# Trust in Social Structures:

# Hobbes and Coase Meet Repeated Games

Robert Gibbons*

*MIT's Sloan School and NBER*

*rgibbons@mit.edu*

March 15, 2000

Prepared for K. Cook (ed.), *Trust in Society*, New York: Russell Sage Foundation

* My exposure to many of these issues began in an endogenous discussion group on "Power and Politics in Governance Structures" at the Center for Advanced Study in the Behavioral Sciences, where my Fellowship was supported by NSF grant SBR-9022192 and Cornell's Johnson School. Neil Fligstein, Tom Romer, and Neil Smelser are especially guilty of broadening my horizons during that year. Since then, I have pursued these issues in work with George Baker and Kevin J. Murphy and with Andy Rutten; anything coherent I say here has benefited greatly from long discussion with them.

# Trust in Social Structures:

# Hobbes and Coase Meet Repeated Games

by

Robert Gibbons

In one of the classic passages in social science, Hobbes argued that without a state, life would be "solitary, poor, nasty, brutish and short," so people should create a state powerful enough to punish malefactors. Three centuries later, Coase (1937) made a similar claim, arguing that where the price system would perform sufficiently poorly, transactions should be conducted inside firms, so that bosses (rather than prices) could direct workers' actions. Both Hobbes and Coase were advocating horse races: comparative analyses of alternative institutions in a given environment.

For Hobbes, the environment was the "state of nature," now often modeled as a group of people playing two-person prisoners' dilemmas with each other. Although Hobbes argued that Leviathan is always the best institution for this environment, others have since suggested that anarchy (*i.e.*, the absence of a ruler) need not produce the "war of all against all." Instead, institutions other than a state might produce "order without law" in Hobbes's environment, and such institutions might be a preferable to a state.

For Coase, the environment was an economic transaction, between buyer and seller. Unlike Hobbes, Coase made a contingent prescription: markets as optimal governance structures for some transactions but firms for others. Coase's market was the anonymous price system of neoclassical economics; others have since suggested that different institutions such as "networks" and "relational contracts" are feasible in Coase's environment and might be preferable to both the anonymous price system and firms.

There are large literatures devoted to each of these institutions—states, firms, networks, and so on. There are also small literatures focused on the Hobbesian and Coasean horse races—explicit comparisons of two or more institutions in a given environment. Recently, repeated-game models have emerged in all these literatures, often with the observation that repeated games allow economists (and methodological fellow-travelers) to analyze self-enforcing institutions. In this essay I first introduce the theory of repeated games and then sketch some of these recent repeated-game models.

I am not the first to cast Hobbes and Coase in something like these terms. For example, Granovetter (1985: 490) offered two main reasons why social relations may discourage malfeasance: "individuals with whom one has a continuing relation have an economic motivation to be trustworthy, so as not to discourage future transactions; and departing from pure economic motives, continuing economic relations often become overlaid with social content that carries strong expectations of trust and abstention from opportunism." I focus only on "pure economic motives," but I am evidently not alone in thinking that such motives are at least part of the story. Indeed, such calculative accounts of trust appear frequently in this book, starting with the opening paragraph of Hardin's (2000) opening chapter, on *The Brothers Karamazov*.

In fact, Granovetter's paper closely parallels this essay. He correctly criticizes Williamson's (1975) conception of a market for being "undersocialized" ( atomized and anonymous, and therefore without any prospect of supporting cooperation through social relations) and Williamson's conception of a hierarchy for being "oversocialized" (the hierarchical superior settles disputes by fiat but never abuses his powers). But economics has made some progress since 1975, and in Granovetter's direction: the more recent repeated-game models summarized in Section 2 begin to explore both social relations in markets and abuse of power in firms. Granovetter also makes analogous criticisms of Hobbes, noting the "undersocialized" nature of his conception of anarchy and the "oversocialized" nature of his conception of Leviathan. But the recent repeated-game models summarized in Section 3 again make progress towards embedding trust and social relations in models of anarchies and states.

## 1. An Introduction to Repeated Games[1]

Game theory is rampant in economics. Having long ago invaded industrial organization, game-theoretic modeling is now commonplace in international, labor, macro, and public finance, and is gathering steam in development and economic history. Nor is economics alone: accounting, finance, law, marketing, political science, and sociology are beginning similar experiences.

Why is this? Broadly speaking, two views are possible: fads and fundamentals. While I believe that fads are partly to blame for the current enthusiasm for game theory, I

[1] This section draws heavily on Gibbons (1997).

also believe that fundamentals are an important part of the story. Simply put, many modelers use game theory because it allows them to think like economists when price theory does not apply. Examples abound: small numbers, hidden information, hidden actions, and incomplete contracts can turn markets into games; in other settings, markets are at most peripheral—such as the relationship between a regulator and a firm, a boss and a worker, and so on. Thus, where markets have become games, and where transactions do not occur in markets, game theory allows economists to study the implications of rationality, self-interest, and equilibrium when price theory would not.

When people interact over time, threats and promises concerning future behavior may influence current behavior. Repeated games capture this fact of life, and hence have been applied more broadly than any other game-theoretic model (by my armchair citation count)—not only in virtually every field of economics but also in finance, law, marketing, political science, and sociology. In this section I describe first a one-shot interaction between two parties (which works out badly) and then an ongoing sequence of such interactions (which work out well because of the parties' concerns for their reputations).

*The One-Shot Interaction*

Suppose that late last night an exciting new project occurred to you. The project would be highly profitable, but is outside your area of expertise, so you would need help in completing it. Furthermore, it would take significant work on your part just to explain the project to someone with the needed expertise. Finally, if you did explain the project to the relevant other, that person could steal your ideas, representing them as substantially his own.

It is not hard to imagine this scenario unfolding in an organization: you work in marketing, and the project is a new product, but you need assistance from someone in engineering, who could later take all (or at least too much of) the credit. To decide whether to pursue the project, it would help to know something about the "trustworthiness" of a particular engineer you could approach. But if you have been buried deep inside marketing, you may not have much information about any of the relevant engineers. In this case, you would be forced to rely either on the average sense of human decency among engineers or on your organization's culture: "how we do things around here" (Peters, 19xx: yy). If the culture emphasizes teamwork over individual

accomplishments, for example, you may have more confidence in approaching an unfamiliar member of the engineering group.

Kreps (1990) captures these issues in the *Trust Game* shown below. (See also Miller's (2000) discussion of this game in the context of Lincoln Electric.) The game begins with a decision node for player 1, who can choose either to Trust or Not Trust player 2. If player 1 chooses Trust then the game reaches a decision node for player 2, who can choose either to Honor or Betray player 1's Trust. If player 1 chooses Not Trust then the game ends—effectively, 1 terminates the relationship. At the end of each branch of the game tree, player 1's payoff appears above player 2's. If player 1 chooses to end the relationship then both players' payoffs are zero. If 1 chooses to trust 2, however, then both players' payoffs are one if 2 honors 1's trust, but player 1 receives -1 and player 2 receives two if player 2 betrays 1's trust.

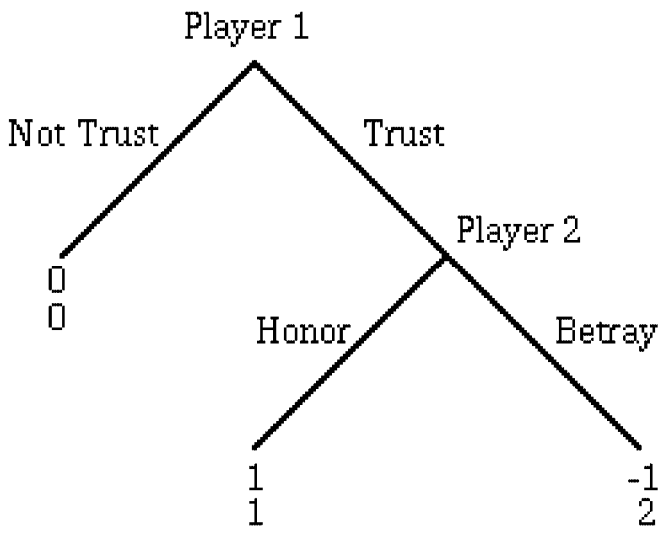


*Figure 1: The Trust Game*

We solve the Trust Game by backwards induction—that is, by working backwards through the game tree, one node at a time. If player 2 gets to move (*i.e.*, if player 1 chooses to trust player 2) then 2 can receive either a payoff of one by honoring 1's trust or a payoff of two by betraying 1's trust. Since two exceeds one, player 2 will betray 1's trust if given the move. Knowing this, player 1's initial choice amounts to either ending the relationship (and so receiving a payoff of zero) or trusting player 2 (and so receiving a payoff of -1, after player 2 betrays 1's trust). Since zero exceeds -1, player 1 should end the relationship.

*The Repeated Game*

Instead of a one-shot interaction, suppose that you and a *particular* engineer will play the Trust Game repeatedly, with all previous outcomes observed by both players before the next period's Trust Game is played. The analysis of this repeated game differs dramatically from the one-shot interaction: the engineer's actions today may affect your expectation of her actions tomorrow, which may affect your actions tomorrow, which affect her payoffs tomorrow. Thus, actions not in the engineer's short-run self-interest (as defined by today's payoffs) may be consistent with her overall self-interest (as defined by the total payoff over time).

I do not mean to imply that this logic is surprising or complicated. To the contrary, I think it is close to ubiquitous. (As Hardin notes, Dostoyevsky clearly understood it.) In this sub-section I describe the simplest possible formalization of this logic. Formally, we will analyze an infinitely repeated game: the game never ends, but both players face an interest rate r per period in discounting their payoffs across periods. (For example, when r is high, a dollar to be received next period is not worth much today—$1/(1+r), to be exact.) We can interpret this "infinitely" repeated game somewhat more realistically by saying that the game ends at a random date. Under this interpretation, the interest rate r reflects not only the time value of money but also the probability that the players will meet again after the current period. (A dollar to be received next period provided that we are still interacting is not worth much if today's interaction is likely to be our last.) Under either interpretation, the present value of $1 to be received every period starting tomorrow can be shown to be $1/r.

Mostly for analytical simplicity (but to some extent for behavioral realism), we will consider the following "trigger" strategies in the infinitely repeated game:

> Player 1: In the first period, play Trust. Thereafter, if all moves in all previous periods have been Trust and Honor, play Trust; otherwise, play Not Trust.
>
> Player 2: If given the move this period, play Honor if all moves in all previous periods have been Trust and Honor; otherwise, play Betray.

These strategies are not forgiving, like Tit-for-Tat. Rather, under the trigger strategies, if cooperation breaks down at any point then it is finished for the rest of the game, replaced by the dictates of short-run self-interest. In most games, reverting to short-run self-interest after a breakdown in cooperation is a middle ground between two plausible alternatives:

forgiveness (*i.e.*, an attempt to resuscitate cooperation) and spite (*i.e.*, going against short-run self-interest in order to punish the other player). Both forgiveness and spite deserve analytical attention, but I will focus on the trigger strategies (with their reversion to short-run self-interest after a breakdown of cooperation) as a tractable compromise.[2]

We will analyze whether these trigger strategies are an equilibrium of the infinitely repeated game. That is, given that player 1 is playing her trigger strategy, is it in player 2's interest to play his? We will see that the trigger strategies are an equilibrium of the infinitely repeated game provided that player 2 is sufficiently patient (*i.e.*, provided that the interest rate r is sufficiently small).

Suppose that player 1 follows his trigger strategy and chooses Trust in the first period. Player 2 then faces a dilemma. As in the one-shot interaction, player 2's one-period payoff is maximized by choosing to Betray. But in the repeated game, if player 1 is playing the trigger strategy then such a betrayal by player 2 leads player 1 to choose No Trust forever after, producing a payoff of zero for player 2 in each subsequent period. Thus, the key question is how player 2 trades off the short-run temptation (a payoff of 2 instead of 1 now) against the long-run cost (a payoff of 0 instead of 1 forever after). The answer depends on the interest rate: if r is sufficiently low then the long-run consideration dominates and player 2 prefers to forego the short-run temptation.

The general point is that cooperation is prone to defection (otherwise we should call cooperation something else—such as a happy alignment of the players' self-interests), but in some circumstances defection can be met with punishment. A potential defector therefore must weigh the present value of continued cooperation against the short-term gain from defection followed by the long-term loss from punishment. If a player's payoffs (per period) are C from cooperation, D from defection, and P from punishment (where D > C > P) then this decision amounts to evaluating two time-paths of payoffs: (C, C, C, ...) versus (D, P, P, P, ...), as shown in Figure 2.

[2] In the Trust Game, unlike most, reverting to short-run self-interest is identical to spite: it achieves the harshest possible punishment of player 2.

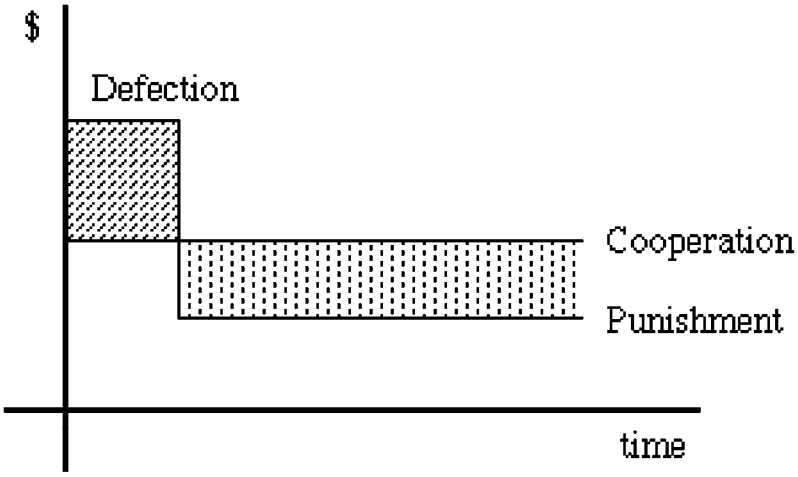


*Figure 2*

Because the present value of $1 received every period starting tomorrow is $1/r, the time-path of payoffs (C, C, C, ...) yields a higher present value than the time-path (D, P, P, P, ...) if

$$(*) \qquad \left\{1 + \frac{1}{r}\right\} C > D + \frac{1}{r} P \ .$$

Rearranging the inequality (*) yields r < (C - P)/(D - C). In the repeated-games literature, this result is often restated as follows: if the players are sufficiently patient (*i.e.*, if r is sufficiently close to zero) then it is optimal to cooperate, foregoing the short-run temptation (D – C now) for the long-term gain (C - P forever after). For purposes of this essay, however, it is more useful to recall that the interest rate r reflects not only the time value of money but also the probability that the players will meet again after the current period. If this probability is high then r is low. Thus, the result that cooperation is optimal if r is sufficiently close to zero can be interpreted in terms of something like social structure: provided the time value of money is not too high, cooperation is optimal today if the players' relationship is sufficiently likely to continue in the future.

The model described here is excessively tidy: cooperation either works perfectly or doesn't work at all, depending on the interest rate. It is natural to ask what happens when the players are not "sufficiently patient." In brief, all is not lost, because it may be possible to achieve partial rather than full cooperation.[3] It is also natural to ask why there

[3] To see how this may work, examine (*). Note that reducing the payoff from cooperation from C to some lower level is no help at all, in and of itself. That is, holding the payoff from defection (D) constant, reducing C makes it harder to satisfy (*). The trick is that reducing C may also reduce D: making due

are never any fights or misunderstandings in this equilibrium. Green and Porter (1984) develop a model in which the players' actions are not perfectly observable and cooperation breaks down periodically (but for a finite time, after which it begins again). Adding such imperfect observability, and its resulting temporary breakdowns of cooperation, would be a step towards realism in many of the settings considered in the next two sections.

It is also natural to ask whether cooperation is the inevitable outcome of this repeated-game, even when the players are sufficiently patient. The answer is that it is not: there are many, many other equilibria of the repeated game. Miller (2000) sketches the range of such equilibria, from perfect cooperation to its polar opposite (perpetual defection); Miller (1992) says more about this plethora of equilibria and, importantly, about the potential role of leadership in creating and sustaining equilibria in repeated games. These latter issues — creating and sustaining equilibria — are at the forefront of research on repeated games and deserve substantial attention in the near future. But my guess is that such future research will enrich the current theory, not overturn its basic insights such as Figure 2 and condition (*). I therefore turn next to applications of the existing theory of repeated games to the problems posed by Coase and Hobbes.

## 2. Coase (and Williamson) Meet Repeated Games[4]

Coase argued that firms would not need to exist if markets were perfect. Williamson (1975) made important progress towards explaining why markets might perform poorly—roughly, because formal contracts are typically incomplete. And as a second prong of his argument, Williamson also suggested why firms might perform better—roughly, because firms might use relational contracts (defined below), as envisioned in Simon's (1951) theory of the employment relationship.

The second prong of Williamson's argument is clearly correct: firms are riddled with relational contracts—informal agreements, unwritten codes of conduct, and norms that powerfully affect the behaviors of individuals in firms. Virtually every collegial and hierarchical relationship in organizations involves important relational contracts,

with partial cooperation may also limit the players' opportunities for profitable deviations. If D falls more than C then (*) may hold.

[4] This section draws heavily on Baker, Gibbons, and Murphy (1999).

including informal *quid pro quos* between co-workers and unwritten understandings between bosses and subordinates about task-assignment, promotion, and termination decisions. Many earlier observers also emphasized the importance of such informal agreements in organizations.[5] Even ostensibly formal processes such as compensation, transfer pricing, internal auditing, and capital budgeting often cannot be understood without consideration of their associated informal agreements.[6]

But business relationships are also riddled with relational contracts. Many transactions do not occur in a pure spot market between buyers and sellers who pass (goods) in the night. Instead, supply chains often involve long-run, hand-in-glove supplier relationships through which the parties reach accommodations when unforeseen or uncontracted-for events occur.[7] Similar relationships also exist horizontally, as in the networks of firms in the fashion industry or the diamond trade.[8] Whether vertical or horizontal, these relational contracts influence the behaviors of parties across firm boundaries. Thus, the first prong of Williamson's (1975) argument seems incomplete: formal contracts may well be incomplete, but relational contracts seem roughly as important between firms as within.

Baker, Gibbons, and Murphy (1999) revisit the Coase-Williamson comparison of markets versus firms, taking into account the ubiquity of relational contracts in both domains. Both within and between organizations, relational contracts can help circumvent difficulties in formal contracting. For example, a formal contract must be specified *ex ante* in terms that can be verified *ex post* by a third party (such as a court), while a

---

[5] See Barnard (1938) and Simon (1947), as well as the case studies by Blau (1955), Gouldner (1954), and Selznick (1949) that inspired American sociology's departure from Weber's emphasis on formal organizational structures and processes. Granovetter (1985: 502) offers a more recent assessment: "The distinction between the 'formal' and the 'informal' organization of the firm is one of the oldest in the literature, and it hardly needs repeating that observers who assume firms to be structured in fact by the official organization chart are sociological babes in the woods."

[6] See Lawler (1990) on compensation, Eccles (1985) on transfer pricing, Dalton (1959) on internal auditing, and Bower (1970) on capital budgeting. See also Blumenstein and Stern (1996) on how the 1700-page contract between General Motors and the United Auto Workers has important gaps that are covered by informal agreements.

[7] Macaulay (1963) emphasized the importance of such "non-contractual relations" in various businesses. Kogut, Shan, and Walker (1992) suggest the prominence of such relationships by relabeling the make-or-buy decision as "The Make-or-Cooperate Decision." Eccles (1981) describes "quasifirms" in the construction industry—long-run relationships between general contractors and independent, specialized subcontractors.

[8] In "Neither Market Nor Hierarchy: Network Forms of Organization," Powell (1990) describes a variety of other examples and emphasizes their differences from spot markets and firms. See Podolny and Page (1998) for a summary and critique of this growing literature.

relational contract can be based on outcomes that cannot be verified *ex post*, and also on outcomes that are prohibitively costly to specify *ex ante*. A relational contract thus allows the parties to utilize their detailed knowledge of their specific situation and to adapt to new information as it becomes available. For the same reasons, however, these relational contracts cannot be enforced by a third party. Instead, relational contracts must be designed to be self-enforcing: each parties' reputation must be sufficiently valuable, relative to the payoffs from reneging on the relational contract, that neither party wishes to lose his reputation by reneging.

What I have called a relational contract seems similar to Heimer's (2000) "trust relation," with its emphasis on "uncertainty and vulnerability." Following Simon (1951) and Williamson (1975), I have tried to emphasize the important role of uncertainty in making relational contracts potentially valuable: in the presence of uncertainty, the parties want to be able to respond to events after they occur. Furthermore, the (necessarily) self-enforcing nature of relational contracts emphasizes at least one kind of vulnerability: even if condition (*) from Section 2 holds (where C is the payoff from abiding by the relational contract, D is the payoff from reneging on the relational contract, and P is the payoff after losing one's reputation for abiding by relational contracts), the prospect that one party could renege on the relational contract makes the other party vulnerable.

Baker, Murphy, and I develop a model in which an upstream party uses an asset to produce a good that can be used in a downstream party's production process. We follow Grossman and Hart's (1986) terminology: when the upstream party owns the asset we call the transaction non-integrated—the upstream party is an independent contractor, working with an asset she owns; when the downstream party owns the asset we call the transaction integrated—the upstream party is an employee, working with an asset owned by the firm. We assume that ownership of the asset conveys ownership of the good. (In fact, the asset could simply be the legal title to the good.) Thus, if the upstream party owns the asset then the downstream party cannot use the good without buying it from the upstream party, whereas if the downstream party owns the asset then he already owns the good.

Under either ownership structure, the downstream party would like the upstream party to take actions that improve the value of the good in the downstream production process. Relational contracts can encourage these actions: for example, the downstream party can promise to pay the upstream party a bonus if she produces a good of high (but non-contractible) value. Because this promise is based on observable but non-contractible

outcomes, it provides incentives to the upstream party only if it is self-enforcing (*i.e.*, the short-run value of reneging must be less than the long-run value of the relationship).

The key question in our analysis is whether a particular promise is more likely to be self-enforcing under integration or non-integration. Under vertical integration, if the downstream party reneges on the bonus, he still owns the good. But under non-integration, if the downstream party reneges on the bonus, he cannot use the good without buying it. In this sense, non-integration gives the upstream party more recourse should the downstream party renege on the promised bonus. But non-integration has a drawback: it creates an incentive for the upstream party to take actions that improve her bargaining position with the downstream party. These two conflicting effects (recourse and bargaining position) produce our main result: integration affects the parties' temptations to renege on a relational contract, and hence affects the best relational contract the parties can sustain. That is, a major factor in the vertical-integration decision is whether integration or non-integration facilitates the superior relational contract. In short, integration is an instrument in the service of the parties' relationship.

So far, I hope to have motivated three assertions: (1) there are many relational contracts that have important influences on behavior, both within and between firms; (2) repeated-game models can capture some of the importance features and tensions of these relational contracts; and (3) the economic incentive to renege on a given relational contract depends on whether the parties are integrated or not. The last of these assertions begins to suggest how social structure affects relational contracts, but we can say more. As in Section 1, social structure enters the model through the interest rate r, which reflects not only the time value of money but also the probability that the parties' relationship will continue in the future. One can imagine two polar cases: a "spot" social structure, in which the parties will never see each other again after their current interaction, and a "relational" social structure, in which the parties will continue to interact for the foreseeable future.

These two polar social structures (spot and relational) can be combined with the two possibilities for asset ownership described above (integrated or not), producing the four governance regimes summarized in Figure 3. Consistent with common usage, one can call the integrated case "employment" (where the upstream party is an employee, working with an asset owned by the firm) and the non-integrated case "outsourcing" (where the upstream party is an independent contractor, working with an asset she owns). The four

governance regimes are then spot and relational outsourcing and spot and relational employment.

**Figure 3**

Combinations of social structure and asset ownership that define four governance regimes: Spot Outsourcing, Spot Employment, Relational Outsourcing, and Relational Employment

| | Asset Ownership | |
|---|---|---|
| Social Structure | Non-Integrated | Integrated |
| Spot | Spot Outsourcing | Spot Employment |
| Relational | Relational Outsourcing | Relational Employment |

Baker, Murphy, and I analyze the efficiency of each of these four governance regimes. In fact, we consider all possible values of the interest rate r, and hence consider all the social structures between the spot and relational polar cases defined above. For any given value of r, we compute whether integration or non-integration is more efficient. That is, for any given probability that the parties' relationship will continue in the future, we conduct a formal version of the Coase-Williamson horse race between markets and firms, taking relational contracts into account in both domains.

Many of the classic contributions to organizational economics can also be described using Figure 3. For example, static analyses of integration in the absence of relational contracting (*e.g.*, Grossman and Hart, 1986) are analogous to our comparison of spot outsourcing to spot employment (the top row in the figure). Similarly, repeated-game analyses of relationships within firms (*e.g.*, Kreps, 1990) are analogous to our comparison of spot employment to relational employment (the right column) and repeated-game analyses of relationships between firms (*e.g.*, Klein and Leffler, 1981) are analogous to our comparison of spot outsourcing to relational outsourcing (the left column). Finally, Williamson's (1975) argument that the comparative advantage of firms over markets lies in the firm's ability to enforce relational contracts is analogous to our comparison of spot outsourcing to relational employment (the main diagonal) and Williamson's (1996, Chapter 4) argument that important relational contracts exist between firms as well as

within is analogous to our comparison of relational outsourcing to relational employment (the bottom row).[9]

I now hope to have motivated a fourth assertion, in addition to the three above: repeated-game models can be used to run the Coase-Williamson horse race between markets and firms, taking relational contracts into account in both domains. The result is a new perspective on vertical integration, in which the "make or buy" decision is seen as an instrument in the service of the parties' relationship. Baker, Murphy, and I also suggest how this approach can be enriched in future work to analyze specific relational contracts within organizations (such as transfer pricing and capital allocation) as well as the roles of relational contracts in specific multi-organizational forms (such as networks and joint ventures). In short, I think one can see on the horizon an economic theory of trust and social relations within and between firms.

## 3. Hobbes (and Axelrod) Meet Repeated Games[10]

The problem of social order—of creating and maintaining the security of persons and property—is central to political economy. Without social order, people will find themselves in the Hobbesian "war of all against all," where life is "solitary, poor, nasty, brutish and short." With social order—that is, when rights are enforced—people can use markets, contracts, and other economic institutions to reap gains from cooperation. Thus, the provision of social order is crucial for prosperity and growth.

Recently, economists and others have used the theory of repeated games to analyze institutions that support social order. This emerging literature revisits Plato's question, "Who will guard the guardians?" The answer is that, in a world of self-interest, there is no disinterested third party to enforce society's rules, so some of those rules must be self-

[9] Williamson (1985) also emphasized relational contracts between firms, but construed them as lying on a continuum between markets and hierarchies (p. 83). Figure 3 suggests that the set of alternative governance regimes is two-dimensional, so it is not possible to locate all governance structures on a line between markets and hierarchies. Compared to Williamson's first book, the second showed much greater concern for limits on the effectiveness of firms and the importance of relational contracts between firms. But there was little analysis of how relational contracts between firms might differ from those within, and discussion of alternative organizational forms was limited to brief mention of "hybrid transactions (franchising, joint ventures, and other forms of nonstandard contracting)" (p. 83).

[10] This section draws heavily on Gibbons and Rutten (1999). Any statements that reveal more than passing acquaintance with the political-economy literature were greatly influenced by Andy Rutten, while any gross errors are surely mine.

enforcing. Put more formally, institutions—that is, the formal and informal "rules of the game" (North, 1990:3)—must be equilibria: they must provide everyone, including enforcers, with an incentive to behave appropriately. Using this approach, Greif (1989, 1993, 1994), Grief, Milgrom, and Weingast (1994), and Milgrom, North, and Weingast (1990) have examined the role of private judges, merchant guilds and other institutions in the revival of long-distance trade in medieval Europe. Similarly, Ellickson (1991), Libecap (1989), and Ostrom (1990) have examined the role of institutions in resolving contemporary commons problems in cattle ranching, oil-drilling, fishing and related industries.

Ironically, these equilibrium institutionalists have so far neglected the most prominent source of social order in the modern world: the state. This neglect stands in stark contrast to the large literature in political economy that takes the state as its central focus. Beginning with Hobbes, Locke, and Madison, this classic line of inquiry has recently developed into three distinct approaches to modeling the state: the rational-predator school, the constitutional economists, and the neo-Hobbesians. While each of these schools has made important progress, each reveals a weakness in the others, as follows.

The rational-predator school of Barzel (1992), Bates (1981), Levi (1988), North (1981) and Olson (1993) treats the ruler as a self-interested sovereign with a monopoly on violence, but argues that such a ruler will typically not be a pure predator: she will find it optimal to provide some services to her subjects. Their reasoning is simple (and akin to the logic of the Laffer curve). For the ruler, subjects are capital goods that provide a flow of income over time. Under a wide range of circumstances, the ruler can expand her tax base by protecting her subjects from each other *and from herself*. Thus, the pursuit of self-interest may induce the ruler to provide a legal system or even representative government.

The rational-predator school shows that even the most powerful predator may want to provide services like those envisioned in the Hobbesian conception of a social contract. In reaching this conclusion, however, these scholars ignore another aspect of the social contract—the constraints it places on the ruler. This omission is remedied by a second approach to models of limited government, that of the constitutional economists. (See Brennan and Buchanan (1985), Buchanan and Bush (1974), and Buchanan and Tullock (1962), for example.) These theorists analyze the effect of constitutional rules on an otherwise self-interested ruler. That is, they model constitutional rules as shaping

politicians' incentives and opportunities, and then ask how political decisions will vary with changes in those rules. Because they are interested in policy choice, their models simply assume the *ex post* enforcement of the rules—for the constitutional economists, the social contract is an exogenously imposed constraint.

A third approach—that of the neo-Hobbesians such as Hampton (1986), Hardin (1990), Heckathorn and Maser (1987)—endogenizes the social contract by modeling its choice. They begin with citizens in a state of nature (sometimes modeled as a prisoners' dilemma). Like Hobbes, they then argue that, to avoid the inefficiency of mutual defection, people come together to create a government to enforce private contracts. The neo-Hobbesians model the process of creating a government as a coordination game, in which citizens can escape the state of nature only if they agree on one of many possible social contracts. In these models, different contracts distribute the benefits of society differently, but all Pareto-dominate the state of nature. From this Pareto-dominance, the neo-Hobbesians conclude that once the citizens have made their choice, the social contract will be self-enforcing. That is, they ignore the possibility that the ruler's misbehavior can destroy social order. Indeed (and ironically, given their debt to Hobbes), their models include no ruler.

In the rest of this section I first sketch some of the recent repeated-game models of self-enforcing institutions other than the state. I then turn to an analogous model of the state, Gibbons and Rutten (1999), that attempts to integrate the strengths of the three prominent theories of the state just summarized. But in this section I discuss only individual horses; as far as I know, a formal analysis of the Hobbesian horse race between the state and anarchy does not yet exist.

### *Anarchy: Order without Law?*

Following Calvert (1994) and Hobbes, suppose that the state of nature is a group of N people playing two-person prisoners' dilemma with each other. In anticipation of the eventual introduction of a state into the analysis, call the players citizens. Each period, each citizen plays a prisoners' dilemma (PD). There is a matching rule specifying who plays whom each period; for simplicity, suppose it is random.[11] In equilibrium, two

[11] Note that random matching every period in a large population is very different from Axelrod's (1984) tournaments, in which two players play each other repeatedly over a long horizon. As a result, the institutions described below that support cooperation in large, decentralized populations are often very different from the Tit-for-Tat strategies Axelrod emphasizes.

citizens playing a PD one time (and then vanishing from the population) would both defect.

This modeling strategy is very abstract, but it is also widely accepted because it captures an important feature of the games of daily life: people are engaged in transactions that they would like to govern by a contract specifying that both Cooperate.[12] Following Hobbes, the role of the state in this environment will be to enforce such contracts. But as has been noted with increasing frequency in economics, law, political science, and sociology, citizens need not suffer the war of all against all, even if no Hobbesian Leviathan exists to punish Defectors. That is, anarchy (literally, the absence of a ruler) does not preclude Cooperation.[13]

Perhaps the simplest mechanism for sustaining Cooperation under anarchy is bilateral reputation. That is, citizen n keeps a record of her past dealings with citizen m, and if n is ever paired with m again, she consults this record and plays as dictated by some strategy (perhaps Tit-for-Tat), given the recorded history between n and m. In this case, the N-citizen dynamic game can be thought of as a collection of two-player repeated games, although meetings between the two players do not happen every period. If the population is large, however, then the expected time between meetings of a given pair of citizens may be quite long, in which case bilateral-reputation incentives cannot be strong.

More optimistically, one might suppose that all information about other citizens' experiences is publicly available, so that citizen n would know the full history of citizen m's play (with n and with all others) before playing m in a given period. In this public-information case, the N-citizen dynamic game can be thought of as a two-player repeated game, where meetings between the two players occur every period. Cooperation can be sustained in equilibrium if the players are sufficiently patient (*i.e.*, the interest rate r is sufficiently small, as derived in Section 1), but remember that this simple analysis applies only in the unlikely case in which all information about all other citizens' experiences is publicly available.

Rather than explore the implications of exogenous information structures, one can endogenize the flow of information that sustains reputation mechanisms. Such analyses

---

[12] See Taylor (1976), Kavka (1986), and Hampton (1986) for discussions of the extent to which Hobbes's state of nature is well modeled by a population playing two-person prisoners' dilemmas.

[13] Hence Taylor's (1976) title, "Anarchy and Cooperation" and Hirshleifer's (1995) "Anarchy and Its Breakdown."

begin from an exogenous but impoverished information structure (such as the own-history information structure from the bilateral-reputation case), then define the opportunities for and costs of communication, and finally determine whether a richer information structure (such as the extreme case of public information) can be sustained in equilibrium, together with a level of Cooperation consistent with that richer information structure. Kandori (1992) and Calvert (1994) provide two such models in abstract settings; the papers cited earlier by (various combinations of) Greif, Milgrom, North, and Weingast build similar applied models to interpret evidence on the development of trade since the 11th century.

Equilibrium information structures, together with the Cooperative behavior they sustain (and threats of punishment following Defections), can be interpreted as self-enforcing social institutions: endogenous rules of the game. A complete analysis must resolve several questions, however, including "Why the institution was needed, what sanctions were to be used to deter undesirable behavior, who was to apply the sanctions, how the sanctioners learned or decided what sanctions to apply, why they did not shirk from their duty, and why the offender did not flee to avoid the sanction" (Greif, Milgrom, and Weingast, 1994: 746). It is important to note that in these equilibrium models two behaviors are endogenous: not only interaction in the PDs but also communication. Thus, all the issues related to sanctioning apply not only to departures from prescribed interaction but also to departures from prescribed communication.

As the applied models of Grief, Milgrom, North, and Weingast illustrate, this equilibrium approach to communication and interaction can be used to interpret real institutions. Furthermore, both decentralized and centralized institutions are within the scope of the theory: Calvert, for example, compares a decentralized mechanism (in which everyone talks to everyone else every period) to a centralized mechanism (in which everyone reports to a center at the end of one period and then asks the center about the next partner's history before play begins in the next period). Unless communication costs are assumed to be zero, however, all these institutions are costly to operate. Furthermore, these costs grow as the size of the population increases. Such communication costs motivate the search for alternative institutions: rules of the game that produce social order but avoid the costs of communication (and the costs of inducing communication) that are an inevitable part of the reputation mechanisms that promise order without law. One possibility is a state.

*Hierarchy: Law without Order?*

Gibbons and Rutten (1999) develop a repeated-game model of the state that parallels the equilibrium analyses of social institutions that produce Cooperation under anarchy. The main points are as follows. First, the state is modeled as an equilibrium institution. Thus, the state's actions in the model (enforcement of contracts between citizens, limited taxation, and provision of a public good) must be in its long-run self-interest.

Second, there are many equilibria. In our basic model, these equilibria range from a maximum tax rate (above which the citizens prefer anarchy and so will revolt) to a minimum tax rate (below which it is in the state's interest to confiscate all the citizens' wealth in one period, even if this means losing power forever after). The problem of selecting one of these equilibria is precisely the coordination game emphasized by the neo-Hobbesians.

Third, the payoffs in the state's "game of politics" with its citizens are derived from the underlying "games of daily life" among citizens (namely, the same two-person prisoners' dilemmas just studied under anarchy). This embedding of the game of politics in the games of daily life exposes a connection between the social benefits from the state's provision of social order and the state's incentive to abuse its power. Furthermore, rooting the analysis of the state in the same environment used to analyze anarchy is a necessary condition for a subsequent formal model of the Hobbesian horse race between state and anarchy.

Fourth, after developing our basic model of the state as the provider of social order, we enrich the model to allow a second role for the state: providing a public good (as suggested by Hume, Rousseau and others) as well as social order. We find that having a single actor provide both social order and public goods may create economies of scope: the state can provide public goods more efficiently if it also is charged with providing social order.[14]

Finally, in every equilibrium, the state is both a contractor (providing social order, as envisioned by Hobbes) and a predator (extracting resources, as feared by Locke). At one extreme (maximal predation), the ruler takes most of the gains from trade; as long as

[14] The interaction between these two roles of the state in our model may be a cousin of another interaction suggested by Hardin (1990), in which a leader first emerges to solve a coordination problem, then finds power in such a leadership role, and finally uses this power to solve social dilemmas.

she leaves the citizens better off than in the state of nature, they will not rebel. At the other extreme (maximal contracting), the citizens keep most of the gains; as long as they do not tempt the ruler too greatly, she will not abuse her power. In all the intermediate equilibria, the state exhibits a mix of predation and contracting. Indeed, even maximal predation includes some contracting, and maximal contracting some predation.

This "predatory contractor" model of the state meshes comfortably with the three prominent theories of the state summarized earlier. Like the rational-predator school, we let our ruler extract resources from her subjects; unlike them, we assume that the ruler has only a temporary monopoly over violence, which the citizens may subsequently revoke. Like the constitutional economists, we treat the social contract as a constraint on self-interested politicians; unlike them, we require that the social contract be self-enforcing. Like the neo-Hobbesians, we treat the social contract as a coordination game, distinct from the prisoners' dilemmas facing ordinary citizens; unlike them, we explicitly link the two—in our model, the payoffs from the social contract derive from the underlying games between the citizens.

In sum, Rutten and I offer a stick-figure rendering of the grand ideas of Hobbes and Locke. In the stationary and spare world we model, our simple state solves the classical problem of providing social order that inspired its creation. But real constitutions set only the basic rules of the political-economic game: they create other institutions (such as the legislature and the courts) that fill in some of the gaps in the social contract. We envision a rich class of equilibrium models of the state, corresponding to the variety of government structures we observe in the world. We hope our model lays the foundation for such a literature on the theory of the state, parallel to the emerging literature on order without law. Further development of these two literatures will make it possible to conduct a formal version of the Hobbesian horse race between state and anarchy, with both modeled as equilibrium institutions.

## 4. Conclusion

Repeated-game models focus on the role of long-run self-interest in overcoming short-term temptation. I have tried to argue that such models can shed light on relational contracts within and between firms and on the self-enforcing institutions of political economy.

According to the repeated-game approach, if you understand my long-run self-interest, you might "trust" me not to yield to certain short-run temptations. As Hardin (2000) documents, there has been much debate over whether this calculative conception of trust is all there is to that slippery term. I am inclined to believe that there is more to trust than calculation, but I have tried to argue here that calculative trust is important in the world and that repeated-game models can help us understand it.

But saying that social structure can help create trust leaves open the question of whether we can trust in social structure. That is, there is abundant evidence that relational contracts often hurt rather than improve organizational performance; see Roethlisberger and Dickson (1939), Roy (1952), and many others. As noted in Section 1, this indeterminacy appears in the economic model as the existence of multiple equilibria in the repeated game. Like Miller (1992), I believe that an economic theory of leadership can and should be developed around the idea that leaders try to create and change equilibria. Such a theory will grope towards ideas such as building and utilizing trust, but even sketching the outlines of such a theory must await another essay.

## References

Axelrod, R. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Baker, George, Robert Gibbons and Kevin J. Murphy. 1999. "Relational contracts and the Theory of the Firm." Unpublished manuscript, MIT's Sloan School.

Barnard, Chester. 1938. *The Functions of the Executive*. Cambridge, MA: Harvard University Press.

Barzel, Yoram. 1992. "Confiscation by the Ruler: The Rise and Fall of Jewish Lending in the Middle Ages." *Journal of Law and Economics* 35: 1-13.

Bates, R. 1981. *Markets and States In Tropical Africa*. Berkeley: University of California Press.

Blau, Peter. 1955. *The Dynamics of Bureaucracy*. Chicago: University of Chicago Press.

Blumenstein, Rebecca and Gabriella Stern. 1996. "UAW, Auto Makers Find Some Things Better Left Unsaid." *Wall Street Journal*, November 25, A1.

Bower, Joseph. 1970. *Managing the Resource Allocation Process*. Boston: Harvard Business School Press.

Brennan, G. and J. Buchanan, 1985. *The Reason of Rules: Constitutional Political Economy*. New York: Cambridge University Press.

Buchanan, J and W. Bush. 1974. "Political Constraints on Contractual Redistribution." *American Economic Review* 64: 153-157.

_____ and G. Tullock. 1962. *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Bull, Clive. 1987. "The Existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics* 102: 147-59.

Calvert, R. 1994. "Rational Actors, Equilibrium, and Social Institutions." In J. Knight and I. Sened (eds.), *Explaining Social Institutions*. Ann Arbor: University of Michigan Press.

Coase, Ronald. 1937. "The Nature of the Firm." *Economica* 4: 386-405.

Dalton, Melville. 1959. *Men Who Manage.* New York: Wiley.

Eccles, Robert. 1981. "The Quasifirm in the Construction Industry." *Journal of Economic Behavior and Organization* 2: 335-57.

_____. 1985. *The Transfer Pricing Problem.* D. C. Heath: Lexington, MA.

Ellickson, R. 1991. *Order without Law: How Neighbors Resolve Disputes*. Cambridge: Harvard University Press.

Gibbons, Robert. 1997. "An Introduction to Game Theory for Applied Economists." *Journal of Economic Perspectives* 11: 127-49.

_____ and Andrew Rutten. 1999. "Hierarchical Dilemmas: Social Order with Self-interested Rulers." Unpublished manuscript, MIT's Sloan School.

Gouldner, Alvin. 1954. *Patterns of Industrial Bureaucracy*. New York: The Free Press.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91: 481-510.

Green, Edward and Robert Porter. 1984. "Noncooperative Collusion Under Imperfect Price Information." *Econometrica* 52:87-100.

Greif, A. 1989. "Reputation and Coalitions in Medieval Trade: Evidence on the Maghribi Traders." *Journal of Economic History* 49: 857-82.

_____. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83:525-48.

_____. 1994. "Cultural Beliefs and the Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies." *Journal of Political Economy* 102: 912-50.

_____, P. Milgrom, and B. Weingast. 1994. "Coordination, Commitment, and Enforcement: The Case of the Merchant Guild." *Journal of Political Economy* 102:745-76.

Grossman, Sanford and Oliver Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Ownership." *Journal of Political Economy,* 94: 691-719.

Hampton, J. 1986. *Hobbes and the Social Contract Tradition*. New York: Cambridge University Press.

Hardin, R. 1990. "The Social Evolution of Cooperation," in K. Cook and M. Levi (eds.), *The Limits of Rationality*. Chicago: University of Chicago Press.

_____. 2000. "Conceptions and Explanations of Trust." This volume.

Heckathorn, D. and S. Maser. 1987. "Bargaining and Constitutional Contracts." *American Journal of Political Science* 31: 142-168.

Heimer, Carol. 2000. "Solving the Problem of Trust." This volume.

Hirshleifer, Jack. 1995. "Anarchy and Its Breakdown." *Journal of Political Economy* 103:26-52.

Kandori, M. 1992. "Social Norms and Community Enforcement." *Review of Economic Studies* 59:63-80.

Kavka, G. 1986. *Hobbesian Moral and Political Theory*. Princeton: Princeton University Press.

Klein, Benjamin and Keith Leffler. 1981. "The Role of Market Forces in Assuring Contractual Performance." *Journal of Political Economy* 89:615-41.

Kogut, Bruce, Weigian Shan, and Gordon Walker. 1992. "The Make-or-Cooperate Decision in the Context of an Industry Network." In N. Nohria and R. Eccles (eds.), *Networks and Organizations: Structure, Form, and Action*. Boston: Harvard Business School Press.

Kreps, David. 1990. "Corporate Culture and Economic Theory." In J. Alt and K. Shepsle, eds. *Perspectives on Positive Political Economy*. Cambridge University Press.

Lawler, Edward. 1990. *Strategic Pay*. Jossey-Bass: San Francisco.

Levi, M. 1988. *Of Rule and Revenue*. Berkeley: University of California Press.

Libecap, G. 1989. *Contracting for Property Rights*. New York: Cambridge University Press.

Macaulay, Stewart. 1963. "Non Contractual Relations in Business: A Preliminary Study." *American Sociological Review* 28: 55-67.

Milgrom, P., D. North, and B. 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.

Miller, G. 1992. *Managerial Dilemmas: The Political Economy of Hierarchy*. New York: Cambridge University Press.

_____. 2000. "Why Is Trust Necessary in Organizations? The Moral Hazard of Profit-Maximization." This volume.

North, D. 1981. *Structure and Change In Economic History*. New York: Norton.

_____. 1990. *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.

Olson, M. 1993. "Dictatorship, Democracy and Development," *American Political Science Review* 87: 567-576.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge University Press.

Peters, Thomas. 19xx. *Xxx*. Yy: zz.

Podolny, Joel, and Karen Page. 1998. "Network Forms of Organization." *Annual Review of Sociology*, 24: 57-76.

Powell, Walter. 1990. "Neither Market nor Hierarchy: Network Forms of Organization." *Research in Organizational Behavior.* 12: 295-336.

Selznick, Philip. 1949. *TVA and the Grass Roots*. Berkeley: University of California Press.

Simon, Herbert. 1947. *Administrative Behavior*. New York: The Free Press.

_____. 1951. "A Formal Theory Model of the Employment Relationship." *Econometrica* 19: 293-305.

Taylor, M. 1976. *Anarchy and Cooperation*. London: John Wiley.

Williamson, Oliver. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications.* Free Press: New York.

_____. 1985. *The Economic Institutions of Capitalism.* Free Press: New York.

_____. 1996. *The Mechanisms of Governance*, New York: Oxford University Press.