in section 3, then, to consider whether rational agents would see their interests as adequately protected by a legal system that sought to maximize wealth rather than utility.

There is some initial reason to think that wealth maximization would be an improvement in this regard. Posner argues that unlike utility maximization, wealth maximization requires voluntary exchange, since wealth cannot exist without markets. And voluntary exchange requires consent. As a result, law and economics will arguably no longer sanction the sacrifice of a smaller number of individuals for the sake of improving the well-being of others. For if society wishes
to make use of the body or the possessions of one of its members, it will have to purchase the right to do so from him. As Posner puts the point:

The great difference between utilitarian and economic morality ... is that the utilitarian, despite his professed concern with social welfare, must logically ascribe value to all sorts of asocial behavior, such as envy and sadism, because these are common sources of personal satisfaction and hence of utility. In contrast, lawfully obtained wealth is created only by doing things for other people—offering them advantageous trades. The individual may be completely selfish but he cannot, in a well-regulated market economy, promote his self-interest without benefiting others as well as himself. (1979, 132)

The thief, in other words, will have to purchase the necklace from its rightful owner rather than simply taking it, since a nonconsensual transfer would bypass a voluntary market.

If Posner is correct, arguably a system predicated on wealth maximization would not permit institutions that left individuals worse off than they would be under their baseline welfare, since, as I have claimed, no rational agent would consent to such an institution. Thus it might be possible to acquire the benefits of a contractualist system within the framework of an approach that seeks to maximize social welfare, simply by measuring social welfare in terms of wealth instead of utility.

One difficulty with this suggestion, however, is that economic analysis conceived in terms of wealth maximization is incomplete. The reason is that the rights respected in a market system are allocated prior to the operation of the market. That is, economic theory cannot itself specify the initial distribution of rights, since the theory actually presupposes such a distribution. In the case of the necklace, for example, we were assuming that the old owner of the necklace has rights to it that protected his entitlement when we said that the thief must purchase the necklace rather than just take it. But that system of rights has not itself been specified in economic terms; we have no apparent justification in terms of wealth maximization for allowing the old owner to reject attempts to remove the necklace from his possession without his consent. As Ronald Dworkin has convincingly argued, if the thief were willing to pay more for the necklace than the owner would accept to sell it, wealth would be maximized if a dictator were simply to give the thief the necklace, without requiring him to pay for it (1980, 196–97).

Posner argues against this point by suggesting that economic analysis is not indifferent to initial entitlements. There are two reasons for this. First, Posner argues that in general it makes sense to give each person the right to his or her own labor, since a person is likely to be more productive if he is his own master. In general, he thinks, there are economic reasons that bodies and labor should be assigned to their “natural” owners, namely that natural owners are likely to value them most. Second, if there are transactions costs, an initial assignment of right to someone who values a good less than another person may be inefficient, since the parties may not be able to bargain for the exchange of the good. Thus Posner writes, “If transaction costs are positive, the wealth-maximization principle requires the initial vesting of rights in those who are likely to value them the most. This is the economic reason for giving a worker the right to sell his labor and a woman the right to determine her sexual partners. If assigned randomly to strangers these rights would generally (not invariably) be repurchased by the worker and woman respectively” (1979, 125).

The argument, however, is not a good one. Why assume that natural owners are likely to value their own labor or bodies more than others? Posner gives no defense of this claim, and on the face of it, it sounds more like an intuitive or moral idea than an economic one. The factory owner may indeed value the worker’s labor more highly than the worker himself, in strictly economic terms; but he has the initial capital to make a large amount of money from the worker’s labor, much more than the worker could ever be paid. Even a law firm will bill an associate’s time at twice what they will actually pay the associate. The same might be said of a woman and her body: a pimp might make more from selling her body than she could ever make from selling her own. By substituting wealth maximization for utility maximization, Posner thus seems to save law and economics from the morally unacceptable results that come from its traditional association with act-utilitarianism. But it is not at all clear that he is able to do this without assuming a system of rights to property and to natural bodily entitlements. And if that is so, then law and economics can provide only a supplementary theory of how goods allocated according to non-economic principles can be fairly exchanged once initial entitlements are already in place.

There is a second problem with the idea of wealth maximization, which Dworkin has also helpfully pointed out (1980, 194–95). In order to be valuable, wealth must either be something of intrinsic worth—that is, an item worth having for its own sake—or it must be instrumentally valuable relative to something else that has value. Utilitarians think of utility (or happiness) as something worth having for its own sake. It is not valuable because it produces other things of value. Rather, all items that have value have it because they contribute to the total amount of utility. In this sense, the value of utility must be understood as self-justifying, since utility is the value in terms of which all other items of value are justified. If Posner is proposing that law and economics seek to maximize wealth instead of utility, he might be thinking of wealth in this same way: It is the ultimate value in terms of which all other items are valued. But this account cannot be defended, and Posner himself has admitted as much. In response to Dworkin’s critique, he writes that wealth is a value because “it is conducive
to happiness, freedom, self-expression, and other uncontroversial goods” (1980, 244). As Posner recognizes, if wealth is a value, it must be an instrumental value—something that is valuable because it allows one to acquire other items of value.

But if wealth is only instrumentally valuable, it is not at all clear why legal economists would want to maximize it. Is there any reason to think it would be better to maximize wealth than to try to maximize whatever is of intrinsic value directly? For example, assume in a given instance that we have reason to increase social wealth because we believe it will increase total societal happiness and not decrease any of the other goods we value. Is there any reason in such a case to seek to maximize wealth rather than happiness? Presumably not. And if not, why would we imagine that the benefits of wealth maximization change when maximizing wealth would maximize several other values as well? Moreover, there is a disadvantage to maximizing wealth as a proxy for maximizing other values: we do not know what quantity of these other values we will achieve if we just aim at wealth, whereas we can strike the balance we prefer if we aim at each value separately. Posner also suggests that by aiming at wealth rather than utility we can “thereby avoid[] certain well-known problems to which utilitarianism gives rise” (1980, 245), presumably the moral problems we discussed earlier. But if we cannot aim at maximizing utility because it produces morally unacceptable results, then surely it does not improve matters to aim at wealth. For if maximizing wealth really does maximize utility, it will produce those same results, and if it fails to produce them, it can only be because we are not, in fact, maximizing utility.

Not only does turning to wealth instead of utility fail to solve the second of the two problems with law and economics we considered above, but it makes particularly clear just how serious the first difficulty is, namely the problem concerning the motivations of legal actors. Is there any reason to suppose that an individual legal actor who is primarily concerned to maximize his own wealth would prefer rules that maximize social wealth? Not especially. It is possible that a rational legislator or judge would believe that his own personal wealth would be maximized if society were maximally wealthy, but the connection between the personal wealth of legal actors and society’s wealth is not a particularly reliable one. A legal system that allowed legal actors to take bribes and kickbacks, for example, would probably maximize the personal wealth of legal actors, but it would be unlikely to maximize social wealth. Thus the gap between rational actor psychology and social welfare seems particularly significant if we substitute wealth for utility in the maximizing conception of value.

5. Conclusion

I have argued that if human beings are rational, self-interested agents, they are most likely to favor institutions premised on the principle of mutual benefit, rather than on the maximization of social welfare. In this sense, rational actor psychology seems at least as likely to lead to contractarian as to utilitarian legal institutions. What it means for a legal institution to be based on contractarian principles will differ depending on the nature of the institution. But unlike a utilitarian approach, I have argued, the contractarian approach to legal institutions precludes legal rules that leave some worse off than they would be without those institutions. It also precludes the creation of institutions premised on the substantive moral commitments of any particular group that desires to force compliance with their own views on others whose views diverge from theirs. What characterizes legal institutions premised on rational agreement is that the individuals that are subject to them can feel that those institutions have been established according to principles they accept as advantageous, even when the specific workings of those institutions are not.

NOTES

My thanks to Michael Abramowicz, Michael Green, Peter Huang, and Leo Katz for comments on earlier drafts and to Geoff Sayre-Mccord for helpful conversations.

1. Although the term “welfare” is often taken to be broader than the notion of utility, I shall use the two terms interchangeably.

2. I do not think that this assumption in any way affects my disagreement with the legal economist. My criticisms would remain even on the broader conception of utility he proposes.

3. This psychological claim is not to be confused with what I referred to above as the legal economist’s “descriptive” claim about the nature of legal rules. While it is a descriptive claim, it is a generic one about human motivation, one that is not particular to the legal economist.

4. Individual maximizers will produce a pareto optimal state of affairs. But that is not necessarily a state in which social welfare is maximized, assuming that each person’s utility counts equally.

5. For a discussion of this point, in connection with Hart’s account of the “internal point of view,” see Siegel 1999, 1581.

6. I have here adapted a criterion of David Gauthier’s (1990, 1994) for determining when entering into an agreement would be rational for an agent. But I make a quite different use of this criterion than Gauthier does, since I am applying it to test the rationality of enforceable agreements. There is arguably no need for such a test in this
context, because the agent can execute a plan he finds attractive on expected benefit grounds, without having to rely on his own willingness to perform a suboptimal action by way of compliance. Gauthier’s test is a way for an agent to rationalize compliance without an agreement or plan that is not strictly speaking in his interest at the moment he must do his part. But arguably the criterion is a useful one for enforceable agreements as well. That is, we might think that the rational enforceable agreement is just the agreement the parties would have entered into in the absence of an enforcement mechanism. Enforcement, on this view, would be no more than a way of bringing somewhat irrational agents into line with the solution they would adopt if they were fully rational. I cannot, however, defend this account in any greater detail on this occasion.

7. Someone might argue that an individual utility function could be constructed whose maximization would satisfy the benefit principle, and that I have not therefore challenged the rationality of straightforward utility maximization at all. But the function being maximized would probably lack at least one of the features that economists typically impose on rational agency, namely transitivity, completeness, and continuity.

8. There are, however, ways to make the benefit more plausible as a general principle of rationality. As I have suggested elsewhere (Finkelstein 2003), it is plausible to think of an ex ante chance of benefit as itself a benefit. And if this is so, then insurance and some gambles would be worthwhile: it would be rational to sign up for insurance, for example, if the ex ante chance of benefit were itself greater than the cost of the premium. Thus if the damage against which one is insuring is sufficiently great, and the chance that the damage will occur sufficiently high, then paying an insurance premium for an insurance policy that one does not end up needing nevertheless conveys a benefit.


10. See supra section 3.

11. See supra section 2.