LEARNING FROM SCHELLING'S 'STRATEGY OF CONFLICT' [1]

by Roger B. Myerson

http://home.uchicago.edu/~rmyerson/research/stratofc.pdf

Abstract. Thomas Schelling's *Strategy of Conflict* (1960) is a masterpiece which should be recognized as one of the most important and influential books in social theory. This paper reviews some of the important ideas in *Strategy of Conflict* and considers some of the broader impact that this book has had on game theory, economics, and social theory. By his emphasis on the critical importance of information and commitment in strategic dynamics, Schelling played a vital role in stimulating the development of noncooperative game theory. More broadly, Schelling's analysis of games with multiple equilibria has redefined the scope of economics and its place in the social sciences.

JEL Classifications B21, C72, D74, F51.

Author's address: Roger Myerson, Department of Economics, University of Chicago, 1126 East 59th Street, Chicago, IL 60637.

Email: myerson@ uchicago.edu.

Internet: http://home.uchicago.edu/~rmyerson/

First version: Aug 2007; *latest revision:* April 2009.

[1]Prepared for a conference at the University of Maryland in honor of Thomas C. Schelling, Sept 29, 2006.

Introduction

Thomas Schelling's *Strategy of Conflict* (1960) is a masterpiece which should be recognized as one of the most important and influential books in history of social science. Indeed, in just the first three chapters, Schelling introduced enough ideas to revolutionize social theory. Chapter 1, entitled "The Retarded Science of International Strategy," begins with a call for a new theory of interdependent decisions, to provide a framework for analyzing problems of conflict and deterrence that arise in international relations. Chapter 2, "An Essay on Bargaining," considers basic questions about how people achieve commitment to strategic threats and promises. Chapter 3, "Bargaining, Communication, and Limited War," examines how people achieve coordination in games that have multiple equilibria. Throughout, Schelling probes these questions by logical thought-experiments about problems of conflict and cooperation among people who are intelligent and rationally motivated by individual self-interest.

This paper reviews some of Schelling's important ideas in this book and considers some of the broader impact that this book has had on the theory of social science. First we examine the vital role that Schelling played in the development of noncooperative game theory, by his emphasis in *Strategy of Conflict* on the critical importance of information and commitment in strategic dynamics. Then, more broadly, we consider how Schelling's analysis of games with multiple equilibria has redefined the scope of economics and its place in the social sciences. For complementary perspectives on Schelling's important contributions, see also Richard Zeckhauser (1989), Vincent Crawford (1991), and Avinash Dixit (2006).

Schelling's impact on the development of game theory

The development of game theory methodology has been greatly influenced by Schelling's work, even though he has approached game theory more as an outsider or critic. Schelling's style is to teach by carefully considering specific examples rather than by formulating general mathematical theories, and chapter 1 in *Strategy of Conflict* includes a negative assessment of game theory's practical applicability in 1960. But, as one of those who like to formulate general mathematical theories, I want to argue here that the development of general models in game theory after 1960 was decisively influenced by Schelling's *Strategy of Conflict*.

To understand Schelling's impact on game theory, we must first review some basic ideas

of game theory. Any analytical discipline needs some general framework to make connections between different applications. Game theory has developed to provide a general framework for analysis of rational incentives in social systems. Modern game theory began with John von Neumann's (1928) definition of three general ways to represent games (see also Myerson, 1999). Von Neumann first defined games in a general dynamic extensive form, but then he argued that any extensive-form game can be reduced to a one-stage normal form, in which players simultaneously choose strategies which are plans of action for all observable contingencies. He then further reduced multi-player games to a simpler coalitional form which specifies only the total worth that each set of players could cooperatively guarantee themselves.

The argument for normal-form generality led John Nash (1951) to see noncooperative equilibrium as a general solution concept. For any game in normal form, a Nash equilibrium is a prediction of a feasible strategy for each player such that each player's strategy maximizes his own expected payoff given what is predicted for the other players. In the 1950s, however, most work in game theory continued to follow von Neumann's emphasis on cooperative coalitional-form analysis. (For a contemporary perspective on the limited role of Nash-equilibrium analysis, see R. Duncan Luce and Howard Raiffa, 1957, pages 104-105 and 180.) Nash equilibrium became the dominant analytical methodology in game theory only after Nash's ideas were extended by John Harsanyi and Reinhard Selten and others in the 1960s and 1970s.

But what started Harsanyi and Selten thinking about noncooperative game theory, a decade after Nash? In Harsanyi's published work, the first indication of his interest in noncooperative game theory is a footnote where, in response to Schelling's criticism, Harsanyi (1961, p. 181) promised to extend his theory of rational bargaining to noncooperative games. In a parallel intellectual development, Selten (1964, p. 626) noted, at the end of a paper about general values for cooperative games, that Schelling (1960) raised important questions about how people can commit themselves to promises or threats, but that such problems of commitment had been assumed away in the general cooperative approach to game theory. As Harsanyi and Selten subsequently noted in their book (1988, p. 365), "the great strategic importance of an ability or inability to make firm commitments was first pointed out by Schelling (1960)." (Selten, 1965, includes a similar remark.) So when we look for a link between Nash's early work and Harsanyi and Selten's subsequent development of noncooperative game theory in the 1960s, what we find

is Schelling's *Strategy of Conflict*. And Harsanyi and Selten were only two of the many readers who were deeply influenced by *Strategy of Conflict*.

Game theorists who read *Strategy of Conflict* could not help noticing that, as Schelling probed important problems of strategic conflict and bargaining, none of his examples made any use of the cooperative coalitional-form models that were the principal object of study in game theory of the 1950s. The coalitional form was mathematically simple but it threw out too much of the vital strategic structure that Schelling knew to be essential. But in *Strategy of Conflict*, we can find many games in normal form, which Schelling analyzes by finding their Nash equilibria.

For other dynamic bargaining examples in *Strategy of Conflict*, however, even normal-form analysis seems inadequate. Von Neumann's reduction to normal form tends to suppress questions about information and timing in games, and Schelling regularly argues that the information and timing of individuals' decisions may be crucial in our analysis of a dynamic bargaining game. Thus, *Strategy of Conflict* demonstrated both the importance of noncooperative equilibrium analysis and the inadequacy of doing it only in the normal form. To match the scope of Schelling's analytical examples, game theorists had to develop Bayesian equilibria for games with incomplete information (Harsanyi, 1967-8), correlated equilibria for games with communication (Robert Aumann, 1974), and sequential equilibria for extensive-form games (Selten, 1975; David Kreps and Robert Wilson, 1982). These great advances in noncooperative game theory were developed only after Schelling (1960) showed what we really needed from a theory of strategic conflict.

The problem of multiple equilibria

By definition, a Nash equilibrium is a prediction of a feasible strategy for each player such that each player's strategy maximizes his own expected payoff given what the other players are predicted to do. So given any game, a prediction of the players' behavior which is not a Nash equilibrium could not be commonly believed by all the players, because if everybody believed such a nonequilibrium prediction then at least one player would rationally prefer to choose some other strategy, different from his predicted strategy. That is, any scenario which is not a Nash equilibrium could not be generally accepted as a solution for the game. This remark is the basic justification for using Nash equilibrium as a general solution concept for games. Thus, when we

study a game that has only one equilibrium, it must be the only rational prediction of players' behavior, and then game theory seems very powerful.

A game can have many Nash equilibria, however, and then the above argument against disequilibrium predictions does not tell us what we should predict. So economists may feel uncomfortable with games that have multiple equilibria, because the multiplicity of equilibria seems to threaten our vested interest in economic determinism. But Schelling saw things differently, perhaps because he was more able to resist the siren calls of methodological determinism.

Schelling saw that the existence of such games with multiple of equilibria is a pervasive fact of life that needs to appreciated and understood, not ignored by economists. Chapter 3 of *Strategy of Conflict* offers a wealth of examples of games with multiple equilibria where the essential problem is for players to coordinate. In such games, Schelling argued, anything in a game's environment or history that focuses the players' attention on one equilibrium may lead them to expect it, and so rationally to play it. This focal-point effect opens the door for cultural and environmental factors to influence rational behavior.

The other obvious way to bring cultural effects into economic analysis would be to assume that individual preferences are culturally determined. Of course, nobody can deny that our tastes are influenced by our social environment. But in a theoretical model where individuals' preferences are endogenously determined, one might "solve" social problems by teaching the poor to love poverty, or by teaching the powerful to love social justice. Economic analysis of institutions could thus be trivialized by such an assumption that individuals could be culturally reshaped to fit institutional requirements. So economists find it more useful to assume that an exogenous selfish materialism characterizes individual preferences. But even without an intrinsic preference for justice, selfish materialists can be influenced by concepts of justice that operate as focal factors to determine focal equilibria in games with multiple equilibria.

From a modern post-Nash perspective, the process of negotiating joint expectations of focal equilibrium may be the only sense in which players can truly "co-operate," because the understanding of a focal equilibrium must be shared jointly by all players. As Harsanyi (1961) recognized, theories about cooperative games may then be posed as theories of how focal equilibria can be identified from the payoff allocations that they yield. Thus, Schelling's focal-

point effect may provide a new basis for cooperative game theory.

The greatest theoretical effort to solve the multiple-equilibrium problem within the mathematical structure of game theory was by Harsanyi and Selten (1988), who labored to find a natural rule for identifying a unique equilibrium for every game. From the beginning of this project, Harsanyi (1961) wanted to extend individual decision-theoretic rationality by adding a strong-rationality postulate that rational behavior in games should depend only on the payoffs, not on payoff-irrelevant environmental factors. So to define strong rationality in games, Harsanyi and Selten sought a natural rule for selecting one unique equilibrium in every game, depending only on the players' payoffs in the game. But Schelling (1960, appendixes B and C) responded that any theorist could see his own equilibrium-selection theory as self-enforcing, once it is generally accepted (in the theorist's hypothetical world, at least).

Imagine for a moment that we tried to induce people in real games to play the equilibrium that is selected for them by a theory such as Harsanyi and Selten's. The players in these games might think that, in our attempt to coordinate them, we were trying to act as leaders with some kind of authority over them. That is, they might recognize something intrinsically political about our attempt to coordinate them on one equilibrium rather than another, because one of the basic functions of political leadership is to coordinate people's expectations in games with multiple equilibria. In this sense, Harsanyi and Selten (1988) were trying to define a neutral political theory, based on a new kind of natural law, which could be compared in scope to the efforts of Thomas Hobbes (1651), albeit in a very different analytical framework. Schelling's response to Harsanyi then could be compared to David Hume's (1748) observation that general public opinion may be the only standard for questions of morals (unlike other areas of philosophical inquiry), because the fundamental basis of social morals is in people's need to coordinate with each other.

Schelling's focal-point effect should be counted as one of the most important ideas in social theory. Recognizing the fundamental social problem of selecting among multiple equilibria can help us to better understand the economic impact of culture on basic social phenomena such as social relationships, property and justice, political authority and legitimacy, foundations of social institutions, reputations and commitment, international boundaries in peace and war, and even the social use of the divine. Let me review each with some simple examples.

Personal relationships, social pathologies, and cultural roots of poverty

Consider the game in Table 1, where players 1 and 2 must simultaneous make independent choices about whether to approach each other in a friendly or aggressive way. If player 2 is expected to be friendly then player 1 can maximize his payoff by being friendly also, as 50>40. But if player 2 is expected to be aggressive then player 1 can maximize his payoff by being aggressive, as 20>0. Similarly player 2's best response is to be friendly if player 1 is expected to be friendly, but 2's best response is to be aggressive if 1 is expected to be aggressive. So both players being friendly is a Nash equilibrium, yielding a good expected payoff allocation (50,50). But both players being aggressive is also a Nash equilibrium, yielding the expected payoff allocation (20,20), which is worse for both players. (In addition, there is a randomized equilibrium in which each player has an independent probability 2/3 of being friendly.)

	2 friendly	2 aggressive
1 friendly	50, 50	0, 40
1 aggressive	40, 0	20, 20

1's payoff, 2's payoff

Table 1: A game with good and bad equilibria. (The stag-hunt game.)

As the strategies' names suggest, the different equilibria here can be interpreted as representing, in a simple model, different kinds of interpersonal relationships. The players here can rationally have a friendly relationship, but they can just as rationally have an unfriendly aggressive relationship, even though it makes them both worse off.

Economists have sometimes suggested that a Nash equilibrium that is Pareto-dominated by another Nash equilibrium should never be predicted by our theories. If we included this Pareto constraint in our analytical methodology, then we would get a unique and pleasant prediction in this game, because the good (50,50) equilibrium Pareto-dominates all other equilibria here. But such a methodological assumption would blind us to the possibility of important social pathologies where people become focused on the bad equilibrium. Suppose that this game is played in a context where, based on cultural expectations and the experiences of players in similar situations, aggressive behavior becomes the normal expectation. Then each player should rationally respond to this expectation by fulfilling it himself. Neither player can

improve the situation by himself. The pathology of the aggressive equilibrium here is a social pathology that derives from their mutual expectations of each other. To improve individuals' prospects in this game requires a social change, that is, a change of generally held social expectations.

Now imagine two different islands, each of which is inhabited by people who are matched in pairs to play this game each day. Suppose that people on the first island are culturally disposed to focus on the good equilibrium, but on the second island the players are culturally disposed to focus on the bad equilibrium. Then we have a simple model of impoverishment that is purely cultural in origin. By modeling local culture as a focal factor, rather than as a factor that affects individuals' preferences, we can meaningfully compare payoffs across the two islands, because moving from one island to the other would not change a player's preferences. The two islands have the same economic fundamentals, but the economic outcomes are worse on the second island because the cultural expectations are different. All individuals on both islands are equally rational, but an individual on the poor island who tried to fix the problem by acting as if he were on the rich island would reduce his own payoff from 20 to 0. If we want to cure the poverty of the second island, we must get everyone's attention there and somehow get them all to focus instead on the better equilibrium. Such pathological social expectations can be changed only by someone who perceived as an authority or leader, and who can identify a better Nash equilibrium for them.

Of course the social problems of poverty actually arise in complex social systems where different equilibria are much harder to identify than in this simple example. In a game that has many equilibria, there are typically many more strategy combinations that are not Nash equilibria. A would-be reformer who wants to improve social welfare by changing people's behavior to a better equilibrium must take care to identify a social plan that is in fact a Nash equilibrium, so that nobody can profit by unilaterally deviating from the plan. If a leader tries to change people's expectations to some plan that is not a Nash equilibrium, then his exhortations to change behavior would be undermined by rational deviations. The point of this example is that, even when the better equilibrium is well understood, there still remains a nontrivial social problem of how to change everyone's expectations to the better equilibrium. Such coordinated social change requires some form of socially accepted leadership, and thus it may depend on

factors that are essentially political.

Property rights and justice

Ownership and transfer of property rights are central concerns of economic theory, but the social mechanisms that sustain property rights are rarely modeled (see Dixit, 2004). Economists regularly assume that transactions costs are negligible, and indeed transfer of valuable property is often effected by a signature on paper or by a simple handshake. How can the ownership of assets be transferred merely by a signal that is commonly witnessed by two or more people, without any substantial movement or transformation of the assets themselves? Such ease of economic transactions can be readily understood when we see the allocation of property rights as a focal equilibrium in a more fundamental ownership game that has multiple equilibria. Then public words could indeed transfer property, by changing the social focus from one equilibrium to another in the ownership game.

For a simple example of an ownership game, consider the game in Table 2, where V and c are given positive numbers (say, V=99, c=1). Each of the two players independently decides whether to claim a valuable asset that lies physically close to both of them. The payoff from winning the asset is V, but the cost of fighting for it against a rival claimant is c.

	2 claims	2 defers
1 claims	−c,−c	V, 0
1 defers	0, V	0, 0

1's payoff, 2's payoff

Table 2: A game among rival claimants to a valuable asset worth V.

This game has three Nash equilibria. There is an equilibrium in which player 1 claims and 2 defers, yielding payoffs V for player 1 and 0 for player 2. This equilibrium corresponds to the social understanding that player 1 "owns" the asset. But the game also has an equilibrium in which player 2 claims and player 1 defers, yielding payoffs 0 for player 1 and V for player 2, and this equilibrium is our model of 2 owning the asset. In addition, the game has a symmetric equilibrium in which each player independently randomizes, claiming with probability $V/(V+c)$, but deferring with probability $c/(V+c)$, so that each player gets an expected payoff equal to 0.

(Notice $0 = (-c)V/(V+c) + (V)c/(V+c)$.) The symmetric equilibrium may be interpreted as a social state where property rights are ill-defined.

(We could extend the game to allow that, if both players claim or both players defer then they will play again, repeating until a round when someone defers while the other claims. This repeated game has sequentially rational equilibria in which the players would act as in the symmetric randomized equilibrium at each round after the first, so that expected payoffs after the first round are zero. Then the game at first round is still essentially as in Table 2, with the same three possible equilibria for their first-round play.)

Plato's *Republic* (book 1) includes a definition of justice as giving each person what is due to him. In this sense, a shared social expectation of who should claim and who should defer in this game is indeed justice. If any shared cultural concept of justice suggests that 1 should claim and 2 should defer in this game, then the players should rationally play that equilibrium. Player 2 may regret this equilibrium, as she gets 0 in it, but the expectation of 1's just claim here should make 2 anticipate that her "unjustly" claiming here would reduce her own payoff to $-c$. Thus Schelling's focal-point effect enables us to understand how social concepts of justice can influence rational behavior, even when people are motivated only by their own self-interest, with no intrinsic desire to act justly. The behavior of selfish materialists can be decisively influenced by justice as a social concept that designates the focal equilibrium in such games of multiple equilibria. With such a multiple-equilibrium model of ownership, it is easy to see how the allocation of property rights can be influenced by symbolic rituals like a verbal promise over a handshake, as these rituals can change the socially expected equilibrium of the claiming game.

When they start from a symmetric state where neither player seems more likely to claim, the players might ask a mutual friend to arbitrate for them, and perhaps such an arbitrator might decide in favor of one player or the other by tossing a coin. When the arbitrator's decision or coin toss has designated 1 (say) as the player who should claim here, this focal arbitration can become a self-fulfilling prophecy, enforced merely by 2's fear that her claiming would cause conflict, and by 1's confidence that his claim will not be contested. Thus the word of a recognized authority or focal arbitrator can allocate valuable property rights without any enforcement mechanism other than people's desire to avoid costly conflict.

Of course, the loser in such an arbitration might prefer that they revisit the question by

going to another arbitrator or by tossing another coin. If some social leader is seen as a higher authority than their mutual friend, then the friend's judgment could be annulled and reversed by a judgment from the leader. So the players can be coordinated in this way only if they have a shared perception of who has jurisdiction over their case and when his judgment is final.

To focus attention on one equilibrium with no higher appeal, it would be best to consult the highest possible authority. If the players share a cultural understanding that certain unpredictable processes may be used by the fundamental divine spirit of the universe to answer questions, and that this divinity cannot be bothered about the same question more than once, then a recommendation that is based on such a sacred randomization can serve as a focal coordination device that cannot be appealed to any higher arbitrator. Then the oracle's recommendations can be self-enforcing, without any further intervention by the divine spirit, provided that the recommendations to the players form an equilibrium. Thus the focal-point effect can admit a socially significant role for oracles and divination, as an effective foundation for social coordination. Indeed, when we look for effective focal factors, what can command people's attention more than the overall pattern of the whole universe? This divine pattern can serve as a focal determinant, however, only when players have a shared understanding about it can be interpreted into a selection among the set of Nash equilibria of their game.

Legitimate authority and the foundations of the state

The many coordination games in chapter 3 of *Strategy of Conflict* teach us to see games with multiple equilibria as an important part of the world. In such a world, strategic coordination with other people becomes one of the principal necessities of life on which people's welfare depends.

For example, imagine a chain of islands, each of which is inhabited by people who are randomly matched every day into pairs who play the rival-claimants game in Table 2 above (as in Myerson, 2004). On one primitive island where social differentiation is lacking, everyone might play the symmetric randomized equilibrium, in which high frequency of conflict dissipates all of the expected resource value V, and so everyone's expected payoff there would be 0. But expected payoffs could be positive on another island where people have cultural notions of justice and ownership that tell them who should claim and who should defer in some of these

rival-claimant games. Where property rights are unclear, people might gather in occasional assemblies to approve new rules for identifying which player should claim in more of these matches; and thus the focal-point effect can explain the natural development of legislative assemblies. Furthermore, the players might agree on a leader who can perform general focal arbitrations, to assign claiming rights for all matches where traditional property rights or other social rules do not apply. Common recognition of such rules and authority is all that is needed to make their designated equilibria become self-enforcing prophecies in these games with multiple equilibria.

Now suppose that the players' payoffs in this model can be interpreted as resources that increase people's long-term reproductive fitness. Then an anarchic island where these resources are wasted in the symmetric equilibrium would sustain a much smaller population than another island where people have systems of rules and authority to coordinate them on better equilibria. If players from highly populated islands can colonize underpopulated islands, taking with them their cultural system of focal-equilibrium selection, then an archipelago of such islands should eventually be inhabited only by people who have systems of authority to coordinate them efficiently in matches where there are multiple equilibria. So from a Darwinian competition among cultures, we derive a general principle that every successful society must have well-developed concepts of justice and authority that enable their members to identify efficient focal equilibria in most of the games that they commonly play.

Such models offer a new perspective on the nature of politics and the state. We may see political leaders as general focal arbitrators for games throughout society where other cultural norms do not apply or require interpretation. In our model, such focal coordination power could be vested in any individual, provided only that each matched pair must share a common understanding about who has jurisdiction over their case. As Russell Hardin (1989) has observed, severe costs of anarchy can make the process of constituting a state into a game with multiple equilibria. So the problem of agreeing about social leadership remains as a social coordination problem, but it is the coordination problem to solve all other coordination problems.

In response to Hobbesian theories of an original social contract being the foundation of the state, Hume (1748) argued that the foundations of political power generally depend, not on any prior consent of the population, but merely on a common recognition by the population.

That is, the establishment of a sovereign government may be effected by a generally shared perception or belief of the population. From our perspective, we can see Hume's argument as a fundamental application of Schelling's (1960) focal-point effect. Our island of rival-claimant matches is a simple model of how the power of a leader to allocate property rights can be derived purely from a general common recognition of his focal-arbitration authority.

Focal factors that bestow such coordination power on a leader may be called legitimacy (or charisma when they are intrinsic to the leader's personality). By the focal-point effect, the selection of legitimate leadership in any society can depend on its particular culture and history, such as a local tradition of identifying a particular family as royal.

Focal arbitration power is power to redefine property rights, which of course may be profitable for the focal arbitrator himself. But a society's generally understood rules for recognizing leaders can impose constraints what a leader must do if he wants to retain his generally recognized position of authority. Thus, constitutional constraints on leaders can be effected when their legitimation is understood to be granted conditionally on their maintaining some standard of appropriate behavior. Such forms of constitutionalism may exist even in societies that do not have written constitutional documents or democratic elections.

In any society, it is vital to maintain a broad general agreement about who has legitimate authority in any situation. Thus, from our earlier remarks about the focal power of the divine, we can see why societies may find it useful or even essential to call for frequent testimony that the local system of rules and authority is compatible with the divine pattern of the universe. Although coordination within a society can be improved by such belief in the universal nature of its principles of justice and legitimate authority, such universalization of local law and authority makes it harder for people in one society to recognize the different forms of justice and authority in other societies (see Myerson, 2006). That is, the same forces that help people to achieve consistent coordinated expectations in a successful society can become forces for inconsistency of expectations across societies in international relations. Indeed, in international conflicts throughout history, people on each side have regularly failed to understand the other side's perception of justice in their conflict.

General theory of institutions enforced in larger games

Leonid Hurwicz (2008), in the culmination of his long analytical inquiry into the design of economic systems and institutions, raised the basic question of how institutional rules are enforced. Game theorists regularly consider institutions as games, so that an institutional reform can be analyzed as a change in the rules of a game. But at a deeper level, we should also ask how such institutional games can be reformed or established in the first place. Having found focal coordination problems at the foundations of the state, we should recognize the essential role of Schelling's focal-point effect in such analysis.

To ask game-theoretically how an institution is established, we must consider the institution as a game that is embedded in a larger and more fundamental game, which Hurwicz (2008) calls the true game. The embedded institution differs from the true game in that some feasible strategies in the true game are not allowed in the institution. For example, when two people are matched in a game of chess, typically each of them is physically able to grab the other's king at any time, but is deterred from chess-illegal moves by the damage such behavior could do to one's reputation in the larger game of life. So the chess game seems supported by some kind of reputational equilibrium in the larger true game of life.

To provide a framework in which different institutions could be established with different patterns of individual behavior, the larger true game must have multiple equilibria. Then an institutional reform and its induced changes in people's behavior can be understood in this framework as a change from one focal equilibrium to another. Indeed, without multiple equilibria, it would be difficult to understand how a revolutionary change of national institutions can dramatically change so much in a nation where the population and resources have not substantially changed in any material aspects.

It might seem, then, that this perspective could be summarized by saying that different institutions correspond to different equilibria in a larger game of life; but this proposition cannot be technically correct. If a chess match were understood as an equilibrium in a larger game of life, then that equilibrium would specify each player's strategy in the chess game itself. But people can be in a chess match without playing optimally. Being in the chess match only means that the players are choosing strategies that are consistent with the rules of chess, not that they are choosing their best strategies.

In Hurwicz's (2008) framework, establishing an institution means ruling out certain strategies that are feasible in the true game of life but are "illegal" under the rules of the institution. So given a normal-form game that describes the true game of life, an institution would be described by listing the nonempty set of strategies for each player that are legal in the institution. Then we may say that the institution's rules are rationally enforceable iff the utility-maximizing best responses for any player are always legal strategies whenever the other players are all expected to use legal strategies (possibly with randomization). With this definition, the legal strategies of an enforceable institution must form what Kaushik Basu and Jorgen Weibull (1991) have called a curb set in the larger true game. (See also Myerson, 2009.)

In a large game, the set of strategy profiles may contain many different subsets that are curb sets, in this sense of being closed under all players' best-response mappings. (Any pure-strategy equilibrium is a one-point curb set, for example.) But this multiplicity is exactly what is required for curb sets to be our basic conceptual model of enforceable institutions, so that reforming to different institutional structures is possible but cannot be achieved without broad coordination of individuals in society.

The new theoretical point here is that Schelling's focal-point effect can be extended to questions of selecting among multiple curb sets, just as among multiple equilibria. In this framework, the focal-point effect can help us to understand how the institutions in any society depend on its particular history and culture. Once everyone understands that everybody else will be restricting themselves to strategies in one particular institutional curb set, it becomes rational for each individual to stay in his or her respective portion of this curb set, and so legal institutional behavior is rationally enforced.

The problem of credible commitment

Schelling approached the theoretical question of strategic commitment from applied questions about bargaining and about deterrence in international relations. In bargaining, a threat to take one's business elsewhere may be used to get a favorable price concession. In international relations, a threat of massive retaliation may be used by a nation to deter another nation from some opportunistic intervention.

In the general framework of a dynamic multi-stage game, player 1 could influence player

2's behavior at earlier stages of the game if 1's strategic response at the later stages would be conditional on what 2 has done earlier. Player 1 may want player 2 to believe that 1 will implement some promises if 2 acts as 1 wishes, but that 1 will implement some threats if 2 deviates from the desired plan of action. If the value for player 1 of these promises and threats is only to influence player 2's choice of action, however, then there is no general reason to expect that 1 would actually want to implement them after 2's action has become an accomplished fact. Thus there is a general credibility problem for strategic promises and threats.

This strategic credibility problem is intrinsically dynamic, and so it is one of those important problems that tended to be obscured by von Neumann's reduction of dynamic extensive games to the one-stage normal form. In raising these questions of strategic credibility questions, Schelling (1960) helped to stimulate the return of game theorists to study dynamic extensive-form games.

This question of how rational agents can commit themselves to costly threats and promises (which benefit the agent only in anticipation by their influence on others) is central in chapter 2 of *Strategy of Conflict*. One easy answer to the commitment problem might be to rely on contract-enforcement services of the state, but such basic reliance on the state only begs the question of how state officials are committed to their legal roles (punishing people if they break contracts, but not if they don't). So we should recognize that the problem of commitment is fundamental to the state and all other great institutions of society.

Schelling probes this fundamental problem by examining it in simple two-person games, where he shows that a player may be able to achieve commitment by staking his reputation with the other player. That is, player 1 may be able to commit himself to implement promises and threats that are costly in the short run by an understanding that 1's failure to do so would adversely change his relationship with player 2.

For example, consider the "prisoners' dilemma" game in Table 3. In this game, a player can gain (+20) by switching from friendly to aggressive, but such a switch causes the other player to lose more (−30). So when this game is played once, there is a unique equilibrium in which both are aggressive and they get the Pareto-dominated payoff allocation (20,20).

	2 friendly	2 aggressive
1 friendly	30, 30	0, 50
1 aggressive	50, 0	20, 20

1's payoff, 2's payoff

Table 3: A game with one bad equilibrium. (The prisoners' dilemma.)

Now suppose that this game will be played twice, with the players moving simultaneously at each round, both learning the first-round outcome before they move at the second round. In this two-stage game, player 1 would like to give player 2 an incentive to be friendly at the first round by promising: "If you are friendly at the first round then I will be friendly at the second round, but otherwise I'll be aggressive at the second round." If this statement were credible and the players do not significantly discount their second-round payoffs, then gaining 30 from 1's friendship at the second round should be worth more to player 2 than the 20 that she could get from being aggressive at the first round. But unfortunately, the above statement is not credible, because player 1 would have no incentive to fulfill his promise to be friendly at the second-round prisoners' dilemma when it is the end of the game.

But suppose instead that the first-round prisoners'-dilemma game in Table 3 will be followed at the second round by the game in Table 1, which has good and bad equilibria. In this two-stage game, there is a sequentially-rational equilibrium in which both players are friendly at the first-round prisoners' dilemma; and at the second round they will play the good equilibrium after both were friendly at the previous round, but they would play the bad equilibrium if either player was aggressive at the previous round. Both the good promise and the bad threat are credible because they are Nash equilibria of the second-round subgame in Table 1, and the net second-round reward of 50−20=30 for first-round friendship is enough to make each player willing to forego the profit of 50−30=20 from first-round aggression. (This argument works even with time discounting if the discount factor per round is greater than 2/3.)

From this example we find further reason to admit both the good and bad equilibria as potentially valid solutions for the game in Table 1. When we first considered this game above, I suggested that the bad (20,20) equilibrium could be interpreted as a model of social pathologies that we need to understand. But in this two-stage model, the possibility of having such a bad

relationship at the second round is what enables the players to have a good trusting relationship at the first round. Thus, social relationships that seem dysfunctional in the short run might actually have a positive social function when viewed from a longer perspective.

Similar equilibria can be constructed for games where the prisoners'-dilemma game in Table 3 is repeated any finite number of times and then is followed by one final round of the game in Table 1. For any such repeated game, there is an equilibrium in which both players will be friendly as long as neither has been aggressive at any past round, but if either ever deviated to act aggressively then both would be aggressive thereafter. (With discount factor greater than 2/3, this construction can also be applied to an infinitely repeated version of the prisoners' dilemma game in Table 3.) In such equilibria, each player effectively motivates the other's good behavior by promising friendly responses to the other's friendship and threatening aggressive responses to the other's aggression. Of course, aggressive behavior at all rounds is also an equilibrium of such games. Indeed, the good equilibria here are sustained by threats of switching to a bad equilibrium.

So the existence of multiple equilibria is essential here. Each player's appropriate behavior in equilibrium is motivated by a combination of promises and threats by others. But if a player's deviation to inappropriate behavior would not actually change anything materially in the subgame that follows, then different promises and threats can all be credible in the subgame only if it has multiple equilibria. Thus, the problem of making threats and promises credible brings us to further appreciate the importance of multiple equilibria in games.

These different equilibria may be interpreted as different kinds of relationships among the players. In long-term social situations where people have many opportunities to observe each others' actions and respond to them, an individual can be credibly committed to a strategy by an expectation that any deviation from this strategy could cause an adverse change in his relationship with others. Generalized versions of this logic have become known as the folk theorem of repeated games.

From this perspective, we may return to the question of how high state officials can become committed to act according to their legal and constitutional roles which are vital to the foundations of the state. The straightforward answer is that such a high official can expect to enjoy a privileged social position if he fulfills his legal and constitutional responsibilities but

could lose these privileges if he does not (see Becker and Stigler, 1974). His possession of this high office and its advantageous social relationships can be understood as a focal equilibrium of a fundamental social coordination game which has multiple equilibria, because there are other people who could hold this office instead. The key is that effective possession of a valuable high office requires general social recognition, and each member of society wants to recognize a powerful official who is recognized by everyone else. Thus constitutional rules depend on general social understanding of how legitimate holders of high office are identified.

Deterrent strategies and mutually recognized boundaries

As an application of the above ideas about credible strategic commitment, let us follow Schelling and consider again the question of how a nation can commit itself to a deterrent strategy that is intended to deter other nations from opportunistic aggression. Such a deterrent strategy must threaten some punishment against such aggressors but it must also promise some better treatment for others who abstain from such aggression. A credibility problem arises if either the punishment or the better treatment would sometimes not be in the nation's short-run interest. But if the long-run costs of not deterring aggression would be sufficiently large, then the long-run advantages of a reputation for acting according to this deterrent strategy can make it worthwhile for this nation to pay the short-run costs of threats and promises. The key is that the nation must have a reputation for acting according to a specific deterrent strategy, and that the nation would risk losing this reputation if it ever deviated from this deterrent strategy.

Notice that the critical reputation here is how this nation is perceived by those other nations that it wants to deter, because it is their behavior that is supposed to change if the nation deviated from the deterrent strategy. Thus Schelling discovered the paradoxical fact that an effective deterrent strategy requires coordination with the adversaries who are supposed to be deterred by it. We are familiar with the need to coordinate with our allies in international conflict, but the importance of coordinating with our adversaries is less intuitive. Our strategy will not have the desired deterrent effect if it is not generally understood and recognized by our potential adversaries. Furthermore, we cannot be credibly held to a deterrent strategy unless our potential adversaries understand it in detail and can judge our correct implementation of it. If the terms of our deterrent strategy are vague or ambiguous, then in any specific case we would be

tempted to deny our obligation to fulfill any costly promise or threat. Rational foundations for international law of war (*jus ad bellum*) can be derived from this basic argument that the strategic conditions for deterrent military responses need to be generally recognized and verifiable.

Of course a nation may have many different strategies which, if anticipated by others, could all have the same desired deterrent effect. Thus, the problem of sustaining a credible deterrent strategy is a problem of negotiating a focal equilibrium in a dynamic game that has multiple equilibria, where the other players include potential adversaries around the world. All the historical and cultural factors that can influence focal equilibrium selection may be critical in defining what kind of deterrent strategy would actually be credible.

A classic example is the Melian debate in Thucydides's *History of the Peloponnesian War*. Here the critical question was whether Athens had a reputation for subjugating all islands in the Aegean Sea or a reputation for subjugating all islands settled by ethnic Ionians. Either interpretation of Athens's reputation could be maintained in an equilibrium, with the understanding that many islands would simultaneously throw off Athens's imperial yoke if it ever lost its reputation. The people of Melos, being ethnic Dorians, advocated the second interpretation and hoped that the Athenians would accept it. But the Athenians felt compelled to destroy Melos, even though it ultimately required a costly punitive expedition, because they believed that the other islands would inevitably focus on the first interpretation and so would rise in revolt if Melos were spared.

From this perspective, Schelling teaches us to see international boundaries as rich game-theoretic phenomena. Each nation needs to maintain a reputation for vigorously defending its boundary. In equilibrium, a nation can be motivated to react strongly against even a tiny violation of its boundary, by the fear that a weaker response could re-focus the world on another equilibrium in which this nation would surrender much more territory without a fight. But if neighbors see their shared boundary differently, then one's defensive reaction may be interpreted as an aggressive provocation by the other, stimulating a worse counter-reaction. So to enjoy the blessings of peace in a world where all nations feel obliged to vigorously defend their boundaries, everyone needs to agree on the exact location of all these boundaries. Thus the world is divided by a focal equilibrium of a coordination game, where even small disagreements can be very costly for all concerned. The focal-point effect explains how the location of

international boundaries can be arbitrarily dependent on accidents of history that would otherwise have no relevance today.

A reconstruction of social theory

Thus, the scope of rational-choice analysis, which has long been fundamental to economics, has been significantly broadened by Thomas Schelling's *Strategy of Conflict*. Schelling showed that practical insights into global conflict can be gained by viewing our adversaries as intelligent rational decision-makers like ourselves, and by logically analyzing our rational decisions and theirs together in a common framework that takes account of how people's preferences and information may differ. The general framework for such analysis of interdependent decisions is the subject of noncooperative game theory, and its development was significantly accelerated and redirected after 1960 by the impact of Schelling's book.

Schelling's emphasis on games with multiple equilibria also transformed the relationship between economic analysis and other areas of social theory. Recognizing the problem of social coordination in games with multiple equilibria may require us to abandon a simple faith in economic determinism, but it opens the door to admit traditional concerns of sociology into the domain of economic analysis. Schelling's focal-point effect enables us to better understand how the cultural environment can affect rational economic behavior, even when people in different cultures have the same individual goals and desires. Different kinds of social relationships, individual reputations, and social positions can be understood as alternative equilibria in an economic model.

Within the traditional domain of economics, questions about how property rights are enforced have been widely recognized as vital but have been rarely analyzed by economists. One reason may be that, when we formulate a fundamental model of how assets are owned, if this ownership game had one unique equilibrium then it would be hard to justify the common economic assumption that ownership can be costlessly transferred from one individual to another. But when the fundamental ownership game has multiple equilibria, then symbolic documents that transfer ownership can function by changing the focal equilibrium.

Schelling's questions about credible strategic commitment also offer deep insights into the foundations of social institutions and the importance of social coordination among multiple

equilibria. The problem of individual commitment is inextricably connected to the problem of committing social institutions to act according to rules of law. People may look to institutions of law to enforce individual contracts and commitments, but the lawful action of any social institution itself depends on its officials and agents being individually committed to their institutional roles. Schelling showed how an individual's problem of commitment can be solved by a reputational equilibrium, in which the individual is generally perceived to have a favorable reputation that he could lose by deviating from expected norms of behavior. But there may be many ways to draw the reputational boundaries that are to be maintained, and so the construction of credible strategic commitment again depends on social coordination within culturally accepted principles. So in the development of a credible deterrent strategy in international relations, or in the creation of a new constitutional democracy, a rational economic analysis must take account of the decisive importance of traditional cultural perceptions.

Thus, basic questions of political theory can be understood in terms of the general focal coordination problem. We have seen how the powers of legislative assemblies, political leaders, and judges in their jurisdictions can all be derived from a fundamental focal arbitration power, that is, the power to create general expectations about which equilibrium will be played in a game with multiple equilibria. The problem of identifying who are leaders in a society is itself a coordination problem where the outcome must depend on cultural concepts of legitimacy. Hume's observation that the foundations of political power generally depend on a common recognition by the population can be understood as an application of Schelling's (1960) focal-point effect. These fundamental political problems of constituting the state and selecting political leaders can be seen as the social coordination problem to solve all other social coordination problems.

This perspective suggests the elements of a general theory of social change. In any successful society, social rules are enforced by high-status individuals whose privileged status depends on their acting according to these rules, in a reputational equilibrium. At the top of this high-status group, there must be some social leaders who have generally recognized authority for focal arbitration in various domains, where they can adjudicate disputes and coordinate plans that go beyond the scope of established rules. These individuals would have achieved their current social status by criteria that were defined by social rules and leaders' judgments in the past. But a

publicly expressed consensus among these focal leaders could modify the social rules for the future. In this sense, the essential state of a society at any point in time may be embodied in its recognized leaders and their reputations.

The focal-point effect may even offer a perspective on some ideas of theology, not about the nature of the divine, but about how societies use the divine. The focal-point effect is about environmental parameters that attract people's attention to one of many equilibria, and no aspect of our environment has a stronger claim on our attention than the divine pattern of the entire universe. Thus, coordination in a society can be strengthened when it culturally portrays local forms of law and authority as derived from universal divine principles.

Even today, in the early twenty-first century, this author still sees much in the daily news that calls for broader application of ideas that Schelling formulated in 1960. A credible deterrent strategy to protect our nations' security requires a national reputation for resolve against provocation, and for restraint in the absence of provocation; and our adherence to this strategy must be recognized by our adversaries and judged in a multinational forum. The restoration of peace and security in a country that we have invaded must depend first and foremost on the legitimation of its new leadership according to that society's deeply-rooted norms and traditions. As we affirm the great universal principles that underlie the system of justice and authority in our nation, we should not assume that our nation's use of military force does not need to be judged by other nations, nor should we assume that our projection of authority in a foreign land will be compatible with the principles of legitimation which prevail there. And we still need to learn Schelling's basic lesson that, in a realistic analysis of international conflict, we should consider our adversaries as rational intelligent decision-makers, whose interests are different from ours, but with whom we share a fundamental problem of coordinating mutual strategic expectations.

References

Robert J. Aumann, "Subjectivity and correlation in randomized strategies." Journal of Mathematical Economics 1:67-96 (1974).

Kaushik Basu and Jorgen W. Weibull, "Strategy subsets closed under rational behavior," Economics Letters 36:141-146 (1991).

Gary Becker and George Stigler, "Law enforcement, malfeasance, and compensation of

enforcers," J Legal Studies 3:1-18 (1974).

Vincent Crawford, "Thomas Schelling and the analysis of strategic behavior," in Strategy and Choice, edited by Richard J. Zeckhauser, MIT Press (1991), pp. 265-294.

Avinash K. Dixit, Lawlessness and Economics, Princeton U Press (2004).

Avinash K. Dixit, "Thomas Schelling's contributions to game theory," Scandinavian Journal of Economics 108(2):213-229 (2006).

John C. Harsanyi, "On the rationality postulates underlying the theory of cooperative games," Journal of Conflict Resolution 5:179-196 (1961).

John C. Harsanyi "Games with incomplete information played by Bayesian players," Management Science 14:159-182, 320-334, 486-502 (1967-1968).

John C. Harsanyi and Reinhard Selten, A General Theory of Equilibrium Selection in Games, MIT Press (1988).

Russell Hardin, "Why a Constitution," in The Federalist Papers and the New Institutionalism, edited by B. Grofman and D. Wittman, NY: Agathon Press (1989), pp. 100-120.

Thomas Hobbes, Leviathan (1651), edited by C. B. MacPherson (Penguin Classics, 1982).

David Hume, "Of the Original Contract" (1748), reprinted in Political Essays, edited by K. Haakonssen, Cambridge U Press (1994), pp. 186-201.

Leonid Hurwicz, "But who will guard the guardians," American Economic Review 98(3):577-585 (2008).

David Kreps and Robert Wilson, "Sequential equilibria." Econometrica 50:863-894 (1982).

R. Duncan Luce, and Howard Raiffa, Games and Decisions, Wiley (1957).

Roger Myerson, "Nash equilibrium and the history of economic theory," Journal of Economic Literature 37:1067-1082 (1999).

Roger Myerson, "Justice, institutions, and multiple equilibria," Chicago Journal of International Law 5:91-107 (2004).

Roger Myerson, "On game-theoretic consistency and the theory of international relations," Journal of Theoretical Politics 18:416-433 (2006).

Roger Myerson, "Fundamental theory of institutions," to appear in Review of Economic Design (2009).

John F. Nash Jr., "Noncooperative games." Annals of Mathematics 54:289-295 (1951)

John von Neumann, "Zur Theorie der Gesellschaftsspiele." Mathematische Annalen 100:295-320 (1928). English translation by S. Bergmann in R. D. Luce and A. W. Tucker, eds., Contributions to the Theory of Games IV (1959), pp. 13-42, Princeton University Press.

Thomas C. Schelling, The Strategy of Conflict, Harvard U Press (1960).

Reinhard Selten, "Valuation of n-person games," in Advances in Game Theory, edited by M. Dresher, L. S. Shapley, and A. W. Tucker, Princeton University Press (1964), 577-626.

Reinhard Selten, "Spieltheoretische Behandlung eines Oligopolmodells mit Nachfragetragheit." Zeitschrift fuer die gesampte Staatswissenschaft 121:301-329, 667-689 (1965).

Reinhard Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games." International Journal of Game Theory 4:25-55 (1975).

Richard J. Zeckhauser, "Reflections on Thomas Schelling," Journal of Economic Perspectives 3(2):153-164 (1989).