A Positivist Account of the Rule of Law

Frank Lovett

Despite extensive debate, accounts of the Rule of Law remain strikingly vague and imprecise. This paper begins the task of remedying this situation through the development of an analytically rigorous theory of the Rule of Law. First, it offers a preliminary sketch of a concept of law; second, it shows how the principles traditionally associated with the Rule of Law can be intrinsically connected with this concept of law; and third, it shows how these principles place meaningful restrictions on what states can and cannot do. Further components of the overall project are gestured toward in the paper's conclusion. Apart from advancing the broader goal of developing a complete theory of the Rule of Law, a paper of even this limited scope will be helpful in imposing some much-needed conceptual discipline over the Rule of Law debate.

The Rule of Law idea has had more than its share of enthusiasts (and critics).[1] Despite extensive debate, however, descriptive accounts of the Rule of Law remain strikingly vague and imprecise. Being myself an enthusiast—though on somewhat unconventional grounds—I find this lack of analytic rigor particularly distressing. Accordingly, in this paper I would like to begin

Frank Lovett is a joint J.D./Ph.D. candidate in law and political science at Columbia University. He would like to thank Jeremy Waldron for very helpful comments and advice on an early draft of this paper and Paul MacDonald for many discussions concerning the paper's themes.

1. I capitalize *the Rule of Law* to distinguish this idea from *a rule of law* (i.e., a statement regarding what the law with regard to some question happens to be in a particular legal system). For example, "contracts to perform illegal activities are not enforceable" is a rule of law in the American legal system. Accounts of the Rule of Law are too extensive to cite comprehensively, but some of the better known include the following: Dicey [1915] 1982, chap. 4; Neumann 1937; Hayek 1944, chap. 6, and 1960, chap. 10; Fuller 1964; Rawls 1971, sec. 10, 38; Raz 1979, chap. 11; and Finnis 1980, chap. 10. For reviews of this literature, see Radin 1989; Waldron 1990, chap. 3; Scheuerman 1994; or Fallon 1997. Oft-mentioned classical sources for the Rule of Law idea include Aristotle, Locke, and Montesquieu. See Shklar 1998 for an overview.

© 2002 American Bar Foundation.
0897-6546/02/2701-41$01.00

the task of remedying this situation by developing an analytically rigorous theory of the Rule of Law.

Suppose one wants to argue, as I do, that the Rule of Law is a good thing. It is worth considering at the outset how such an argument might fail to be interesting or useful. One way it might fail is if it turned out to be superfluous. Suppose, as some have argued, we regard the principles traditionally associated with the Rule of Law (generality, publicity, prospectivity, and so forth) as independent normative standards or ideals legal systems may or may not conform to in varying degrees. These Rule of Law principles might be considered a subset of all the virtues applicable to legal systems; or else the Rule of Law might simply be defined as the complete set of such virtues. But in either case, the point (according to this view) is that understanding what a legal system is, and determining whether it conforms to Rule of Law principles are two separate and independent questions.[2] Now if this view were correct, then our account would not be very successful, for *the Rule of Law* would then serve as little more than shorthand for the various ways we might regard legal systems as good. Arguing that the Rule of Law is a good thing would then amount to little more than arguing good laws are a good thing, a decidedly uninteresting and not very useful claim.

Therefore, a successful account of the Rule of Law idea ought to reveal some *intrinsic connection* between legal systems on the one hand and Rule of Law principles on the other; in other words, it ought to show that for a political community to have something recognizable as a legal system already entails some degree of conformity to the Rule of Law. This, by contrast, *would* be an interesting and useful claim. But even if our account succeeded in showing an intrinsic connection between legal systems and Rule of Law principles, it might fail in another way. Namely, it would fail if the Rule of Law placed no actual limits on what states can do. "If the state is comprehended as a legal order," suggests Hans Kelsen, "then every state is a state governed by law (*Rechsstaat*), and this term becomes a pleonasm" (1989, 313). A successful account must therefore show how the Rule of Law places meaningful restrictions on what states can do: in other words, Rule of Law principles must somehow express the limits of what sorts of state activities count as legal in nature.

By now it should be clear that developing a complete theory of the Rule of Law would be quite an extensive undertaking. First, we must develop a concept of law as such. Second, we must offer an account of how the principles traditionally associated with the Rule of Law idea are somehow intrinsically connected with this concept of law. Third, we must show how these principles place meaningful restrictions on what states can and cannot do. Once we have completed these tasks, it still remains on the one hand to

2. Raz (1979, chap. 11) and Waldron (1990, chap. 3) more or less endorse the first version of this view; Finnis (1980, chap. 10) advances something like the second.

produce a normative argument for why states *should* conform to Rule of Law restrictions—and on the other hand, to suggest how political and social institutions might be designed so to ensure that states *actually do* conform as we would like them to. Obviously, all this is far more than can be accomplished in one paper. Therefore, I focus on the second and third tasks here. A preliminary sketch of a concept of law is offered below, but it is developed only to the point where it becomes feasible to work out the possibility of an intrinsic connection between legal systems and Rule of Law principles. The normative and institutional components of the project are only gestured toward in the conclusion.

Apart from advancing this broader theoretical project, I hope a paper of even this limited scope will impose some much-needed conceptual discipline over the Rule of Law debate. One potentially interesting payoff from this (discussed in part three) will be a division of the principles traditionally associated with the Rule of Law idea into two groups: those that can be intrinsically connected with the concept of law as such, and those that cannot. In my view, the latter should be disaggregated from the Rule of Law idea and viewed as a distinct set of virtues applicable to legal systems.[3] This, I believe, will greatly clarify further debate.

I

In this and the following part of the essay, I present an account of legal systems, which will then serve as the basis for a discussion of the Rule of Law in part three. As stated in the title of the paper, this will be a positivist theory of law, in that it conforms to two broad commitments shared by all legal positivists: (1) the "social thesis" that determining what counts as a legal system must be a matter of social fact and (2) the "separability thesis" that law and morality are not necessarily connected (Coleman and Leiter 1996, 241 ff.). Obviously, my argument for an intrinsic connection between the Rule of Law idea and the concept of law hinges on the reader's willingness to accept at least provisionally my particular account of legal systems. This account is similar to other well-known positivist theories of law, though it is presented in a somewhat less familiar vocabulary. Nevertheless, legal positivism has its competitors, and I make no attempt to address such debates here.

Let me stress again what I said in the introduction, that my account of the concept of law is only a preliminary sketch, developed no further than is necessary to work out the possibility of an intrinsic connection between legal systems on the one hand, and Rule of Law principles on the other. The

3. Fuller (1964) also wanted to show an intrinsic connection between the concept of law and Rule of Law principles, but he failed to notice this important division.

discussion is divided into two parts: this part concentrates on the idea of a social convention, the next on the idea of a legal system.

The Idea of Social Conventions

Let us imagine some real or hypothetical political community.[4] In any such community, the general behavior of community members will be guided by a large array of what might be called *social conventions*. A social convention is a publicly known and regularly followed rule for action, sustained ultimately by community members' mutual expectations regarding each others' behavior. For example, in many communities a social convention exists of standing in line and waiting in turn for service at banks, checkout counters, post offices, and so forth. This social convention is supported by the mutual expectation of being disapproved of when one violates the established rule and approved of when one conforms to it. The idea of a social convention is similar in certain ways to what H. L. A. Hart calls a social rule (1994, 51–61), Kelsen a norm (1989, 3–23), and Frederick Schauer a prescriptive rule (1991, 1–6), though not without certain differences, as we shall see.

FIGURE 1

		Wife: Opera	Wife: Boxing
Husband	Opera	2, 1	0, 0
	Boxing	0, 0	1, 2

As I will use the term, a social convention is equivalent to what game theorists mean by a Nash equilibrium, and a rule for action is equivalent to what they mean by a strategy.[5] Consider the well-known game called "battle of the sexes" (see figure 1). In this game, a husband and wife must decide whether to go to the opera or a boxing match. The husband prefers the opera to boxing matches, and the wife boxing matches to the opera, but both prefer spending the evening together, at either event, rather than apart. Suppose the social convention of going to the opera already exists: in

4. By *political community* I mean only a community that is, or potentially could be, fully self-sufficient. I assume the political community is closed (i.e., that we can ignore anything that goes on outside the community). A fully developed theory of law, of course, would have to take into account the role of, for example, international relations and international legal regimes.

5. For the association of social conventions with Nash equilibria I am indebted to Calvert 1995. In his recent book, Posner (2000, chaps. 2–3) offers a somewhat similar account of social conventions.

other words, suppose the rule for action or strategy "always go to the opera" is known to both the husband and wife, and is regularly followed by both. Neither the husband nor the wife has an incentive to violate the established rule by going to the boxing match, because both operate under the expectation that the other will be going to the opera. When no one has an incentive to unilaterally change his or her strategy in this way, we have a Nash equilibrium. A social convention, as I define it here, is simply a Nash equilibrium writ large.

All actual social conventions are Nash equilibria, but not all Nash equilibria are actual social conventions.[6] In the game above, always going to the opera is only one of several possible Nash equilibria: always going to the boxing match is another, as is a "mixed strategy" equilibrium in which the husband goes to the opera with a probability of $1/3$ and the boxing match with a probability of $2/3$, while the wife does the opposite. Further equilibria are possible if we introduce publicly observed events. For example, going to the opera on even-numbered calendar dates and a boxing match on odd-numbered calendar dates could be a Nash equilibrium. The *actual* social convention is whatever equilibrium the husband and wife happen to have coordinated on. The existence of some social convention in a particular community I take to be a matter of descriptive fact, at least in principle amenable to empirical verification.

To repeat, a Nash equilibrium is a situation in which no player has an incentive to unilaterally change his or her strategy, given the expected strategies of the other players. Let me stress that the incentives at work here must be construed very broadly. For example, if we are interested in the social convention of waiting in line for service, the relevant incentives might include (a) the material costs and benefits of waiting in line or not, (b) the internal psychological costs and benefits of following the rule to wait in line or not, and (c) the social costs and benefits of being approved of or disapproved of by others depending on whether one waits in line or not. If the social convention of waiting in line for service actually exists, then we must presume all the benefits minus all the costs of complying with the rule, given what everyone else is expected to do, outweigh all the benefits minus all the costs of not complying with it.

Many people hear talk of incentives and assume one must be referring to narrowly economic or otherwise purely self-regarding costs and benefits, but this is by no means necessarily the case.[7] For example, many people wait in line for service not because they fear social disapproval, but rather because they would feel guilty if they did not, or even because they genuinely

6. All Nash equilibria might be *possible* social conventions, except for the fact that bounded rationality problems make some extremely complex Nash equilibria practically unlikely.

7. Elster (1989a, 99–100, 125–40, and 1989b, 119–23) persuasively argues against the reduction of social conventions to narrowly economic or self-regarding incentives.

believe it is the right and fair thing to do. These sorts of incentives are not narrowly economic or otherwise purely self-regarding, but considering social conventions from a purely descriptive point of view for the moment, it should be clear that they too can be considered incentives in a broad sense. Feelings of guilt are part of the internalized reward and punishment structure human beings acquire through socialization. Similarly, the desire to do the right thing, from a strictly descriptive point of view, is an example of a goal-directed preference, like the desire or preference to avoid disapproval, or the desire or preference to avoid feeling guilty. If it is a descriptive social fact that some social convention exists, then it must also be a descriptive social fact that incentive structures are in place to support it.[8] It is an entirely separate question whether the social convention in place is good or just, or whether some other possible social convention might be normatively superior.

Having uncoupled the description of social conventions from their moral standing, we can easily see that people might conform to social conventions they disapprove of. In the battle-of-the-sexes game above, the social convention is certainly not a matter of normative indifference to the husband or wife. Although the wife has a good reason to go to the opera, this reason is certainly not that she believes the observed rule is fair or just in any normative sense. Indeed, she might well conform to the rule while vocally criticizing it as unfair. The existence of a social convention is thus quite compatible with widespread—even universal—public criticism.[9]

Four Objections

The equation of social conventions with Nash equilibria suggests, among other things, a commitment to what is sometimes called the practice theory of social conventions—that is, the theory that social conventions can be exhaustively explained with reference to descriptive social practices. Since the practice theory is subject to a number of well-known criticisms, it

8. This does not mean people always, in each individual instance, comply with social conventions because they have made the necessary incentive calculations. Often people act as if on "autopilot." This will be discussed further below.

9. A social convention might be universally criticized if it is a Pareto sub-optimal Nash equilibrium. In this situation, there are several possible equilibria, but the players happen to be coordinated on one of the inferior ones, worse for everyone than some alternative. They are stuck, however, because no one player can switch strategies unilaterally without making herself worse off: all the players have to switch together. The coordination problem involved in switching together may obstruct reform. Many have remarked on the persistence of undesirable social conventions: for example, see Elster 1989a, chap. 3, and 1989b, chap. 12; Hart 1994, 257.

may be appropriate to consider some of these in detail before turning to consider legal systems in part two.[10]

FIGURE 2

		Player 2	
		Action A	Action B
Player 1	Action A	1, 1	1, 0
	Action B	0, 1	0, 0

1. One objection to associating social conventions with descriptive social practices is that not all observed patterns of behavior correspond to a rule for action. For example, we might observe patterns of cuisine or dress, and yet these are not necessarily the product of rules for action sustained by mutual expectations. Rather, they might simply result from the fact that members of a given community happen to have similar tastes.[11] In response to this problem, let us distinguish between social conventions and social habits. Consider the game shown in figure 2. In this game, both players will take action A, and so strictly speaking the top left-hand box represents a Nash equilibrium. However, the incentive structure of each player is completely independent of the other player's actions, and so one would not ordinarily describe the observed outcome as being the product of "following a rule." Rather, the pattern is merely a byproduct of the preferences the players happen to have—in this case, they both happen to prefer doing A to doing B. Game theorists call this a decision-theoretic situation, as contrasted with a strategic situation: in a decision-theoretic situation, each player can maximize her own payoff without concerning herself with the strategies other players might or might not adopt. Thus the "battle of the sexes" game is a strategic, and not a decision-theoretic situation, because what the husband wants to do depends on what the wife is doing, and vice versa.

Patterns of behavior that arise from similarities of individual preference alone I will refer to as *social habits*. Only patterns sustained at least in part by mutual expectations regarding the behavior of others count as genuine

10. One criticism I do not address is that the practice theory cannot explain unobserved rules—as, for example, moral rules people accept as valid but do not comply with (see Dworkin 1977, 52–53; Raz 1999, 53–55). This criticism is not serious in my view, because for the limited objectives of this essay, we need only a theory of social rules (i.e., social conventions), and not a theory of all sorts of rules.

11. For discussions of this problem, see Hart 1994, 51–61; Raz 1999, 55–56. I say "not necessarily" because in fact these examples may turn out to rest on social expectations after all: for example, many dress codes might be supported not primarily by individual tastes, but rather by the fear of disapproval or the desire for approval. Indeed, I am not convinced cases of pure social habit exist.

social conventions.[12] One way of determining whether a particular pattern of behavior arises from a social habit or a social convention is to see whether nonconforming individuals are subject to social criticism: Generally speaking, people will not take a critical attitude towards those who fail to conform to social habits. (If they did, then presumably some individuals at least would start to conform not because they want to intrinsically, but because they prefer avoiding social criticism; at this point, the pattern ceases to be a social habit, and becomes a social convention after all.) The distinction appears in ordinary conversation as well. Suppose we ask people their reasons for acting in a particular way. If their answer takes the form, "it is the rule to ϕ," or "because one ought to ϕ," we may assume the pattern of behavior rests on social convention. If their answer takes the form, "because I prefer to ϕ," then we may assume the pattern of behavior rests on social habit. Of course, these tests are not perfect, but they capture the main point well enough, and in any case distinguishing social habits from social conventions will turn out to be of minor importance for the concept of law developed below.

Although social conventions are associated with patterns of behavior, one must be careful not to think that a social convention *is* a pattern of behavior. The observed pattern of behavior is only the by-product of a social convention (i.e., it is a by-product of the social fact that people are following a particular rule for action or strategy). In the battle-of-the-sexes example above, the observed pattern of both husband and wife going to the opera is a by-product of the fact that both are following the rule for action "always go to the opera." In the social convention of waiting in line for service, the observed pattern of everyone waiting in line is a by-product of the fact that everyone is following the rule for action "wait in line, approve of others who wait in line, and disapprove of others who do not wait in line." The content of the social convention is the rule for action itself, aspects of which may often appear only as counterfactuals and thus not as part of the usually observed pattern.[13] To put it in game theory language, a Nash equilibrium is an equilibrium of *strategies*, not of *outcomes*. Appendix B describes a game in which this difference is quite clear.

2. A second problem with the practice theory, related to the first but more subtle, is that it may be impossible to say *which* particular rule for action is being followed simply by looking at the observing behavior. This

12. Hart (1994, 51–61) draws a similar distinction between social rules and social habits. In his view, however, social rules are set apart from social habits by the fact that those who conform to the former have a critical reflective attitude toward the convergent behavior in question, which he calls the "internal point of view." My theory does not depend on this notion. The distinction also appears in Schauer 1991, p. 1–3, there between descriptive rules and prescriptive rules. Although the set of descriptive rules is more encompassing than the set of social habits (it includes nonsocial descriptive regularities—for example, "as a rule the Alps are snow-covered in May"), our respective lines of division are otherwise equivalent.

13. Postema (1982, 176–77) correctly notes this point.

objection is often associated with Wittgenstein (1958, esp. §§ 143–242; see also, Postema 1982, 188–89; Radin 1989, 797–810; Schauer 1991, 64–68). The difficulty he noticed is that any observed rule-guided behavior sustained up to time *t* might always result from several different rules for action. For example, suppose the husband in our earlier example is following the rule "always go to the opera," whereas the wife is following the rule "go to the opera 100 times, then go to the boxing match 100 times, and so on." Unless we have observed 100 or more cases, we cannot tell which rule the players are following; indeed, the players themselves might believe they are both following the same rule, when in fact they are mistaken.

As I suggested in the case of social habits, a large part of this problem can be solved as a practical matter simply by asking people what they are doing. When there is a genuine social convention, everyone should offer the same rule as the reason for their action. But this is not a perfect solution, for the players may have different interpretations of what the rule statement means without knowing it. Suppose the opera house one night is closed for repairs: In this situation, the husband might believe the rule implies going to the boxing match instead, whereas the wife believes the rule implies staying at home. So long as this particular question of interpretation does not arise, the fact that they interpret the rule differently in some cases might escape notice.

This presents some challenging philosophical problems, but not very serious practical ones, because in the vast majority of cases, difficulties of this sort will not arise. This must be so because in fact people do seem to observe many social conventions without much difficulty or fanfare, and this would not be possible unless problems of this sort were relatively minor. In many cases, cultural similarities ensure that even when facing novel situations, most people in a given political community will agree the rule should be extended in one way rather than another.

When it comes to the legal systems specifically, this problem flows from gaps in the law, and a case at law analogous to the situation above, where the observed rule is underspecified, is a "hard case" in the jurisprudential parlance. Clearly, hard cases exist in all legal systems; Ronald Dworkin argues that because convention-based accounts of legal systems fail to describe how judges respond to hard cases, such accounts are undermined (1977, chaps. 2–3; 1986, chap. 4). While I agree with Dworkin that any complete account of the concept of law has to deal in some way with the problem of hard cases, I do not believe a convention-based account like the one sketched here would have much difficulty in doing so. A complete and satisfactory response would take us far afield, however, so I must set the

problem aside.[14] Fortunately, most cases are not hard cases, nor could they be in a legal system that functions well at all.

3. A third objection to the practice theory is that no account of social conventions can fully explain their "normative character" with reference only to descriptive social practices. In order to assess this objection, we must first understand a distinction sometimes made between two sorts of social convention.[15]

Sometimes people follow rules merely because other people follow them; standard examples are the rules of grammar, the rules of chess, or the rule of driving on one side of the road. There is no substantive reason for driving on one particular side of the road rather than the other, but it is useful for everyone to be doing the same thing. Thus the fact that people drive on the right in the United States is a good reason for driving on the right oneself, but there is no additional reason beyond this for driving on the right per se. A social convention of this sort is variously called a consensus of convention, a conventional rule, or a convention equilibrium.

For the most part, it is easy to explain this first sort of social convention with reference to descriptive social practices alone. But other times people follow rules for different reasons. For example, many people do not follow a rule of waiting in line for service, or a rule prohibiting theft merely because others do. Rather, they follow the rule because they believe it is the right thing to do, regardless of what others are doing. In other words, people view *the rule itself*, and not merely the behavior of others, as a reason for acting in some particular way. A social convention of this second sort is sometimes called a consensus of conviction or a social norm. This latter sort of social convention, it is argued, cannot be explained with reference to descriptive social practices alone, and therefore the practice theory falls short.

It is important to be clear about where the supposed difficulty lies, for saying the practice theory cannot fully explain the "normative character" of many social conventions can be misleading. Obviously, social conventions may have moral properties: Some social conventions correspond to the requirements of justice, as for example social conventions against lying or stealing; others correspond to morally permissible but not obligatory principles, as for example the ancient Greek social convention of always offering

14. Kelsen's response to the problem of gaps seems to me wholly inadequate (1989, 245–50). My view is closer to Hart's (1994, 124–36): roughly, that most well-developed legal systems have social conventions according to which people coordinate on the interpretations issued by designated authoritative interpreters (e.g., judges). As we shall see, legal systems may include many social conventions that are not themselves laws, and so there may be widespread gaps in the laws without there being very many gaps in the legal system considered broadly.

15. The discussion in this section follows Dworkin 1977, 50–51, 53–58, and 1986, 135–39; Raz 1999, 56–58; Elster 1989b, chap. 11; and Hart 1994, 256–59. I must confess that I agree with Hart in finding this objection "tantalizingly obscure" (1994, 257).

strangers food and shelter in one's home; and still others are quite unjust, as for example social conventions of discrimination. The practice theory of course does not explain moral properties of this sort, because it is not a theory of moral philosophy. For a descriptive account of social conventions—which is all the practice theory purports to offer—the justice or injustice of that social convention is neither here nor there. Thus, the objection to the practice theory cannot be that it fails to explain the normative character of social conventions in this sense.

Nor can the objection be that people often comply with social conventions because they believe they ought to—that is, because they each individually believe following the rule is the morally right thing to do, regardless of what others are doing. This can easily be incorporated into the practice theory as an instance of social habit, wherein a person's desires or preferences are such that regardless of what others are doing, she will prefer following the rule to not in nearly every situation. The practice theory is strictly neutral toward the content of people's beliefs.

The real objection must concern the nature of what people who comply with certain social conventions are *actually doing* when following the rule for action in question. For example, when following the rule of waiting in line for service, or the rule prohibiting theft, people may not actually undertake anything like a cost-benefit analysis in light of their beliefs, preferences, and relevant incentives (external or internal). Rather, they follow a rule blindly—on "autopilot," as it were. The problem, then, is that the practice theory inaccurately describes what people are doing much of the time.

As a matter of empirical fact, I do not doubt that people often behave in this manner, following rules for action blindly without considering reasons for or against doing so (or, as it is often put, taking the rule itself as a sufficient reason for following it). This should hardly be surprising, for human beings are creatures of habit. In the course of ordinary life, we face far too many decisions to consider in detail what the best course of action would be, all things considered, in every case. Habit often takes the place of repetitive decision-making calculations, operating as a much-needed time-saving device. But suppose it were no longer the case that waiting in line for service, for example, were the best thing to do, all things (including informal social criticism) considered. For a short time, habit might maintain the usual pattern of behavior; but as people reflect on their decisions, once in a while they will abandon the previous rule for action and adopt a better one. Before long, a new pattern of behavior will emerge, corresponding to some new social convention. Therefore, while it is true that patterns of behavior corresponding to social conventions might often be proximately caused by blind habit, a social convention must nevertheless be sustained *ultimately* by community members' mutual expectations regarding each others' behavior.

4. A final objection to the practice theory, which I consider only briefly, relates to the mistaken notion that the Nash equilibrium concept cannot handle the fact that real-world social conventions are generally imperfectly complied with. Quite the contrary, it is easy—though not very profitable—to incorporate deviance (and not merely counterfactual deviance) into a Nash equilibrium model. In appendix B.2, I present a model of this sort to show it can be done. Often, however, the most useful model is not the most descriptively accurate, and this is probably one of those cases.

II

The idea of a social convention should be clear enough by now. The next question is What do we mean by law? Like Kelsen (1989), I find it convenient to divide my analysis into two stages: the first considers legal systems as if they were unchanging or "static," the second removes this assumption and incorporates their "dynamic" aspect.

The Static Aspect of a Legal System

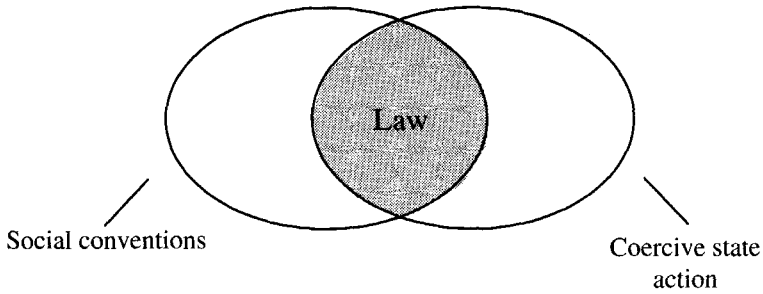
No doubt many social conventions operate solely of the basis of everyday social pressures and individual preferences, as in the example of waiting in line for service mentioned earlier; this will especially be the case in political communities that are small, close-knit, and relatively undeveloped. In all reasonably complex political communities, however, at least some social conventions are backed by the coercive powers of the state or state-like institutions, generally in addition to the incentives provided by social pressures and individual preferences. In other words, with regard to some social conventions, people expect that breaches or observances of the rule for action will be punished or rewarded not only by social disapproval, but also by state enforcement. This is roughly what I mean by a law, or at least the normal form of law: a social convention backed by the coercive powers of the state.[16]

Let us define sanctions to include both the punishments (negative sanctions) and the rewards (positive sanctions) attached to a particular rule for action. Informal sanctions are those produced by social pressure; formal sanctions are those produced by the coercive powers of the state.[17] A law is

16. In this paper I will not further analyze the concept of the state; obviously, a fully developed theory of law would have to do this. For our purposes, we can assume something like the well-known definition given in Weber 1919, with the caveat that political communities may have legal systems backed by other sorts of institutional public authorities as well.

17. All social conventions have sanctions as here defined, but they might only be informal. A pattern of behavior supported *only* by internal sanctions, without either informal or formal sanctions in addition, is a social habit, not a social convention, as per our earlier

FIGURE 3



simply a social convention enforced by formal sanctions (in addition to informal sanctions, or not). The basic idea can be seen in figure 3. Note from this figure that, *pace* Kelsen (1989, p. 286–319), not all coercive state action counts as law—or, to put it another way, the connection between coercive state action and law is contingent (Cf. Raz 1999, 137–41; Schauer 1991, 10–12, 167–74; Hart 1994, 141–47). Obviously, I will have more to say about this later, in part three. In addition, I will argue shortly that the legal system understood broadly must include social conventions not themselves laws in the narrow sense.

The exact relationship between formal state sanctions and the underlying social convention will naturally vary from case to case. Consider a number of examples.[18]

First, there will be cases in which, regardless of formal state sanctions, a firm social convention or social habit would exist, either because only one Nash equilibrium is realistically possible or because in the set of possible Nash equilibria, one is clearly better for everyone. For example, it is hard to imagine any political community that did not place at least some prohibitions or restrictions on the use of violence against community members. Social conventions or social habits to this effect will certainly exist, with or without formal state sanctions; to some extent, then, one might regard laws supporting them as superfluous. So what is their point? First, the addition of formal state sanctions may improve the breadth and depth of compliance: breadth, by deterring those few individuals insufficiently motivated by internal and informal sanctions alone, and depth by reducing temptations to violate the rule under extraordinary circumstances. Second, laws may serve additional purposes, not directly related to enforcement. These might include, for example, an educative purpose: teaching community members in

discussion. Since many people have *formal* sanctions in mind, it might seem that sanctionless social conventions exist. Cf. Kelsen 1989, 27–28, 50–54; Raz 1999, 155–62; Finnis 1980, 325–37.

18. Postema (1982, 183–86) offers a list similar in some respects to the one that follows. By way of contrast, see Posner 2000, chaps. 4–9.

a more explicit and public manner what is expected of them (Finnis 1980, 262–63).

Second, there will be cases where several Nash equilibria are more or less equally possible, and people are indifferent among them; game theorists call this a pure coordination problem. Examples of this might include driving on one side of the road, or setting a standard width for railroad tracks. In some cases, the community might spontaneously settle on one particular solution, in which case the extra weight of law adds little because the rule is self-enforcing. (Perhaps the law about driving on the right signals potential drivers what the convention around here actually is, in those rare cases where they do not already know.) In other cases, the community might fail to settle on one particular solution, and the law can then serve as a coordination device by signaling the equilibrium of choice.

Pure coordination problems are probably rare. Indeed, even the examples given above are only genuine pure coordination problems before people sink costs in some particular solution (e.g., by building train cars of a particular width). This leads us to our third, and much more important group of situations: impure coordination problems. These exist whenever there are several possible Nash equilibria, but people disagree with respect to which is best for the community or themselves (even when all agree that coordinating on any one solution is better than not coordinating at all). The battle-of-the-sexes game represents a generic example. Impure coordination problems constitute perhaps the single most important class of problems legal systems are capable of resolving. The bulk of property law, for example, falls under this heading: For the most part, people agree that any of a wide range of systems of property would be better than none at all, but obviously different property systems will have wildly different distributional consequences, and it is far from clear (especially as one gets down to the details) that any one system is best for everyone. The legal system plays a vital role in resolving such problems: On the one hand, laws can serve as mechanisms for settling on one particular equilibrium; on the other hand, by redistribution, laws can compensate those who are unhappy with the solution adopted (Finnis 1980, 231–33; Waldron 1999, 101–18).

FIGURE 4

		Player 2	
		Action A	Action B
Player 1	Action A	9, 10	11, 9
	Action B	10, 10	10, 11

A fourth category may not be so significant as the third, but it is nevertheless quite important: In some cases, the introduction of formal state

sanctions may induce social conventions not formerly available to the community. This will be the case when some hypothetical social convention is almost but not quite a Nash equilibrium on its own terms, but with formal state sanctions factored in may become one. A simple model of this is given in figure 4. In this game, there is no pure strategy Nash equilibrium. No matter what the players are doing, someone has an incentive to do something else. But now suppose the state decides to punish action A with some penalty "costing" the players –2 units of utility. In this case, the bottom-right-hand box becomes a Nash equilibrium, and the rule "always B" can then be a social convention enforced in part by formal sanctions. The most common real-world instances of this would probably be cooperation problems where the temptation to free ride is too strong to be overcome by internal and informal sanctions alone. Examples might include contract law or tax law: In large and complex communities, the likelihood of being informally sanctioned is low enough that the extra muscle of formal state sanctions is needed to sustain cooperation.[19]

It is probably the case that to be effective laws must enforce latent or near-latent informal social conventions, for rarely does a state have the resources necessary to induce a specific pattern of behavior on the basis of the threat of punishment alone (Tyler 1990, esp. chaps. 3–5). This, perhaps, is the lesson of Prohibition in the United States: Even the most effective and potent police force will have difficulty imposing a rule for action on a large population unless the rule is largely self-policing.

This preliminary list is not meant to be exhaustive, and of course there will be a number of interesting borderline cases. International laws, to the extent they are effective, I take to be instances of social conventions between states, and thus only law-like by analogy so long as these conventions are not actually enforced by some super state organization. Ineffective international laws are neither laws nor social conventions. Lapsed laws not observed in a community present a more challenging puzzle. On the one hand, they clearly cannot be described as social conventions, and so do not count as laws on the definition I have offered here. But on the other hand, insofar as state agents can begin enforcing lapsed laws at any time (at least in theory[20]), some legal positivists *would* count them as laws, and intuitively

19. In all but the smallest communities, formal state sanctions will probably be needed to induce citizens to pay their taxes. Note that the sanctions need not be very strong, nor even commonly enforced, if people generally respect the law as such, but it is hard to imagine that many people would voluntarily contribute for long if taxes were officially made optional. On contract law, there is some dispute as to whether a social convention of performance would exist in the absence of formal state sanctions: Elster 1989b, p. 149, argues no; Milgrom, North, and Weingast (1990) and Calvert (1995) argue yes. See also appendix B.

20. In some cases, a lapsed law may seriously contradict existing legal or political theory and practice, such that were some state agent to attempt to enforce it, the said law would be hurriedly repealed or voided. When it is clear to everyone that enforcement is not realistically possible, the lapsed law on the books can safely be regarded as not a law. The difficulty is, of course, that this is not always clear.

perhaps we regard them as such. The former route is more theoretically elegant for my concept of law, for reasons that will be especially clear once I have defined the Rule of Law in part three; but for the limited purposes of this paper, I must leave this question aside.[21]

To get a clearer idea of how social conventions and laws operate, I work through in appendix B an example drawn from contract law. First I show how a rule supporting the performance of contracts might operate as a social convention without the aid of formal sanctions (B.1). Second, I show that the existence of this social convention is compatible with rather high levels of actual deviance, given appropriate empirical assumptions (B.2). And third, I recast the model as one of a state-enforced social convention, noting some of the advantages of this latter situation over the first (B.3).

The Dynamic Aspect of a Legal System

Considered statically, laws are simply that subset of social conventions backed by formal state sanctions (in addition to informal sanctions, or not). In order to describe the dynamic aspect of law, we must expand our view of the legal system to include some social conventions not themselves law in the strict sense.

One advantage of using the language of game theory is that it enables us to see in a simple and elegant way how these important social conventions can exist without being enforced in the usual sense of being backed by the coercive powers of the state. In particular, I am thinking of something like Hart's rule of recognition or Kelsen's basic norm: a foundational rule such as "take the pronouncements of X as valid law," where X might be, for example, a monarch or a national legislature. Let me reiterate that the social convention in question here is not a behavior pattern according to which some members of the political community (e.g., state officials) happen to take the pronouncements of X as valid law—though a behavior pattern of this sort will in fact be observed. The social convention is the Nash equilibrium constituted by people having adopted the strategy to follow the rule. In other words, if everyone else in the political community is taking X's pronouncements as law, it will probably be a good idea to do so yourself (see Hart 1994, 94–95, 100–110; Kelsen 1989, esp. 193–214; and Postema 1982, 197–200).[22]

21. In order to address this problem, I would probably need to introduce a separate notion of validity, such that rules of law could be legally valid without actually being laws (unless there happened to be another rule in some legal system voiding lapsed laws). I am grateful to an anonymous *Law and Social Inquiry* reviewer for pointing out this problem.

22. By defining the foundational legal norm as a Nash equilibrium, we avoid having to think of it as a "transcendental-logical presupposition" as Kelsen did. The foundational legal norm is no more or less mysterious or problematic than Nash equilibria generally.

As we observed above, the existence of particular social conventions by no means implies anything about their intrinsic fairness or justice, nor does it suggest that people will not want to criticize them. Obviously, these properties also apply to the foundational legal convention "take the pronouncements of *A* as valid law" as well: It is in no way necessary that anyone actually following the foundational legal convention believe it to be just or fair, though often many will in fact believe this.[23]

The foundational legal convention cannot in the ordinary sense be enforced by anyone (i.e., it cannot be enforced by formal state sanctions), nor can it be regarded as the will or command of anyone. Rather, it must be a self-enforcing rule for action observed by the members of the community at large, or at least by enough of the officials in charge of administering the legal system to make it work.

The foundational legal convention can include effective limiting provisions, such that it takes the form "take the pronouncements of *X* as valid law so long as they conform to conditions *A*, *B*, etc., and not otherwise." It has been shown that a foundational legal convention of this sort—we might think of a constitution including a bill of rights—is sustainable as a Nash equilibrium in a sovereign-subject game, where constitutional resistance serves as the sanction on sovereigns who might otherwise be tempted to transgress the specified limiting provisions (see Hampton 1994; Weingast 1997).[24]

Other social conventions not laws in the strict sense might also be a part of a legal system, but I will not go into further detail here. The point is to imagine a hierarchy of social conventions constituting the complete legal system. At the base level, there are state-enforced social conventions specifying rules of action for the general population, such as "perform legally valid contracts."

On top of these base-level social conventions, there is a dense middle layer of social conventions relating to their application, such as "judges should enforce legally valid contracts by awarding damages," and "police officers should enforce judicial rulings by coercive force."[25] These latter rules for action must be Nash equilibria just like the former, though the players and the rules of the game will be different: Carrying out the formal sanctions backing base-level social conventions will be part of an equilibrium in a sort of meta-game including both the agents of the state and those subject to its laws. If the latter equilibria are supported in part by formal

23. Raz correctly emphasizes this point (1999, 147–48).

24. Thus the persistent difficulty in understanding how constitutional limits could possibly be enforced against the sovereign (Jean Bodin and Thomas Hobbes made much hay of this puzzle) is handily clarified by the Nash equilibrium concept.

25. The cut between base-level and mid-level social conventions is analogous to Hart's cut between primary and secondary rules (1994, esp. chap. 5), and Postema's cut between first- and second-level coordination problems (1982, 182–94).

sanctions, then they too will count as laws. An example of this might be the social conventions governing police conduct, where these are backed by formal sanctions. Otherwise, they are just social conventions, though still a part of the legal system broadly defined, insofar as they relate to the implementation of formal sanctions attached to the base-level social conventions. An example of this might be the so-called rule of four used by the American Supreme Court in deciding when to grant writs of certiorari. The middle layer of any well-developed legal system will probably be a complicated mixture of these types.

The hierarchy of equilibria is capped by a foundational legal convention like the one mentioned above. As we have said, this final social convention is necessarily not itself a law, though it is a part of the legal system broadly understood.

It is this hierarchy that makes dynamic legal change possible. Consider again the battle-of-the-sexes game discussed above, except that in addition to the husband and wife there is a third player. The third player does not himself go to the opera or the boxing match. Instead, he moves before the husband and wife, and his action set includes publicly saying either "go to the opera" or "go to the boxing match." It now possible for the husband and wife to adopt strategies of the sort, "go to the opera if the third player says 'go to the opera,' and go to the boxing match otherwise." These strategies then form a Nash equilibrium in the three player meta-game. (Interestingly, they can form a Nash equilibrium even if the third player is unable to enforce his command. The expectation that the wife will do what the third player says is itself enough to make the husband want to do the same, and vice versa.) By such means, the actions of certain players can be taken by others as signals to change what they are doing. This, roughly, is the basis of dynamic legal processes such as legislation and adjudication. Of course, if we were to fully spell out this sort of game, we would have to consider the incentive structure faced by the third player, and so forth.

Legal systems that empower certain members of the community to bring about changes in the general web of social conventions have important advantages over those that do not (cf. Hart 1994, 196–97). A game-theoretic analysis of social conventions suggests that the problem of multiple equilibria is probably endemic (for reasons briefly rehearsed in appendix C). Political communities may often find themselves stuck on inferior, inefficient, or unjust social conventions, and state action can then serve as a signal for everyone to move from coordinating on one equilibrium to coordinating on another (hopefully) better one. Obviously, people can be mistaken about whether the new equilibrium is actually better, and the signalers must have incentives to move toward equilibria that actually are better if the system is going to work properly. In a paper of this scope, I cannot hope to address these issues in any detail.

III

A law, as defined in part two, is a social convention enforced by formal state sanctions, generally in addition to informal sanctions. Now, it is certainly not the case that actual states limit their exercise of coercive powers to the enforcement of social conventions. For example, a state might coercively confine persons of a particular ethnic or racial group to internment camps; command the destruction of condemned property or dangerous animals; seize lands for public use by exercising eminent domain; delegate to immigration officials the discretionary authority to naturalize or not naturalize; and so on. From the point of view of those persons who actually feel the brunt of these sorts of coercive state action (those who are interned, those whose property is condemned, etc.), there is no social convention to which they are being asked to conform or not conform. This, however, is precisely the concern of the Rule of Law idea.[26]

As a first pass, let us simply define the Rule of Law as the rule *by* law, and not by some other means. As A. V. Dicey puts it, the Rule of Law requires that "no man is punishable or can be made to suffer in body or goods except for a distinct breach of law established in the ordinary legal manner before the ordinary courts of the land" ([1915] 1982, 110). In other words, a political community observes the Rule of Law to the extent that its members experience the brunt of coercive state power only when they have failed to comply with a law, defined as a state-sanctioned social convention. Naturally, the actual existence of the Rule of Law in a given political community will be a matter of degree, and could probably never be absolute. The extent or degree to which a given political community actually does conform to this ideal type is a descriptive social fact theoretically amenable to direct observation.

This definition of the Rule of Law is only preliminary, and needs to be considerably refined and explained. In particular, it may appear in one sense an almost pointlessly weak conception of the Rule of Law, and yet in another sense, an almost impossibly strong one. I will address these concerns in the third section of this part of the paper. Before doing so, however, I will first examine how far the principles traditionally associated with the Rule of Law can be grounded in the positivist concept of law developed in this

26. In some cases, state actions may have no real impact on members of the political community. Innocuous examples include meaningless declarations, such as "the official state bird shall be . . . " (cf. Fuller 1964, 91–92). Not so innocuous are state actions impacting (perhaps severely) persons outside the political community, such as many acts of foreign policy. There will be good or bad reasons for engaging in such actions, but these reasons will not have to do with the Rule of Law as such—unless perhaps one wants to develop a theory of the Rule of International Law analogous to the theory developed here for national legal systems. Such a theory is possible, in my view, so long as international laws are clearly understood as social conventions between states. Whether such a theory would be useful or not is a different question.

paper and second respond to a common criticism of the Rule of Law idea, namely that it is incoherent because rules of law are necessarily indeterminate in their meaning.

Traditional Principles of the Rule of Law

As a matter of social fact, a number of things must be the case in order for a social convention to exist. Because the Rule of Law means rule by law and not other means, and because laws are state-sanctioned social conventions, these necessary characteristics of social conventions must carry over into the idea of the Rule of Law. In this rather roundabout way, we can say that the Rule of Law "requires" or "demands" these characteristics. In other words, *if* a political community has something recognizable as a legal system, *then* it must be the case that Rule of Law principles are at least to some extent being observed. As many discussions concentrate on enumerating these principles (often without carefully linking them to a concept of law), it will be useful to sort out in a more rigorous way what they are exactly.[27]

a. For state-enforced social conventions to exist, of course, there must be underlying rules for action that members of the political community are expected to follow. In a contract law regime, for example, the rule for action might be "perform valid contracts." In order for a system of contract enforcement to count as part of a legal system in the sense here defined, state coercive action must be directed toward community members only when they fail to conform to the underlying rule for action. In other words, the contract law regime must present itself to the ordinary members of the political community as a set of rule-like propositions backed by formal sanctions.

I say "ordinary members of the community" because the law will often present itself differently in certain regards to officers of the state. Suppose the legislature in some political community enacts the statute "the police shall confine persons of class X to internment camps." This may count as a law, or at least as the fragment of a complete law, in a sense—namely from the point of view of the police. Since conforming to this rule must be a Nash equilibrium strategy for the said police (assuming the legal system is functioning), it must be the case that formal or informal incentive structures are in place to enforce it. If those incentive structures include coercive sanctions enforced by other state officials, then this social convention, "the police shall confine persons of class X to internment camps," counts as a law addressed to the police. From their point of view there is no violation of the Rule of Law. But from the point of view of those members of the

27. These criteria will be similar to those typically found in contemporary accounts of the Rule of Law (see note 1 above). For a comparative compilation of the specific requirements found in other accounts, see appendix A.

community in class X, the enforcement of this statute *is* a violation of the Rule of Law, because there is no rule for action *they* have been asked to conform to.

The requirement that the state limit itself to enforcing rule-like propositions is sometimes taken to be a requirement of generality, on the assumption that rules are by their nature general in form. This does not seem to me correct, however. A complete rule must have three parts, specifying (1) to whom it applies, (2) the factual conditions under which it applies, and (3) the sort of action(s) required (cf. Raz 1999, 50; Schauer 1991, 23–24).[28] Often, many aspects of a rule will be implied. For example, the rule "perform valid contracts" may imply that it applies to all members of the community, from now until the rule is changed. Thus, the rule is general over persons and times. Also, it applies to the general class of situations in which there is a "valid contract," however that is defined. Generality is not necessary, however. For example, a rule could state that "in exactly one week, Mary must perform her contractual obligation A owed to Paul, on the condition that it is possible for her to do so at that time." In no ordinary sense can this be considered a *general* rule, but it is a rule, to which sanctions for noncompliance may be assigned. Of course, in a very technical sense, even this rule might be thought general with respect to some narrow set of actions that would count as performing the contractual obligation, to some temporal window that would count as exactly one week from now, and so on. At the limits of linguistic specificity, then, we might say rules are necessarily general. The crucial point, however, is that properly formulated commands may count as rules, *pace* Hayek (1960, 149–51).[29] Bills of attainder fall afoul this first Rule of Law requirement not because they name specific persons, but because they are not formulated as rules to which compliance is possible.[30]

b. The second requirement is that the rule for action members of the political community are expected to conform with must actually be known to them. In game theory language, a Nash equilibrium cannot exist unless the strategies available to the players are public information. If the players cannot form expectations regarding the strategies others will adopt, they cannot make decisions concerning what strategies to adopt themselves. As the existence of informal social conventions makes clear, explicit promulgation is not necessary (though perhaps desirable in many circumstances), but the rule must at least be *capable* of being explicitly stated.

---

28. Raz terms (1) the "norm subject" and (3) the "norm act." Schauer terms (2) the "factual predicate."

29. Hayek believes it essential to distinguish laws from commands, and yet he does not do so in a theoretically satisfying way.

30. There is some confusion on this point, however, for if a bill of attainder is the expected sanction for some legally defined delict, it might under certain conditions count as the fragment of a law.

Likewise, the rule for action obviously must be clear enough and stable enough to be understood. On the one hand, if the observed rule is "perform valid contracts," it must be the case that people have a clear idea what sort of behavior constitutes performance and what sort of contracts count as valid according to the rule. If the rule were very complicated, not only would people have difficulty complying themselves, but their belief that others will also comply—essential to sustaining a Nash equilibrium—would begin to break down. On the other hand, even simple and clear rules defining valid contracts cannot be effectively known if they change too rapidly. The exact bounds clarity and stability place on the effectiveness of social conventions are a practical matter, and thus cannot be defined as a matter of principle.

c. The third requirement is that the rule for action be performable. Knowability and performability are logically separate requirements insofar as one can theoretically know what a rule for action is without being able to comply with it, and vice versa. Obviously, the rule for action must not impose impossible or unduly heroic demands. In game theory language, this is only to point out that a Nash equilibrium must be an equilibrium of strategies actually available to the players. The rule for action must also not be in contradiction with other rules for action in the general system of social conventions. If one thinks of the complete system of social conventions as one very large and complicated game, then the players cannot adopt strategies that involve undertaking incompatible actions. And finally, the rule for action must be prospective and not retroactive. A retroactive social convention simply does not make sense, for one cannot employ strategies that involve changing past actions.

Because social conventions are by necessity rule-like, knowable, and performable, so too are laws by our stipulated definition. Thus, in a sense, we can say that the Rule of Law "requires" that the coercive powers of the state be used only to enforce knowable and performable rules. Let me be clear about the sort of claim I am making here. This is not directly a normative claim. There may be (and I believe there are) normatively desirable consequences of conforming to the Rule of Law at least to some degree, but whether or not a given political community does so makes no difference when it comes to describing the Rule of Law as such. The claim here is not that laws *should* be rule-like, knowable, and performable, but rather that a coercive state action would fail to be an instance of rule by law if it did not have these properties. And this is only because something without these properties cannot, as a matter of social fact, be a social convention. The features of the Rule of Law are simply the features of social conventions.

The Indeterminacy Thesis

In this section, I would like to briefly respond to one of the most common criticisms of the Rule of Law idea: namely, that it does not take seriously enough what is sometimes called the indeterminacy thesis.[31] Roughly, the objection is that concepts like performance or valid contract are so pervasively plagued by indeterminate meanings that no stated definition, no matter how clear or explicit, could ever resolve all the questions of its own interpretation. This objection, of course, is closely related to the first two objections we considered when discussing social conventions earlier, except that here the problem is often taken to be one of legal language. As we noted before, if this problem were taken too seriously, it is difficult to see how social conventions could operate at all. The social convention of waiting in line for service at a bank obviously exists in some communities, and so it *must* be the case that people in general know that it means to wait in line, even if they cannot offer an unassailably explicit definition of it. Regardless of whether, or to what extent, the indeterminacy thesis is true as a philosophical question, we can simply insist that if it makes no difference in practice, there is not much point to disputing it in theory.

In fact, however, the indeterminacy thesis is largely misdirected. Much confusion results from failing to distinguish a rule of law from the law itself. The law itself is a state-sanctioned social convention existing in the world of social facts. A *rule of law* is a propositional statement in some natural language (e.g., in English), made by a legislature, a judge, a lawyer, or whomever, regarding what the law is in a particular community. The actual law is the intended referent of a rule of law proposition. A rule of law proposition can be viewed as more or less valid depending on how accurately its semantic meaning corresponds to the actual law to which it refers (i.e., to the social fact of a state-sanctioned social convention).

If we take the indeterminacy thesis to apply only to the rule of law, and not to the law itself, then it presents no difficulty for the Rule of Law idea. As a claim about the impossibility of the semantic meaning of some proposition *P* ever perfectly corresponding to its object of reference, a social fact *S*, the indeterminacy thesis is probably correct. But this, of course, does not matter for a theory of the Rule of Law—or at least to a theory of the Rule of Law grounded on a positivist account of law. The Rule of Law idea is concerned with what is *actually* going on from the point of view of those persons subject to a state's coercive authority, and not with the linguistic possibility of our *saying* what is going on. The fact that our description of *X*

31. This sort of objection is typically associated with legal realists such as Karl Llewellyn and Jerome Frank, and more recently with the critical legal studies movement. For good overviews of the extensive critical legal studies literature, see Kelman 1987 or Altman 1990.

and its properties will necessarily be approximate does not say anything about the nature of X itself.

Limits of the Rule of Law Idea

As I mentioned above, the Rule of Law idea presented here may seem in one sense pointlessly weak, and yet in another sense impossibly strong. While unfortunately I will not be able to provide a definite solution to either problem here, I can at least illuminate both as clearly as possible. It is my hope that by formulating the Rule of Law idea in rigorous positivist language, I have at least exposed those issues that stand in need of further analysis.

The Rule of Law idea set out in this paper may seem pointlessly weak to those like Hayek, Fuller, and others, who expect Rule of Law principles to positively exclude a fairly wide range of what are viewed as unjust state activities. The main reason for this lies in confusion regarding the first Rule of Law principle mentioned above, that the legal system present itself as a set of rule-like propositions.

As I noted there, this is sometimes described as a requirement of generality. This description is misleading, however, because a perfectly sensible rule for action can restrict its application to a narrow, even more or less unique, range of times, persons, and situations. In light of this, we might imagine a (quite implausible) legal system in which each individual member of the political community is subject to a different and unique personal code of rules. No community would ever attempt such a thing, but if it did, the arrangement would not be a violation of the Rule of Law. For those who hoped that the Rule of Law idea would exclude, for example, discriminatory laws, the theory developed here may seem very weak indeed. (Of course it *would* exclude those forms of discrimination carried out by direct state coercion.) By no means do I want to suggest that legal equality is not a good thing, but in my view, it stands in need of an independent normative justification. Many previous accounts of the Rule of Law trade on their failure to separate legal equality from other principles intrinsically connected with the concept of law, suggesting first as a conceptual point that law is necessarily egalitarian to some degree, and then drawing on our egalitarian moral intuitions when subsequently arguing the Rule of Law is a good thing. The analysis here exposes this as mere sleight of hand.

Nevertheless, even this weak conception of the Rule of Law idea remains impossibly strong in an important sense. Consider again Dicey's statement of the Rule of Law: that "no man is punishable *or can be made to suffer in body or goods* except for a distinct breach of law." If we are to take this claim seriously, then *any* state action causing some member of the political community to suffer in body or goods would be a violation of the Rule of

Law, unless that suffering were the punishment for a breach of law. But clearly this is impossibly demanding. For example, if Congress revises parts of the tax code, some members of the political community will inevitably be made to suffer (and others to profit) in goods; if Congress cuts funding for public hospitals or military bases, there will be corresponding injuries to those who benefited from such expenditures; if the Federal Communications Commission alters its licensing guidelines, this will differentially impact actual and potential radio carriers; and so forth. In each case, these actions would seem to be violations of the Rule of Law.

It should be immediately clear that very nearly anything the state decides to do or not do will have some deleterious impact on *someone* in the political community, and there is simply no obvious way to connect each such instance to a distinct breach of law. This has always been true, of course (even Medieval European states raised taxes, regulated commerce, etc.), but perhaps it has only become starkly apparent with the advent of the modern welfare state. Instead of being exceptional, these sorts of activities now constitute a large part, perhaps even a majority, of what modern states do.

At least four possible responses to this problem have been suggested. The first two are obvious, but rather drastic: On the one hand we might say, "so much the worse for the welfare state"; on the other hand we might say, "so much the worse for the Rule of Law idea." The first path is taken by Hayek and like-minded libertarians. The second is taken by a variety of radical leftists and, on some interpretations, also by some critical legal studies adherents. Fortunately, more nuanced middle-of-the-road responses are available.

The first accepts and even emphasizes the distinction between the traditional sort of law-enforcing state activities and the now more common sorts of activities variously described as regulative, directive, and so forth. If these two sorts of state activity are viewed as fundamentally different, then it is perhaps plausible to argue that the Rule of Law idea is relevant only with regard to the former. In somewhat different ways, both Fuller (1969) and Edward Rubin (1989) argue for this view. If one agrees with Rubin that the newer sorts of state activities are likely to largely or completely supplant the older, then the Rule of Law idea might be viewed as correspondingly obsolete (though not false, per se).

The second middle-of-the-road response also accepts that the newer sorts of state activity differ in certain regards from the older, but not so much that the Rule of Law idea no longer applies: rather, it applies in a somewhat different manner.[32] For example, suppose those officials charged with making budget decisions, setting regulatory policy, and so on must carry out these activities according to certain rules that are publicly known

32. I would like to thank Jeremy Waldron for suggesting in conversation what follows.

to the ordinary members of the community. Further suppose these rules are actionable by persons made to suffer in body or goods as a result of such decisions by state officials; in other words, the injured citizen would have a case at law if some state official failed to follow the rules in making the decision leading to the injury in question. This might be regarded as a sufficient equivalent to satisfying the Rule of Law in the traditional sense.

It is beyond the scope of this paper to argue at length for one or the other of these options, though naturally I incline toward the last two.

IV

In this paper I have attempted only to describe the Rule of Law idea, and to lay bare its precise connection with the concept of law as such. Naturally, a great many questions of interest have of necessity been left aside. In concluding this paper, I would like to gesture toward two such questions in particular.

First, what sorts of institutional arrangements are necessary for some degree of the Rule of Law to operate in a given political community? It is sometimes doubted that the genuine Rule of Law could ever be secured, given the apparent impossibility of restraining power by mere rules; the view that such restraint is indeed impossible is sometimes called *rule skepticism*.[33] Institutional mechanisms, in my view, are the means of getting around this. One idea behind the separation of powers system, for example, was precisely that since only power can effectively oppose power, we must cleverly design institutions such that the battle lines drawn between competing powers happen to coincide with boundaries set by the Rule of Law (see the *Federalist Papers*, esp. nos. 47–48, 51). Other institutional structures associated with the Rule of Law include an independent judiciary, the jury trial, judicial review, universal access to courts, procedural due process, and so on.

In one sense, if the Rule of Law as a practical matter cannot exist without institutional supports of the sort mentioned, we might regard the guarantee of these institutions as additional principles of the Rule of Law along with those discussed in part three. It is useful, however, to keep these two sets of requirements distinct, because they differ in the sense in which they are requirements of the Rule of Law. The Rule of Law *means* to be governed by knowable and performable rules. These are logical requirements, much in the sense that being unmarried is a logical requirement for being a bachelor. While securing the Rule of Law might be impossible as a practical matter without an independent judiciary, universal access to courts, and so on, the Rule of Law is not identical with having these

33. The indeterminacy thesis is sometimes taken as one of the many grounds for rule skepticism.

institutions. If in the future we discover some quite different set of political institutions that manage to secure government by knowable and performable rules even more effectively than those institutions we presently have, we would not then say that abandoning the latter in favor of the former amounts to abandoning the Rule of Law.

The second major question I have left aside in this paper is the normative one regarding whether it is good for a political community to maximize the Rule of Law. I have argued that describing the Rule of Law is a task separate—or at least separable—from the task of showing why or to what extent it is a good thing. Of course if it were truly a matter of indifference, we would have no reason to be interested in describing the Rule of Law in the first place, much less determining whether some particular community conforms to it. Unfortunately, however, I must conclude this paper without indicating much more than the form such an argument would take.

Many normative arguments against the Rule of Law consist in showing on the one hand that many desirable public policies do not conform to Rule of Law principles (antitrust law might be an example of this), and on the other hand that many undesirable public policies do (one could imagine, perhaps, a scrupulously legal system of racial segregation). In my view, these arguments compare apples and oranges: good policies in violation of the Rule of law with bad policies in conformity with the Rule of Law. The proper comparison is between the two modes of securing any given policy: Given a good policy, should the members of a political community prefer that its government pursued that good policy via the Rule of Law? Given a bad policy, should the members of a political community prefer its government pursued that bad policy via the Rule of Law? I believe an affirmative answer could be strongly defended in both instances.[34] It is for this reason that the Rule of Law idea remains of interest to legal and political philosophy.

REFERENCES

Altman, Andrew. 1990. *Critical Legal Studies: A Liberal Critique*. Princeton, N.J.: Princeton University Press.

Calvert, Randall L. 1995. Rational Actors, Equilibrium, and Social Institutions. In *Explaining Social Institutions*, ed. Jack Knight and Itai Sened. Ann Arbor: University of Michigan Press.

Coleman, Jules L., and Brian Leiter. 1996. Legal Positivism. In *A Companion to Philosophy of Law and Legal Theory*, ed. Dennis Patterson. Cambridge, Mass.: Blackwell.

Dicey, A. V. [1915] 1982. *Introduction to the Study of the Law of the Constitution*. Reprint, Indianapolis, Ind.: Liberty Fund.

34. I take Thompson (1975) to be making an argument for the second conclusion, for example.

Dworkin, Ronald. 1977. *Taking Rights Seriously*. Cambridge, Mass.: Harvard University Press.

———. 1986. *Law's Empire*. Cambridge, Mass.: Harvard University Press, Belknap Press.

Elster, Jon. 1989a. *The Cement of Society*. Cambridge, England: Cambridge University Press.

———. 1989b. *Nuts and Bolts for the Social Sciences*. Cambridge, England: Cambridge University Press.

Fallon, Richard H. 1997. The Rule of Law as a Concept in Constitutional Discourse. *Columbia Law Review* 97:1–56.

Finnis, John. 1980. *Natural Law and Natural Rights*. Oxford, England: Clarendon Press.

Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge, Mass.: MIT Press.

Fuller, Lon L. 1969. *The Morality of Law*. Rev. ed. New Haven, Conn.: Yale University Press.

Hampton, Jean. 1994. Democracy and the Rule of Law. In *The Rule of Law: Nomos 36*, ed. Ian Shapiro. New York: New York University Press.

Hart, H. L. A. 1994. *The Concept of Law*. 2d ed. Oxford, England: Clarendon Press.

Hayek, Friedrich A. 1944. *The Road to Serfdom*. Chicago: University of Chicago Press.

———. 1960. *The Constitution of Liberty*. Chicago: University of Chicago Press.

Kelman, Mark. 1987. *A Guide to Critical Legal Studies*. Cambridge, Mass.: Harvard University Press.

Kelsen, Hans. 1989. *Pure Theory of Law*, trans. Max Knight. Berkeley and Los Angeles: University of California Press, 1967. Reprint, Gloucester, Mass.: Peter Smith (page references are to reprint edition).

Neumann, Frantz L. [1937] 1996. The Change in the Function of Law in Modern Society. In *The Rule of Law Under Siege: Selected Essays of Frantz L. Neumann and Otto Kirchheimer*, ed. William E. Scheuerman. Berkeley and Los Angeles: University of California Press.

Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs. *Economics and Politics* 2:1–23.

Posner, Eric A. 2000. *Law and Social Norms*. Cambridge, Mass.: Harvard University Press.

Postema, Gerald J. 1982. Coordination and Convention at the Foundations of Law. *Journal of Legal Studies* 11:165–203.

Radin, Jane. 1989. Reconsidering the Rule of Law. *Boston University Law Review* 69:781–819.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, Mass.: Harvard University Press, Belknap Press.

Raz, Joseph. 1979. *The Authority of Law*. Oxford, England: Clarendon Press.

———. 1999. *Practical Reason and Norms*. 2d ed. Oxford, England: Oxford University Press.

Rubin, Edward L. 1989. Law and Legislation in the Administrative State. *Columbia Law Review* 89:369–426.

Schauer, Frederick. 1991. *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*. Oxford, England: Clarendon Press.

Scheuerman, Bill. 1994. The Rule of Law and the Welfare State: Toward a New Synthesis. *Politics and Society* 22:195–213.

Shklar, Judith. 1998. Political Theory and the Rule of Law. In *Political Thought and Political Thinkers*, ed. Stanley Hoffman. Chicago: University of Chicago Press.

Tyler, Tom R. 1990. *Why People Obey the Law*. New Haven, Conn.: Yale University Press.

Thompson, E. P. 1975. *Whigs and Hunters: The Origin of the Black Act*. London, U.K.: Allen Lane.

Waldron, Jeremy. 1990. *The Law*. London, U.K.: Routledge.

———. 1999. *Law and Disagreement*. Oxford, England: Clarendon Press.

Weber, Max. [1919] 1946. Politics as a Vocation. In *From Max Weber: Essays in Sociology*, ed. H. H. Gerth and C. Wright Mills. New York: Oxford University Press.

Weingast, Barry R. 1997. Democracy as a Self-Enforcing Equilibrium. In *Understanding Democracy: Economic and Political Perspectives*, ed. Albert Breton, et al. Cambridge, England: Cambridge University Press.

Wittgenstein, Ludwig. 1958. *Philosophical Investigations*, trans. G. E. M. Anscombe. 3d ed. Englewood Cliffs, N.J.: Prentice Hall.

APPENDIX A
Compilation of Rule of Law Theories

	Fuller (1969, 46–91)	Rawls (1971, 236–39)	Raz (1979, 214–18)	Finnis (1980, 270)
I. Rule of Law requirements				
A. Rule-like	[1] There must be rules	[2] Similar cases . . . treated similarly* [3c] General both in statement and intent*		
B. Knowable	[2] Promulgation [4] Clarity of laws [7] Laws should not be changed too frequently	[3a] Known and expressly promulgated [3b] Meaning clearly defined	[1b] Laws should be . . . open and clear [2] laws . . . relatively stable	[3] Rules are promulgated [4] Rules are . . . clear [6] Rules are sufficiently stable
C. Performable	[3] System of rules . . . generally prospective [5] No contradictions [6] Must not require the impossible	[1a] Law . . . must not impose a duty to do what cannot be done [1c] Recognize impossibility of performance as a defense [3e] Penal laws not retroactive	[1a] Laws should be prospective	[1] Rules are prospective [2] Rules . . . not . . . impossible to comply with [5] Rules . . . coherent one with another
II. Restatements of the definition of the Rule of Law	[8] Congruence between official action and the law		[3] The making of particular laws . . . should be guided by open, stable, clear, and general rules [8] Crime-preventing agencies should not be allowed to pervert the law	[7] The making of decrees and orders . . . guided by rules [8] Those people . . . in an official capacity are accountable . . . and do actually administer the law
III. References to the institutional requirements of the Rule of Law		[1b] Those who enact laws and give orders do so in good faith	[4] Independence of the judiciary . . . guaranteed [6] Courts should have review powers [7] Courts should be easily accessible	
IV. Miscellaneous		[3d] Severe offenses strictly construed [4] Natural justice observed**	[5] Principles of natural justice must be observed**	

* These requirements fit in this box only to the extent that they are interpreted narrowly, as requiring rule-like propositions. To the extent that they imply a stronger notion, e.g., some form of legal equality, they belong in box IV (or possibly III).

** To the extent that these principles refer to (procedural) due process, they might belong in box III.

APPENDIX B: MODELS OF CONTRACT PERFORMANCE

The models developed in this appendix are drawn from those found in Milgrom, North, and Weingast (1990), Fudenberg and Tirole (1991, ch. 5), and Calvert (1995), and in no sense can be considered particularly original. I will begin by showing the general conditions under which a social convention of contract performance can exist in the absence of the formal sanctions of a legal system.

B.1. Suppose a community N with $\{1, 2, \ldots n\}$ players, such that $n \geq 2$ and even. These players engage in an indefinite series of one-on-one contractual encounters with one another; each such encounter is modeled as a simple stage game.

In each period of the overall game $t = 0, 1, 2 \ldots$, has two steps. First, the players are paired randomly with one another, such that the probability of any particular player i being paired with any particular player j is $1/(n-1)$. Let π^t be a particular pairing of N in period t, and let Π be the set of all possible such pairings for any given period.

Once this random pairing has occurred, the players in each pairing play the stage game—a standard prisoners' dilemma as in figure B.1,

FIGURE B.1

		Player j	
		Perform	Cheat
Player i	Perform	1, 1	$-b$, a
	Cheat	a, $-b$	0, 0

such that $a > 1$ and $-b < 0$. Here a represents the incentive to cheat, and b represents the severity of the risk performing parties expose themselves to.

This completes one period; in the next period, the players are again paired randomly, each pairing plays one round of the stage game, and so on. The players discount future payoffs by a factor of $0 < \delta < 1$, measuring the degree to which they care about the future. For example, if each period is taken to represent one year, a discount rate of 0.9 means that one would pay \$9 today to receive \$10 one year from now. Another way to think of it is that a discount rate of 0.9 is roughly equivalent to an interest rate of 11%. The higher their discount rate, the more a person cares about the future.

Assume that the players have complete and perfect information about their own actions and the actions of the other players in all past periods. Let h^t represent everything that occurred in period t, and let $\mathsf{h}^t = h^t + h^{t-1} + \ldots h^0$ represent the complete history of the game up to period t. Let H be the set of all possible single-period histories, and H be the set of all possible game histories. Note that while H is closed and bounded, H is not because the game has an indefinite number of periods.

Let s_i be a strategy for player i, and let $s = (s_1, s_2, \ldots s_n)$ be a profile of strategies for all the players in N. A complete strategy is a mapping s_i: $\mathsf{H} \times \Pi \rightarrow$ {Perform, Cheat}, and the set of strategies available to player i is S_i. Mixed strategies will not be allowed, but nevertheless the strategies space S_i is quite large indeed. In fact, it is not bounded, because H is not.

Fortunately, we are interested in one particular strategy, the well-known tit-for-tat strategy. According to s_i^{TFT}, player i will perform in period t if the player j she is paired with at t performed in period $t-1$, or if j cheated in period $t-1$ in order to punish an earlier offender. If j cheated in period $t-1$, of if j failed to punish an earlier offender in that period, i will cheat in period t in order to punish j for her actions in the pervious period. Note that punishments will typically not be levied by the initially cheated player, unless the latter happens to be paired with the cheater two periods in a row. This strategy generally does not require perfect recall on the part of the players. In most cases, they need only remember the events of the past few periods in order to distinguish between genuine cheaters and those cheating only to punish earlier offenders. (If the players are

unable to make this distinction, strange modeling problems arise, which I gloss over by allowing perfect recall [see Fudenberg and Tirole 1991, 172–74, for a discussion]).

Under what conditions will the strategy profile s^{TFT} constitute a sub-game perfect Nash equilibrium? First note that our game is continuous at infinity because with discounting, payoffs approach zero as $t \to \infty$. According to the optimality principle of dynamic programming it will be sufficient to show that, assuming all $i \in N$ are playing s_i^{TFT} already, no one player has an incentive to deviate for one period from her equilibrium strategy. This is because discounting ensures that deviations will only become more costly as they are carried beyond one period. If we can show this, then s^{TFT} is a sub-game perfect Nash equilibrium (see Fudenberg and Tirole 1991, 108–10 for a statement and proof of this principle, and a discussion of how it applies to repeated games).

Consider an arbitrary player $i \in N$ who has been paired with $j \in N$ in period t. There are a finite number of possible one-period deviations to consider. First, i could deviate in period t by failing to punish j if (counterfactually) j had deviated from playing s_j^{TFT} in period $t - 1$; obviously no incentive exists for this deviation because $a > 1$: punishing offenders carries its own reward.

Second, we must show that if player i cheated in period $t - 1$ (and not just to punish some earlier cheater), it is always in her advantage to accept punishment now by performing in this and future periods, and not postpone punishment by continuing to cheat for one extra period. Suppose she decides to postpone punishment. The sequence of events from i's perspective are represented in table B.1.

TABLE B.1

Strategy (s_i, s_j^{TFT}):	$t - 1$	t	$t + 1$	$t + 2$	. . .
Accept punishment now	C, P	P, C	P, P	P, P	. . .
Postpone punishment	C, P	C, C	P, C	P, P	. . .

Note that the sequence of events resulting from accepting punishment in period t and accepting it in period $t + 1$ will be the same from period $t + 2$ on. Therefore, we need only compare i's payoffs in the two periods t and $t + 1$. This being so, i will not have an incentive to postpone punishment so long as $-b + \delta \geq -b\delta$, or

$$\delta \geq \frac{b}{1 + b} \qquad [1]$$

In other words, cheating players will return to cooperation right away so long as their short-term punishment $-b$ is not too severe relative to their discount rate.

Suppose that condition [1] holds. We must next show that assuming both players performed in period $t - 1$, player i does not have an incentive to cheat in period t. Since we have assumed that cheating players will return to equilibrium as soon as possible, from period $t + 2$ on the payoff to player i will be the same whether she cheats in period t or not; therefore, we need only be sure that $1 + \delta \geq a - b\delta$, or

$$\delta \geq \frac{a - 1}{1 + b} \qquad [2]$$

In other words, the one-time cheating gain cannot be too high relative to the player's discount rate, though the temptation of large potential gains can be counterbalanced by

severe punishments (to the point, of course, where condition [1] comes into action). Therefore, the strategy profile s^{TFT} is a sub-game perfect Nash equilibrium if $\delta \geq \max\{b/(1+b), (a-1)/(1-b)\}$. The only absolute limit placed by these conditions is that $a - b < 2$. Provided this is the case, there exists a discount rate high enough to support an equilibrium no matter how great the temptations to cheat. This suggests that social conventions against cheating are possible in the absence of formal legal institutions, so long as people care strongly enough about the future.

One interesting note: from the external perspective, what one observes in the tit-for-tat equilibrium is all the players performing all the time. This pattern of behavior is not the equilibrium: the equilibrium is between the strategies, not between the observed actions. To put it another way, the rule being followed is not "always perform," but rather "always perform with past performers, always cheat with past cheaters." The social convention is the latter rule, not the former. It just so happens that since cheating is never observed, the pattern of behavior looks the same.

B.2. In the tit-for-tat equilibrium described above, all players perform their contractual obligations all the time. This may seem unrealistic given that even widely held social conventions are at least occasionally violated in the real world. It is thus sometimes complained against practice-based theories of social convention that they require what in fact never exists: perfect compliance.

Suppose therefore that players randomly deviate from their equilibrium tit-for-tat strategy with a probability of ε, such that $0 < \varepsilon < 1$. This random deviation can be interpreted as capturing imperfect information, as for example when players mistakenly believe that they have been paired with a cheater who needs to be punished; or it can be interpreted as capturing irrational overestimates of the gains from cheating, as miscalculations of future payoffs, or as something else, or as a combination of these.

Now in any period of the game there will be four relevant groups of players: those who randomly deviated last period and will randomly deviate this period as well; those who randomly deviated last period but will not this period; those who did not randomly deviate last period but will now; and those who did not randomly deviate last period and will not this period either. The proportions of players in N falling into each group are $\varepsilon \cdot \varepsilon$, $\varepsilon \cdot (1-\varepsilon)$, $(1-\varepsilon) \cdot \varepsilon$, and $(1-\varepsilon) \cdot (1-\varepsilon)$ respectively. Interestingly, these proportions will turn out to be the same every period after the first.

Taking into account this error factor, condition [1] becomes:

$$-b(1-\varepsilon) + p + \delta((1-\varepsilon) - b\varepsilon) \geq (1-\varepsilon)(-b\delta) + \delta\varepsilon + 0(1-\varepsilon) + a\varepsilon$$

$$\delta \geq \frac{b + \varepsilon(a - b - 1)}{1 + b - 2\varepsilon(1 + b)} \qquad [3]$$

which, as we should expect, at the limit $\varepsilon \to 0$ reduces to condition [1] above. Similarly, condition [2] becomes:

$$(1-\varepsilon) - b\varepsilon + \delta\,((1-\varepsilon) - b\varepsilon) \geq a(1-\varepsilon) + (0)\varepsilon - \delta b(1-\varepsilon) + \delta\varepsilon$$

$$\delta \geq \frac{a - 1 + \varepsilon(1 - a + b)}{1 + b - 2\varepsilon(1 + b)} \qquad [4]$$

which likewise reduces to [2] as $\varepsilon \to 0$.

Random deviations from the tit-for-tat equilibrium make the social convention against cheating harder to sustain. Table B.2 gives some representative examples.

TABLE B.2

Discount Factor (δ)	Temptation to Cheat (a)	Punishment (b)	Max Error (ε)
0.90	2.0	1.0	$\varepsilon \leq 0.22$
0.90	2.5	1.5	$\varepsilon \leq 0.17$
0.90	3.0	2.0	$\varepsilon \leq 0.13$

Thus, for example, given the moderately high discount factor of 0.9, when a player can double her one-period payoff by cheating ($a = 2$), a social convention against cheating can still be supported as an equilibrium despite the fact that players expect to be cheated more than once in every five encounters. As the temptation to cheat grows, however, the acceptable level of deviance falls, but does not vanish.

B.3. Although we showed in B.1 that a contract-performance social convention can be self-enforcing without the aid of formal sanctions, some difficulties should be noted. First, the equilibrium places demanding information requirements on the players. They are expected to be able in each case to determine on their own initiative whether their current contractual partner performed last period; this may be difficult to do. Second, the equilibrium relies on what may be a rather stringent balance of incentives and discount factors; thus a wide range of profitable but risky contracts may be impossible to carry out. These limits become even more pressing once random deviance is taken into account.

Let us therefore modify the game to model a situation in which contract performance is enforced by formal sanctions. Our community N now has $n + 1$ players, $\{1, 2, \ldots n, g\}$, where g represents the government player—here taken to represent the state-enforcement apparatus (in an admittedly abstract way).

The stage game is now somewhat more complex: In particular, each stage is now composed of five steps. First, the nongovernment players $1, 2, \ldots n$ are randomly paired as before. Second, they play one round of the stage game in Figure B.1 above. So far, the game is exactly the same.

Third, however the stage game turns out, each nongovernment player can opt to appeal to the government for contractual relief. For each pairing in which one or both of the players appeal, the government is presented with a case. Let us assume that the government can hear any number of cases.

Fourth, the government player issues a ruling in each case of that period. There are three possible rulings in a case between i and j: leave things as they are, rule in favor of i by forcing j to pay damages d to player i, or rule in favor of j by forcing player i to pay damages d to player j. For the moment, the amount of d is unspecified.

Finally, in a fifth step, each nongovernment player may opt to pay a tax c to the government player, where $0 \leq c \leq 1$.

A complete strategy s_i for player i must now include what do to at each step of each stage of the game. Similarly, a complete strategy s_g for the government player must include how to rule in each possible case that might be brought to its attention. Since these strategies can be very complex, I will not attempt to specify them fully.

We are interested in the conditions under which one particular strategy profile s^{CLR}, or "contract law regime" profile, constitutes a Nash equilibrium. The player strategies in this profile are as follows:

s_i^{CLR}: Always perform in the stage game, unless the player one is paired with did not pay her tax last period and should have—in which case, cheat. If cheated in the stage game, always appeal to the government for relief, but not otherwise. Always pay the tax, unless the government failed to provide relief in a case brought by the player last round.

s_g^{CLR}: Always rule in favor of a player who performed when her opponent cheated, if the performing player paid her tax in the previous period. Otherwise, leave things as they are.

These are rather informal descriptions of the strategies, but they should be sufficient for the task at hand.

As in B.1, we need only consider the possibility of single-period deviations to be sure that s^{CLR} is a Nash equilibrium. First, note that given all the nongovernment players' strategies, the government player will never have any incentive to deviate, because each deviation will cost her c and gain her nothing. Provided she does not deviate from her equilibrium strategy, the government player will receive a payoff of $c \cdot n$ each period.

The only question is whether the nongovernment players ever have an incentive to deviate. Table B.3 indicates the anticipated payoffs to player i for five different deviations from s^{CLR}, supposing no other player deviates.

TABLE B.3

Strategy (s_i, s_j^{CLR})	$t-1$	t	$t+1$	$t+2$	...
Equilibrium strategy (s_i^{CLR})	$1-c$	$1-c$	$1-c$	$1-c$	...
Cheat once; pay tax	$1-c$	$a-d-c$	$1-c$	$1-c$	...
Don't pay tax once	$1-c$	1	$-b-c$	$1-c$	...
Cheat once; don't pay tax	$1-c$	$a-d$	$-b-c$	$1-c$	...
Don't pay tax; cheat in anticipation	$1-c$	1	$-c-d$	$1-c$	...
Cheat once; don't pay tax; cheat in anticipation	$1-c$	$a-d$	$-c-d$	$1-c$	...

As we can see, all one period deviations have different payoffs only in periods t and $t+1$. Consider the deviation of cheating once but paying one's tax (this ensures that the player one is paired with next period will perform, according to the equilibrium strategy). There is no incentive for this so long as $a - d - c \geq 1 - c$, or

$$d \geq a - 1 \qquad [5]$$

In other words, the damages paid by the cheating player must be greater than the gains from cheating minus what one would have received without cheating.

Let us suppose that condition [5] holds, as we would certainly expect in any contract law regime. Now consider the second deviation of not paying one's tax. Since the government will not protect players who did not pay their taxes in previous periods, one's partner in the following period effectively has a license to cheat. There is no incentive for this so long as $1 + \delta(-b - c) \leq 1 - c + \delta(1 - c)$, or

$$\delta \geq \frac{c}{1+b} \qquad [6]$$

Now consider the third possible deviation. This is the same as the first, but now the cheating player does not pay her tax. There is no incentive for this so long as $a - d + \delta(-b - c) \leq 1 - c + \delta(1 - c)$, or $\delta \geq (c + a - d - 1)/(1 + b)$. But since we have supposed that condition [5] holds, $a - d - 1 \leq 0$, and this last inequality will always be satisfied whenever condition [6] is.

Next, consider the fourth possible deviation. This is the same as the second, but now i anticipates being cheated in $t + 1$ and so cheats as well. There is no incentive for this so long as $1 + \delta(-c - d) \leq 1 - c + \delta(1 - c)$, or

$$\delta \geq \frac{c}{1 + d} \qquad [7]$$

Finally, we need to consider the fifth deviation, which effectively combines all the previous deviations. There is no incentive for this so long as $a - d + \delta(-c - d) \leq 1 - c + \delta(1 - c)$, or $\delta \leq (c + a - d - 1)/(1 + d)$. Again, however, since we have supposed condition [5] holds, $a - d - 1 \leq 0$, and thus the satisfaction of condition [7] will entail the satisfaction of this additional inequality. In short, we need concern ourselves only with the two conditions, [6] and [7] so long as we assume [5] holds.

To get a sense of what sort of conditions are required to support s^{CLR} as a Nash equilibrium, consider table B.4.

TABLE B.4

a	*B*	*d*	*c*	[6]	[7]
1.50	1.00	$d > 0.50$	0.10	$\delta \geq 0.05$	$\delta \geq 0.07$
1.50	1.50	$d > 0.50$	0.10	$\delta \geq 0.04$	$\delta \geq 0.07$
2.00	1.00	$d > 1.00$	0.10	$\delta \geq 0.05$	$\delta \geq 0.05$
2.00	1.50	$d > 1.00$	0.20	$\delta \geq 0.08$	$\delta \geq 0.10$

As can be seen, this equilibrium is very easy to support. Curiously, the most important factor turns out to be the tax rate c paid to the government. The absolute possibility of equilibrium will break down only where $c > 1$, which would mean the players were being asked to pay more in taxes than they could earn by performing their contracts.

APPENDIX C: SOCIAL CONVENTIONS AND THE FOLK THEOREM

In this appendix, we will consider a basic repeated game similar to the one described in appendix B.1, except that the stage game in figure C.1 takes the place of the prisoners' dilemma game used previously.

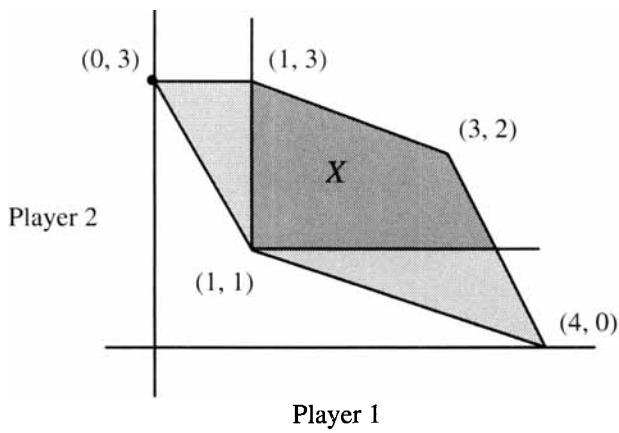
FIGURE C.1

		Player 2		
		Left	Center	Right
	Top	0, 3	4, 0	1, 3
Player 1	Middle	3, 0	3, 2	1, 3
	Bottom	1, 1	1, 1	1, 1

This game is not meant to have any real-life analogue. It is merely a device intended to illustrate a point about social conventions in general.

Imagine the stage game in figure C.1 is indefinitely repeated, but ignore the effect of discounting for the moment. This game can then be displayed in cross section, as it were, by figure C.2.

FIGURE C.2



In this figure, each point represents a stream of single period payoffs. Thus the point (1, 1) represents the stream of payoffs to the players if player 1 is playing "bottom" in every stage game, and player 2 is playing "left" in every stage game. Let us now permit mixed strategies. It then becomes possible to achieve any payoff stream in the shaded area above. This area is called the "convex hull" of the pure strategy payoff stream coordinates.

Now suppose the players use what is called a "grim trigger" strategy, s^{GT}. According to this strategy, the players coordinate on one particular stage-game strategy profile—let's say, s = (middle, center). According to the grim trigger strategy, each player will continued to play middle or center each period so long as the other does, but as soon as one player defects to some other action, from that time on the other player punishes the first relentlessly. For player 1 this means playing "bottom" every round, because then no matter what player 2 does, she will be held to a single period payoff of 1. For player 2 this means playing "right" every round, by the same reasoning.

Under what conditions would player 1 not want to deviate from s^{GT}? If she continues to play middle, her payoff stream will be 3, 3δ, $3\delta^2$, . . . etc.; if she deviates by playing top, her payoff stream will be 4, 1δ, $1\delta^2$, . . . etc. She will not have an incentive to deviate so long as

$$\sum_{t=0}^{\infty} 3\delta^t \geq 4 + \sum_{t=1}^{\infty} \delta^t$$

$$\frac{3}{1-\delta} \geq 4 + \frac{\delta}{1-\delta}$$

or $\delta \geq 1/3$. Things do not look much different from player 2's point of view.

It turns out that as $\delta \rightarrow 1$, any point in the more darkly shaded region of Figure C.2 labeled X can be a possible Nash equilibria with the help of the grim trigger strategy. In other words, there are literally thousands and thousands of possible equilibria. This result is known as the folk theorem (see Fudenberg and Tirole 1991, 150–60 for a more elaborate discussion). The folk theorem suggests that when it comes to real social conventions, equilibrium selection may represent a serious problem.