AEA Continuing Education Program – January 2011

Game Theory

Avinash Dixit, Princeton University and David Reiley, Yahoo! Research

I. Topics in Basic Theory

by Avinash Dixit

The last 20-25 years have brought several important new ideas into basic game theory. Most of them have to do with relaxing some restrictive or undesirable features of the simplest or '101-leve' Nash equilibrium. In some cases the new concepts help resolve the problem of multiplicity of Nash equilibria; in other cases they point to a plausible outcome when no Nash equilibrium exists; in still other cases they point to a plausible alternative when the Nash equilibrium is unrealistic or implausible. In the time available, I will touch upon just a few of these developments that I find particularly interesting.

1 Cheap Talk

Given an actual game with its players, sets of their feasible strategies, and their payoff functions, a cheap talk game is one where one or more rounds of pre-play communication are added. This communication is cheap in two senses: it has no direct payoff consequences, and there is no enforceable requirement for it to be truthful. Players may attempt to communicate their private information, or their intended actions in the game to come. The question is whether this cheap talk can affect the choices and outcomes in that game, in the sense that for any Nash equilibrium of the game without communication, there is a subgame perfect Nash equilibrium of the game with communication that is different (and Pareto-better). Of course the game with communication always has a subgame perfect equilibrium that replicates any Nash equilibrium without communication: at the stage of actual play the receiver ignores the communication, and rolling back, the sender sends any arbitrary message. This is called a "babbling" equilibrium; the focus of interest is on whether non-babbling equilibria are possible.

This may seem doubtful. Unlike Spencian signals that are differentially costly to send and therefore may serve to separate players with different private information (different "types"), cheap talk has no cost, let alone cost differentials across types. And without any enforcement of truthtelling, nothing seems to stop a bad type from pretending to be a good type, or a player sending deceitful messages about what his intentions. However, this omits consideration of possible alignment of interests of the players. In the extreme case where their objective functions are identical (perfect alignment of interests), there is every reason for a better-informed player to reveal his information truthfully, and for all players to communicate their intentions about actions truthfully.[1] In the opposite extreme, if the game is constant-sum (perfect opposition of interests), any statement by another player should be ignored. The interesting cases are in the spectrum between these extremes, where interests are partially aligned and partially in conflict.

In a pathbreaking paper, Crawford and Sobel (1982) demonstrated the possibilities and limits of cheap talk. In their framework, one player (the sender) has private information about the game to come, in the form of a number from a given interval, and can send to the other player (the receiver) a message that is also a number, which may or may not be the true number. If the players' interests are partially aligned, it is possible to convey partial information truthfully, in the sense that there is a partition of the interval into a number of subintervals, such that the message "the true number is in the subinterval k" can be truthfully conveyed by the sender and believed by the receiver. The better the alignment of the players' interests, the finer the partition can be.

This model is well known albeit beyond most undergraduate curricula, and there are good elementary surveys of related literature, for example Farrell and Rabin (1996). I will briefly mention one (two audiences) in Session V; here I focus on just one new interesting contribution, namely Aumann and Hart (2003) on multi-stage cheap talk. The question is to examine when and how more can be done with multiple stages than with just one.

They introduce a mnemonic convention that you will find useful in many other contexts. In their two-player games, the row player is named Rowena and the column player is Colin. In their paper, the actual game is chosen by "nature" from among several possibilities. Rowena knows which game is the true one, and the two players can exchange messages as long as they wish. Thus there is a potential infinite sequence of talk stages, followed by an action stage. In most of their examples, the point can be made with just two or three rounds of talk, but in the theory usually it is not possible to put a finite bound on this number in advance.

[1] Of course it is important that this alignment is common knowledge among the players; even slight departures from such common knowledge can alter the equilibria drastically, as we will see in the next section on global games.

The first example from Aumann-Hart that I will offer is shown in Table 1. If the true game is A, the players can achieve the expected payoff (4,4) by a round of "compromise" talk where they agree to activate a public randomization device that generates one of two equally likely events (toss of a fair coin), and then play (U,L) or (D,R) depending on the outcome. If the true game is B, they agree that Colin plays Z, yielding payoffs (4,4). When only Rowena knows the true game, without any communication the Nash equilibrium would be for Colin to play R, yielding expected payoffs (1,3) to the two. With communication, they can still get (4,4) but this needs two rounds of communication: first Rowena tells Colin what game it is, and then if it is A, they agree to compromise as above. The communication must occur in this order. Suppose the true game is A, but they agree to randomize first, and the device calls on Colin to play R. Then truthful disclosure in the second communication round would yield 2 to Rowena. Knowing this, she would be tempted to say that the game is B, in which Colin is to play Z, yielding Rowena 3. And knowing this in turn, Colin would not believe Rowena's report.

Table 1: Two-Stage Cheap Talk – Signal, then Compromise

			Colin L	Colin R	Colin Z
Rowena	Game A, Prob. $\frac{1}{2}$	U	6 , 2	0 , 0	3 , 0
		D	0 , 0	2 , 6	3 , 0
	Game B, Prob. $\frac{1}{2}$	U	0 , 0	0 , 0	4 , 4
		D	0 , 0	0 , 0	4 , 4

My second example from Aumann-Hart is in Table 2. With no communication at all, Colin would choose M resulting in expected payoffs (0,5). Suppose two channels of communication exist, this is common knowledge, and Colin can observe which channel Rowena is using. One channel is 100% accurate. The second is noisy: it conveys Rowena's message unaltered with probability 0.75, and reverses it with probability 0.25. With one round of communication, Rowena can get 3 using a "partial revelation" strategy. She uses the noisy channel. If this is to be an equilibrium where Rowena sends a truthful message, then using

Bayes' formula, Colin's posterior probabilities are that the game is what Rowena reports with probability 0.75, and it is the opposite with probability 0.25. If he receives signal A, his best action is to choose L, while if he receives B, it is to choose R, in each case for the expected payoff 6. Then Rowena stands to get 3 no matter what she reports, but with the usual "goodwill" assumption of breaking ties in favor of the opponent, there is a subgame perfect equilibrium of the cheap talk game with expected payoffs (3,6), which is Pareto superior to (0,5).

Table 2: Two-Stage Cheap Talk – Compromise, then Signal

			Colin				
			LL	L	M	R	RR
Rowena	Game A, Prob. $\frac{1}{2}$	U	1 , 10	3 , 8	0 , 5	3 , 0	1 , − 8
	Game B, Prob. $\frac{1}{2}$	D	1 , − 8	3 , 0	0 , 5	3 , 8	1 , 10

But there is another cheap talk equilibrium supported by its own beliefs. This one has full revelation: Rowena uses an error-free channel and reports the game truthfully, and Colin plays LL or RR according as whether she reports A or B. Payoffs are (1,10).

Each prefers a different one-stage talk game, but can compromise further using a two-stage cheap talk game: compromise in the first stage of talk, using a public randomization device to decide which communication game to play, resulting in payoffs that are an average between (1,10) and (3,6). For example tossing a fair coin yields expected payoffs (2,8). Again the order of the conversation stages matters.

Readings

Robert J. Aumann and Sergiu Hart. 2003. "Long cheap talk." *Econometrica* 71(6), November, 1619–1660.

Vincent P. Crawford and Joel Sobel. 1982. "Strategic information transmission."' *Econometrica* 50(6), November, 1431–1451.

Joseph Farrell and Matthew Rabin. 1996. "Cheap talk." *Journal of Economic Perspectives* 10(3), Summer, 103–118.

2 Global Games

In reality, people can never be absolutely sure that others share exactly the same information about their interaction. How sensitively does the equilibrium outcome of a game depend on the assumption that the game – its full specification including the payoff functions – is common knowledge among the players? The question is made trickier by the fact that if you think it possible that another player has a lower estimate of some payoff than you do, you must also entertain the possibility that he thinks that you have an even lower estimate of the same payoff than that. The logic continues to further steps of such doubt. Can such fears of misestimation compound to produce a very different set of actions and outcomes? The theory of global games examines such situations, and finds that may indeed be the case.

As usual, the general theory is complicated, but the issues are very well illustrated by the earliest examples developed by Carlsson and Van Damme (1993) and surveyed and exposited by Morris and Shin (2003).

2.1 Two-person Investment Game

Begin with a two-person symmetric game of investment, with the payoff matrix in Table 3. If $\theta < 0$, then `NotInvest` is dominant for both firms, and if $\theta > 1$, then `Invest` is dominant for both. If $0 < \theta < 1$, this is an assurance game with two pure strategy equilibria – each player wants to invest if and only if the other invests – and (`Invest`,`Invest`) is the Pareto-preferred. The actual θ is drawn from a known (and common knowledge; to avoid awkward writing I will henceforth leave this additional requirement implicit and unstated except when it is important to draw attention to it) distribution. If the realization of θ is publicly observable, the mutually preferred equilibrium is likely to be focal as soon as $\theta > 0$.

But now suppose the actual realization is not publicly observed. Instead, each player gets a signal with some error. Player i's signal is

$$x_i = \theta + \epsilon_i \, ;$$

each knows that the other is getting such a signal. Suppose θ is known to come from a normal distribution with mean y and variance $\mathrm{V}[\theta]$. (Thus all possibilities – `NotInvest` dominant, `Invest` dominant, and the assurance game with two equilibria – can arise with positive probability.) The errors are known to be normal with means 0 and variances $\mathrm{V}[\epsilon]$, and independent of θ. The usual regression formula gives, from player 1's perspective

$$\mathrm{E}[\theta \,|\, x_1\,] - y = \frac{\mathrm{V}[\theta]}{\mathrm{V}[\theta] + \mathrm{V}[\epsilon]} \; (\, x_1 - y\,) \, .$$

Table 3: Investment Game

		Player 2	
		`Invest`	`NotInvest`
Player 1	`Invest`	θ, θ	$\theta - 1$, 0
	`NotInvest`	0, $\theta - 1$	0, 0

and

$$\mathrm{V}[\theta \,|\, x_1\,] = \frac{\mathrm{V}[\theta]\ \mathrm{V}[\epsilon]}{\mathrm{V}[\theta] + \mathrm{V}[\epsilon]}\,.$$

At this point, the mathematics is greatly simplified if we consider the limit where $\mathrm{V}[\theta]$ is infinite. That means the prior distribution of θ is improper – uniform over $[-\infty, \infty]$. Students may complain about this, but they may like the underlying substantive assumption that the players are completely ignorant about θ. In any event, the assumption can be justified with mathematical rigor; Morris and Shin show how, and also treat the more general case with finite $\mathrm{V}[\theta]$ in their Section 3. The great simplification from assuming $\mathrm{V}[\theta] = \infty$ is that:[2]

$$\mathrm{E}[\theta \,|\, x_1\,] = x_1 \qquad \text{and} \qquad \mathrm{V}[\theta \,|\, x_1\,] = \mathrm{V}[\epsilon]\,.$$

Then, iterating expectations,

$$\mathrm{E}[\,x_2 \,|\, x_1\,] = x_1 \qquad \text{and} \qquad \mathrm{V}[\,x_2 \,|\, x_1\,] = 2\ V[\epsilon]\,.$$

The distribution of x_2 conditional on x_1 is therefore normal with mean x_1 and variance $2\,\mathrm{V}[\epsilon]$. Its cumulative distribution function is

$$\Phi\left(\frac{x_2 - x_1}{\sqrt{2}\ \mathrm{V}[\epsilon]}\right)\,,$$

where $\Phi(.)$ is the cumulative distribution function of a standard normal variate.

[2]Note that when $\mathrm{V}[\theta] = \infty$ the conditional expectation is the same regardless of y; thus the specification of the unconditional expectation becomes irrelevant.

The natural equilibrium to seek for this game is given by a threshold k and a strategy specifying action as a function of the player's signal x:

$$\texttt{Invest} \text{ if } x > k$$
$$\texttt{NotInvest} \text{ if } x \leq k$$

Suppose player 2 is following such a strategy with threshold k_2. What is player 1's best response? Upon receiving the signal x_1, calculates the probability that player 2 invests:

$$\Pr\{\, x_2 > k_2 \,|\, x_1 \,\} = 1 - \Phi\left(\frac{k_2 - x_1}{\sqrt{2}\ \mathrm{V}[\epsilon]}\right).$$

His expected payoff is $\mathrm{E}[\,\theta\,|\,x_1\,] = x_1$ if the other invests, and $\mathrm{E}[\,\theta\,|\,x_1\,] - 1 = x_1 - 1$ if the other does not invest. Therefore player 1 calculates his expected payoff as

$$x_1 \left[1 - \Phi\left(\frac{k_2 - x_1}{\sqrt{2}\ \mathrm{V}[\epsilon]}\right)\right] + (x_1 - 1)\ \Phi\left(\frac{k_2 - x_1}{\sqrt{2}\ \mathrm{V}[\epsilon]}\right) = x_1 - \Phi\left(\frac{k - x_1}{\sqrt{2}\ \mathrm{V}[\epsilon]}\right).$$

Player 1 will invest if and only if this is positive. It is easy to see that the expected payoff is increasing in x_1; it is negative when $x_1 = 0$ and goes to ∞ as x_1 goes to ∞. Therefore for each k_2 there is a unique positive value of x_1, call it k_1, such that player 1's expected payoff is positive if and only if $x_1 > k_1$. Thus player 1's best response is also a threshold strategy. The best response function is found by solving

$$k_1 - \Phi\left(\frac{k_2 - k_1}{\sqrt{2}\ \mathrm{V}[\epsilon]}\right) = 0\,.$$

for k_1 in terms of k_2. Figure 1 shows this function; this is Figure 3.1 from Morris and Shin (2003). It is easy to see from the figure, and to verify algebraically by implicit differentiation, that its slope is positive and less than 1 everywhere.

There is a similar best response function for player 2. Equilibrium occurs at the intersection of the two. Because of the slope properties of the best responses, there is a unique equilibrium, and it is easily verified to be symmetric with

$$k_1 = k_2 = \tfrac{1}{2}\,.$$

Thus, even though mutual investment is Pareto-preferred as soon as $\theta > 0$, it will not occur in equilibrium unless the players' signals are both $> \frac{1}{2}$. And this large difference opens up no matter how small is the variance $\mathrm{V}[\epsilon]$ of the errors in the signals! To see the intuition

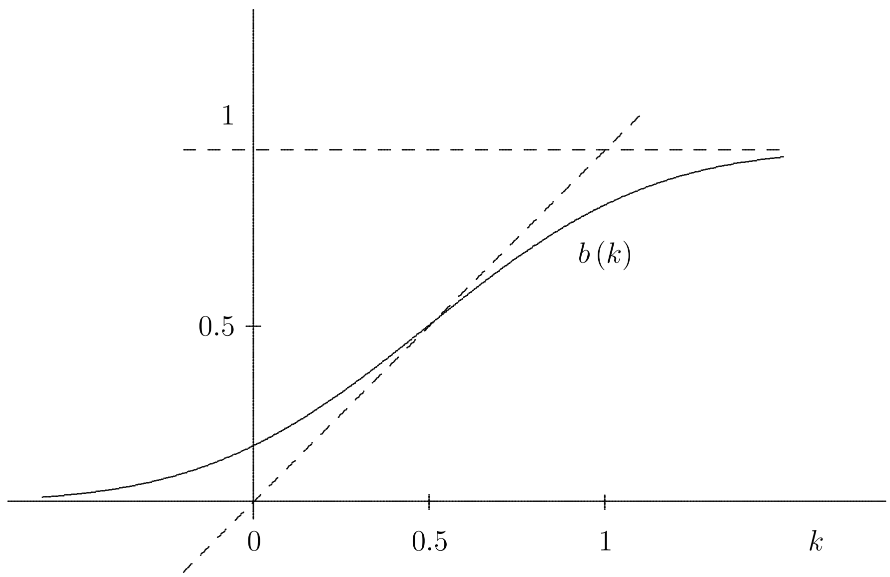

Figure 1: Best-response Thresholds

for this, suppose player 2 is using the threshold zero, as he may if θ could be observed perfectly and with common knowledge. But with our more realistic assumption of the slight failure of common knowledge, player 1 would find it optimal to respond to such a strategy of player 2 by being more cautious and using a positive threshold, as shown in the figure by the intercept of the best response function on the vertical axis. Of course player 2 should recognize this, which would then make it optimal for him to be even more cautious, as we see by iterating the best response one more time. This succession of best responses converges to $\frac{1}{2}$. That is how the possibility that the other player may have received a negative signal, or may think you have, ... exerts its influence on the outcome, even with very small error variance. With a little lack of common knowledge of what game is being played, extreme possibilities can have substantial effect on outcomes. This also motivates the use of the title "global games" for this branch of the theory.

The same argument can be made even more forcefully using dominance solvability (iterated elimination of dominated strategies) or rationalizability (iterated elimination of strategies that are never best responses). Player 2 is not going to invest if his signal is negative, because then he calculates $\mathrm{E}[\theta \,|\, x_2\,] < 0$ and `Invest` becomes his dominated action. But then player 2 should not invest when his signal is less than the vertical axis intercept of the best response function. And so on.

An interesting non-economic application of the same idea is formalization of Schelling's discussion of surprise attack, where one country (or gunslinger) shoots first in "self-defense," out of fear that the other was about to shoot in "self-defense" thinking that ... ; see Van Damme (1997).

2.2 Many-person Collective Action Game

The same ideas also extend to many-person games and problems of collective action. Consider a continuum of players, each of whom chooses whether to invest. The payoff from investing is $\theta + l - 1$, where l is the proportion of others who invest. Thus, given θ, it becomes optimal to invest if enough others are also investing ($l > 1-\theta$), and everyone investing is the Pareto optimal outcome if $\theta > 0$. But as before, θ is not publicly observable, and each player i gets a signal $x_i = \theta + \epsilon_i$. As before, θ is normal with infinite variance, each ϵ_i is normal with zero mean and variance $\mathrm{V}[\epsilon]$, and the ϵ_i are independent across players. (There are technical issues about using such probabilistic notions with a continuum of random variables, but they can be handled rigorously, and at the level of undergraduate exposition are best left out.)

Consider a symmetric equilibrium where each player invests if and only if his signal exceeds a threshold k. This means that any one player should be indifferent between investing and not investing when he gets signal k, doing the calculation assuming that everyone else is following the same strategy. If his signal is h, doing the same calculation as in the two-person case, his conditional expectation of θ is $\mathrm{E}[\,\theta \,|\, h\,] = h$, and he expects others' signals to be distributed normally with mean h and variance $2\,\mathrm{V}[\epsilon]$. Therefore his conditional probability that any one other player j invests is

$$1 - \Pr\{\, x_j \leq k \,\} = 1 - \Phi\left(\frac{h-k}{\sqrt{2}\,\mathrm{V}[\epsilon]} \right) .$$

Invoking the law of large numbers for the continuum, this is also the fraction of others that invest. Therefore the player we are focusing on should invest if

$$h + \left[\, 1 - \Phi\left(\frac{h-k}{\sqrt{2}\,\mathrm{V}[\epsilon]} \right) \,\right] - 1 > 0 \,.$$

For this to be indifference at $h = k$,

$$k = \Phi(0) = \tfrac{1}{2} \,.$$

Exactly as in the two-person case, even a small error in the signals weighs heavily on the outcome.

2.3 Risk-Dominance

The general theory is of course much more complex, but has an important implication for equilibrium selection. Consider the investment game above, but modify it to make it common knowledge that $0 < \theta < 1$. The actual value of θ picked by "nature" is still not revealed to the player, let alone common knowledge between them; each gets a private signal about θ with some small random error. Thus it is common knowledge that the game has two Nash equilibria in pure strategies, and that the (`Invest`,`Invest`) equilibrium is Pareto-better than the (`NotInvest`,`NotInvest`) equilibrium. Despite this, a little lack of common knowledge may prevent it from being selected.

Consider a more general game of this kind, shown in Table 4.

Table 4: Risk-Dominance In Assurance Game

		Player 2	
		`A`	`B`
Player 1	`A`	$a_1,\ a_2$	$d_1,\ c_2$
	`B`	c_1, d_2	$b_1,\ b_2$

Suppose $a_1 > c_1$, $a_2 > c_2$, and $b_1 > d_1$, $b_2 > d_2$, so (`A,A`) and (`B,B`) are both Nash equilibria. Suppose further that $a_1 > b_1$, $a_2 > b_2$ so (`A,A`) is Pareto-preferred, or *payoff-dominant* over (`B,B`). All this is common knowledge, but within that there is the slight uncertainty about the actual values of the payoffs.

For each equilibrium, calculate the loss of payoff a player would suffer if he deviates unilaterally, and calculate the product of the two players' losses. The equilibrium for which the product is higher is defined to be the *risk-dominant* one. For the (`A,A`) the losses are $(a_1 - c_1)$ and $(a_2 - c_2)$; for (`B,B`) they are $(b_1 - d_1)$ and $(b_2 - d_2)$. Therefore the (`B,B`) equilibrium is risk-dominant if

$$(a_1 - c_1)\,(a_2 - c_2) < (b_1 - d_1)\,(b_2 - d_2)\,. \tag{1}$$

The appeal of risk-dominance comes from two results. Carlsson and van Damme (1993) showed that under quite general conditions, iterated dominance in a global game with slight lack of common knowledge forces selection of the risk-dominant equilibrium, and Young (1998, Theorem 4.1) showed that adaptive learning dynamics in a game with such small errors converge to it.

In the investment game, the losses of the two players are θ each for the (`Invest`,`Invest`) equilibrium and $(1-\theta)$ each for the (`NotInvest`,`NotInvest`) equilibrium. Thus the latter is risk-dominant if

$$\theta^2 < (1-\theta)^2\,, \quad \text{or} \quad \theta < 1-\theta\,, \quad \text{or} \quad \theta < \tfrac{1}{2}\,.$$

Therefore, if $0 < \theta < \frac{1}{2}$, the (`NotInvest`,`NotInvest`) equilibrium may get selected even if (`Invest`,`Invest`) provides better payoffs for both players.

To get a rough intuition for the selection of a risk-dominant equilibrium, calculate player 1's net gain from playing `A` rather than `B` when player 2 is mixing using probability α_2 for his action `A`:

$$[\ \alpha_2\ a_1 + (1-\alpha_2)\ d_1\ ] - [\ \alpha_2\ c_1 + (1-\alpha_2)\ b_1\ ] = \alpha_2\ [\,(a_1 - c_1) + (b_1 - d_1)\,] - (b_1 - d_1)$$

Let α_2^* denote the value of α_2 that sets it equal to 0. (This is just player 2's mixture probability that makes player 1 indifferent between his two pure actions. Recall that the assurance game has a third Nash equilibrium in mixed strategies; in fact α_2^* is just the equilibrium mixture probability for player 2 in that equilibrium.) Similar calculations yield player 1's mixture α_1^*.

If α_2^* is small, then there is a large range of player 1's beliefs about payoffs for which he will choose action `A`. Similarly for player 1. It is easy to verify that the risk-dominance criterion (1) is equivalent to

$$\alpha_1^* + \alpha_2^* < 1\,,$$

so it corresponds to α_1^* and α_2^* being small, and thus provides a rough rationale for the selection.

While the risk-dominance criterion for equilibrium selection has rigorous theoretical support based on the possibility of small failures of common knowledge, other equilibrium selection criteria exist, and some experimental evidence favors a very simple "level-1" procedure where each player simply regards all actions of the other player as equally likely (a "Laplacian" belief in the face of ignorance) and chooses his action in response to maximize his own expected payoff; see Camerer (2003, pp. 398–399). In the investment game, if player 1 thinks that player 2 is equally likely to invest or not, his expected payoff from investing is

$$\tfrac{1}{2}\ \theta + \tfrac{1}{2}\ (\theta - 1) = \theta - \tfrac{1}{2}\,,$$

and that from not investing is 0, so he should invest if $\theta > \frac{1}{2}$. In this example the two criteria (risk-dominance and Laplacian reasoning) coincide, and both go against the simple assumption that the Pareto-dominant equilibrium should be focal. Morris and Shin (2003) discuss the relationship between the theory of global games and the Laplacian assumption.

The result that very small failures of common knowledge can have drastic consequences for the outcome of a game, and can prevent achievement of a Pareto-superior outcome in an assurance game, can help us understand why societies develop elaborate procedures to make the Pareto superiority a matter of common knowledge. Chwe (2001) offers an excellent analysis and examples of these procedures, for example public announcements where the participants are arranged in an inward-facing circle so all can be sure that all are hearing the same message.

Readings

Carlsson, Hans and Eric Van Damme. 1993. "Global games and equilibrium selection." *Econometrica* 61(5), September, 989–1018.

Colin Camerer. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction.* New York, NY: Russell Sage Foundation and Princeton, NJ: Princeton University Press.

Michael Chwe. 2001. *Rational Ritual: Culture, Co-ordination and Common Knowledge.* Princeton, NJ: Princeton University Press.

Morris, Stephen and Shin, Hyun Song. 2003. "Global games: Theory and applications." In *Advances in Economics and Econometrics: Theory and Applications (Eighth World Congress of the Econometric Society),* eds. Mathias Dewatripont, Lars Peter Hansen and Stephen J. Turnovsky, Cambridge, UK and New York, NY: Cambridge University Press.

Van Damme, Eric. 1997. "Equilibrium selection in team games." In *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten,* eds. Alberts, Güth, Hammerstein, Moldovanu and Van Damme, Springer-Verlag.

Young, H. Peyton. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions.* Princeton, NJ: Princeton University Press.

3 Repeated Games

Many interactions in reality extend beyond a single encounter, and this raises the possibility of influencing behavior using the shadow of the future, often for mutual benefit of the players as in the prisoners' dilemma. The theory has evolved far beyond the simple examples given in elementary textbooks. Here I give a very brief statement of some of the methods and results.

A purely repeated game is the same game played for many stages, with no change in the specification of the game from one stage to the next. Each stage game is specified by [1] the players and their feasible actions, [2] the payoff functions. These should stay the same in the play at any stage regardless of what happened in previous stages, although of course the actual actions chosen may depend on the available history of play in past stages and the anticipation of actions in future stages. Many situations of infinitely or indefinitely ongoing interactions are modeled as repeated games; most prominently, self-enforcing resolution of prisoners' dilemmas by creating implicit cooperation among selfish players is usually explained using this framework.

Such pure repetition should be distinguished from a dynamic situation in which actions taken now affect what actions are *available or feasible* at later stages, and the payoff *functions* specifying how future outcomes depend on future actions. Such dynamic games are the staple of Schelling's strategic moves – commitments, threats, and promises – where actions at one stage are chosen to preclude some future choices or to alter their costs and benefits for other players or for one's own future self. Analyses and solutions for dynamic games in simple settings are well known in various contexts. With infinite or indefinite repetition, such games usually have a state variable that affects the payoff functions at each stage, and whose "law of motion" is a differential or difference equation. Such fully dynamic games are harder to analyze. For repeated games, however, much more general and powerful techniques have become available. Here I offer a very brief summary of some of them.

The analysis of repeated games is simplified by the fact that the continuation game after any history of previous plays looks exactly like the original game, again so far as the feasible strategies and payoff functions are concerned. This allows the use of dynamic programming methods in a stationary environment.

3.1 Notation

The game played at each stage is defined by

[1] the set of players labelled $i = 1, 2, \ldots\ n$;

[2] their sets of available actions A_i, together forming a subset $\mathbf{A}$ of the n-dimensional space R^n, with typical element $\mathbf{a} = (a_1, a_2, \ \ldots\ a_n)$;

[3] the payoff function of each, $\Pi_i(\mathbf{a})$, forming the vector payoff function $\mathbf{\Pi} : \mathbf{A} \rightarrow R^n$

This stage game is repeated infinitely, with stages or periods labeled $t = 0, 1, 2, \ldots$. The discount *factor* δ applies between any two successive stages or periods. Just as a reminder and for emphasis, $\delta = 1/(1+r)$ where r is the discount *rate*. If a sequence of plays yields a player the sequence of payoffs $u(t)$ for $t = 0, 1, 2, \ldots$, write the present value of the payoff from the repeated game as

$$\begin{aligned} V(0) &= u(0) + \delta\ u(1) + \delta^2\ u(2) + \ldots \\ &= u(0) + \delta\ V(1)\,, \end{aligned} \tag{2}$$

where $V(1)$ denotes the present value at time 1 of the payoff sequence starting with $u(1)$.

The case of high patience or high regard for the future corresponds to r being close to 0 or δ being close to 1. A lot of the theory of repeated games deals with the limit as $\delta \rightarrow 1$, or by suitable redefinition of the objective function, even the case of total patience, $\delta = 1$. This gives rise to the so-called folk theorems that show all the payoff outcomes that can be sustained as subgame perfect equilibria of the repeated game. Your students will be rightly skeptical of the assumption of extremely high patience. Therefore I will not go into the recent research that continues to produce ever more new folk theorems, but will stick to characterization of what is possible for a given $\delta < 1$.

Cooperation in a repeated game can be sustained by punishing the deviators at any one stage in some or all of future stages. In this line of theory, the best cooperation can be sustained by threat of the worst punishment. However, these threats must be credible, and that is usually interpreted in the sense of subgame perfectness. Designing severe subgame perfect punishments is the name of the game.

The algebra of the theory is easier to express and more intuitive in terms of the hypothetical constant flow of payoffs v that will yield the same present value as a given V. Using (2) when all the $u(t)$'s are equal to v, we see at once that $v = (1-\delta)\,V$. So let $v(0)$ be the hypothetical constant flow equivalent to $V(0)$ etc, we can rewrite (2) as

$$v(0) = (1-\delta)\ u(0) + \delta\ v(1)\,. \tag{3}$$

This is the form I will use in what follows.

Two qualifications should be mentioned up front. [1] Subgame perfectness may not be enough for credibility. After the fact of a deviation, all players may do better to forgo the punishment and make a fresh start to the repeated game. In other words, the punishment may not be renegotiation-proof. Of course one episode of renegotiation and forgiveness may merely encourage future cheating, so the players may find it better ex ante to commit to not renegotiating if they can do so. There are various alternative formulations of the renegotiation concept, which attempt to deal with these problems in different ways, but none of them finds universal acceptance. So I will leave it to you to explore them at your leisure. [2] The most severe punishment a game theorist can devise, often consisting of trigger strategies that end all cooperation at the first instance of any deviation, are rarely observed in practice. This is especially emphasized by Elinor Ostrom in her research on the reality of collective action. She finds that punishments start small, and are stepped up only in the face of repeated infractions. But theoretical explanations of graduated punishments are not yet well formulated.

The worst punishment that can be inflicted on player i in any stage game is for all the others to gang up and choose coordinated actions to minimize i's payoff, knowing that i is going to act as best he can to maximize it. Call the result $\underline{u}_i$ the *security level* of player i:

$$\underline{u}_i = \min_{a_j \in A_j \text{ for all } j \neq i} \; \max_{a_i \in A_i} \; \Pi_i(a_1, a_2, \ \ldots \ a_n) \, . \tag{4}$$

The basic (two actions, perfect observation) prisoners' dilemma game has a very special property in this regard: the outcome that holds all players down to their security levels is the unique Nash equilibrium of the stage game. Therefore reversion to Nash is automatically the harshest subgame perfect punishment. In other games, this is not so. Therefore reversion to Nash does not in general support Pareto-preferred outcomes as well as more cleverly designed punishments. Characterizing the best feasible punishments to support the best feasible cooperation is the aim of this line of theory. However, it can be extended to include some concepts of renegotiation by imposing appropriate renegotiation-proofness constraints on the sets of feasible payoffs.

3.2 Unimprovability

The key to many of the results on repeated games is the idea that if the stage game has bounded payoffs, and a player can gain by deviating from a specified strategy[3] in the repeated

[3] Remember that a strategy is a complete plan specifying his actions at all nodes (or to be more precise, at all information sets) that can arise in logically conceivable paths through the game tree.

game with discounting, then he can gain from a deviation at just one node. (The converse, of course, is trivial.)

To see this, first note that because the payoffs are bounded and are being discounted, the present value of the gain from deviations that are far distant in time becomes arbitrarily small. Therefore, if any complicated deviation at multiple nodes at an infinite sequence of periods produces a positive discounted present value gain to the player, so does a deviation at a finite number of nodes. Consider a deviation that involves the smallest number of nodes, and its last time of deviation. Then that last deviation must bring a gain by itself; otherwise deviation at a smaller number of nodes would have been profitable. But then we have found the gainful one-shot deviation we seek.

A strategy that cannot be bettered (for the player playing it) with a one-shot deviation is called *unimprovable*.

3.3 Punishments Better Than Nash-Reversion: Example

Abreu (1988) offers the example of a Cournot game where each of two firms has the choice of three levels of output, low (L), medium (M), and high (H). The payoff matrix for each stage game is shown in Table 5. The stage game has a unique Nash equilibrium, (M,M) with payoffs (7,7). To sustain the collusive strategy profile (L,L) with payoffs (10,10) using the threat of Nash reversion, each firm's gain from one-period deviation (we can restrict attention to such deviations because of the property of unimprovability) to M, namely $15 - 10 = 5$, must be less than the present value of the subsequent loss of $10 - 7 = 3$ every period, or

$$5 < 3 \ \left(\delta + \delta^2 + \delta^3 + \dots \right) = 3 \ \delta/(1 - \delta) \, ,$$

or $\delta > 5/8 = 0.625$.

Consider an alternative punishment structure: If firm 1 deviates from its prescribed strategy, the play will switch to (M,H) for T periods and (L,M) thereafter. The effects of firm 2's deviation are symmetric. Any deviation from the prescribed actions, including participation in the punishment of the other firm, is itself a deviation and will be punished starting in the very next period (so any previously ongoing punishment will be abandoned). As this specifies the punishments for both firms after all conceivable deviations from prescribed actions, it should be called a complete "penal code."

Table 5: Payoff Matrix of Three-output Cournot Example

		Firm 2		
		L	M	H
	L	10 , 10	3 , 15	0 , 7
Firm 1	M	15 , 3	7 , 7	− 4 , 5
	H	7 , 0	5 , − 4	− 15 , − 15

Check when this will sustain a prescribed cooperative path of playing (L,L) all the time. In a situation where no deviations have occurred so far, firm 1 would see the following payoff consequences of the two actions it could take:

$$\begin{array}{lccccccc} \text{Conform:} & 10 & 10 & \dots & 10 & 10 & 10 & \dots \\ \text{Deviate:} & 15 & -4 & \dots & -4 & 3 & 3 & \dots \\ & & \multicolumn{3}{c}{\underbrace{T \text{ periods}}} & & & \end{array}$$

Then firm 1 will not want to deviate if the first-period gain, 5, is less than the discounted present value of future losses:

$$[10-(-4)]\ \left(\delta+\delta^2+\ldots+\delta^T\right)+(10-3)\ \left(\delta^{T+1}+\ldots\right)=14\ \frac{\delta-\delta^{T+1}}{1-\delta}+7\ \frac{\delta^{T+1}}{1-\delta}\,,$$

which simplifies to

$$5<\delta\ \left(19-7\ \delta^T\right)\,.$$

This is the condition for a simple Nash equilibrium in the repeated game, but it is not enough for subgame perfectness. For that, we also need to check that the continuation after any deviation, namely the punishment path, is itself a Nash equilibrium. This raises two questions: [1] Why should firm 1 accept its punishment? [2] Does firm 2 have the incentive to participate as prescribed in the punishment of firm 1?

Firm 1 could try to respond to firm 2's punishment action H by choosing L instead of its own prescribed M to get 0 instead of −4. But that would count as a fresh deviation and restart the punishment with a fresh sequence of −4's for T periods, which would merely postpone the start of the better outcome 3:

$$
\begin{array}{lccccccc}
\text{Conform:} & -4 & -4 & \dots & -4 & 3 & 3 & \dots \\
\text{Deviate:} & 0 & -4 & \dots & -4 & -4 & 3 & \dots \\
 & & \multicolumn{4}{c}{\underbrace{\quad T \text{ periods} \quad}} & &
\end{array}
$$

Firm 1 does not want to deviate if $4 < 7\,\delta^T$. This will indeed turn out to be the crucial constraint.

Firm 2 refuses to play its prescribed part in the punishment (M,H) of firm 1, and instead play its own M to get 7 instead of 5. But this will count as a deviation of is own. Firm 1's punishment will be abandoned, and a phase of punishment for firm 2 will start, giving it -4 instead of 5 for T periods, and then only 3 instead of 15:

$$
\begin{array}{lccccccc}
\text{Conform:} & 5 & 5 & \dots & 5 & 15 & 15 & \dots \\
\text{Deviate:} & 7 & -4 & \dots & -4 & -4 & 3 & \dots \\
 & & \multicolumn{4}{c}{\underbrace{\quad T \text{ periods} \quad}} & &
\end{array}
$$

This is indeed an unattractive prospect, easy to rule out for the values of δ that meet the other conditions.

The best way to satisfy the condition for firm 1's going along with its punishment is to take $T = 1$, so requiring $\delta > 4/7 = 0.571$. Then the condition for Nash equilibrium is satisfied: $5 < (4/7) * (19 - 7 * 4/7) = 60/7$. In fact, solving the quadratic inequality $5 < \delta\,(19 - 7\,\delta)$ shows that the Nash condition needs only $\delta > 0.295$.

Therefore the proposed penal code yields a subgame perfect equilibrium for a larger range of discount factors than Nash reversion can: values of δ in the interval $(4/7\,, 5/8) =$ (0.571,0.625) are added.

Observe that when $\delta = 4/7$, the present value of the payoff for a firm in the punishment phase is

$$-4 + 3\,\delta + \delta^2 + \ldots = -4 + 3\,\delta/(1 - \delta) = 0\,,$$

which is the worst possible that can be inflicted on a firm. (The minimax in each stage game is 0.) In comparison, Nash reversion would give the present value payoff 7 in the punishment phase.

This example illustrates several general properties of optimal penal codes that Abreu establishes: [1] Punishments are player-specific rather than deviation-specific; any deviation is punished in the same manner. [2] Punishments have a stick-and-carrot structure; a more severe early phase is followed by a less severe later phase to create the incentive to participate in one's own punishment. [3] Failure to participate in another firm's punishment is itself a deviation, leading to an immediate start of one's own punishment.

3.4 Self-generation

In a repeated game, the actions of a player at time t are influenced not only by his calculation of the payoffs at that time, but also by his calculation of the payoffs that will accrue in future periods in so far as they are affected by his current action. In our present context, the most pertinent consideration of this kind is the punishment he can be made to suffer as a result of any deviation from a prescribed "cooperative" action. We are requiring that these punishments be sustainable as a subgame perfect equilibrium in the game that will commence the next period. But the whole game is a repeated game with constant discount factor; therefore the continuation game looks just like the original game.

The strategies that constitute a subgame perfect equilibrium in the repeated game could in general be very complicated, but what matters to the player is the resulting payoff. Let $\mathbf{V}$ denote the set n-dimensional vectors of the n players' payoffs that can arise from all its subgame-perfect equilibria. Thus $\mathbf{v} = (v_1, v_2, \ldots v_n) \in \mathbf{V}$ if and only if there exists a vector of immediate actions $\mathbf{a} \in \mathbf{A}$, and a function $\mathbf{U} : \mathbf{A} \rightarrow \mathbf{V}$ such that [1] $\mathbf{U}$ is the result of the stage game payoffs $\mathbf{\Pi}(\mathbf{a})$ and the (credibly threatened or promised, subgame perfect) continuation payoff values $\mathbf{U}(\mathbf{a})$, that is, for all players i,

$$v_i = (1-\delta)\ \Pi_i(\mathbf{a}) + \delta\ U_i(\mathbf{a})\,, \tag{5}$$

or in vector form collected for all players,

$$\mathbf{v} = (1-\delta)\ \mathbf{\Pi}(\mathbf{v}) + \delta\ \mathbf{U}(\mathbf{a})\,; \tag{\mathbf{5}}$$

and [2] no player can do better by unilaterally deviating from his prescribed action, that is, for all i and for all $a'_i \in A_i$,

$$(1-\delta)\ \Pi_i(\mathbf{a}) + \delta\ U_i(\mathbf{a}) \geq (1-\delta)\ \Pi_i(\mathbf{a}'_i) + \delta\ U_i(\mathbf{a}'_i) \tag{6}$$

where $\mathbf{a}'_i$ is obtained by replacing the i^{th} component of $\mathbf{a}$ by a'_i leaving all other components unchanged. Note that the unimprovability result allows us to consider only one-shot deviations for any one player.

Thus payoffs in $\mathbf{V}$ can be achieved, or "generated," using threats of punishments based on continuation payoffs in $\mathbf{V}$, as specified by the function $\mathbf{U}$. More generally, we could think of punishments based on continuation payoffs in a set $\mathbf{W} \subseteq R^n$. These would generate a set of payoffs

$$(1-\delta)\ \mathbf{\Pi}(\mathbf{a}) + \delta\ \mathbf{U}(\mathbf{a})$$

for some $\mathbf{a} \in \mathbf{A}$ and some function $\mathbf{U} : \mathbf{A} \to \mathbf{W}$ satisfying (6) for all players i. Call this set $\mathbf{B}(\mathbf{W})$; thus $\mathbf{B}$ is a function that maps subsets of R^n to subsets of R^n. In particular, it maps the whole set of values in subgame perfect equilibria, namely $\mathbf{V}$, to itself; thus $\mathbf{V}$ is a fixed point of the mapping $\mathbf{B}$.

More generally, call a set $\mathbf{W}$ *self-generating* if $\mathbf{W} \subseteq \mathbf{B}(\mathbf{W})$, that is, if every payoff vector $\mathbf{W}$ can be achieve using values in $\mathbf{W}$ itself as continuation payoffs. Self-generating sets are a useful step in finding all subgame-perfect outcomes of the repeated game, because they have a key property which I state and explain heuristically; rigorous proofs are in Abreu, Pearce and Stacchetti (1990):

If $\mathbf{W}$ is self-generating, then $\mathbf{B}(\mathbf{W}) \subseteq \mathbf{V}$, that is, payoff vectors that can be sustained using punishments in $\mathbf{W}$ can be achieved in a subgame-perfect equilibrium.

Reasoning: For any payoff vector in $\mathbf{B}(\mathbf{W})$, there is a first-period action vector $\mathbf{a^0}$, and a continuation payoff $\mathbf{U}(\mathbf{a^0}) \in \mathbf{W}$ such that no player can gain by deviating from his component action in $\mathbf{a^0}$. Now the payoff starting at date 1, $\mathbf{U}(\mathbf{a^0})$, is in $\mathbf{W}$, which is $\subseteq \mathbf{B}(\mathbf{W})$ because $\mathbf{W}$ is self-generating. Thus $\mathbf{U}(\mathbf{a^0})$ is in $\mathbf{B}(\mathbf{W})$. Therefore there is a period 1 action $\mathbf{a^1}$ and a continuation payoff $\mathbf{U}(\mathbf{a^1})$ such that no player can gain by deviating from his component action in $\mathbf{a^1}$ at date 1. And so on. Thus we have found a complete strategy profile that is immune to one-period deviations, i.e. is unimprovable. Therefore it is immune to any deviations, i.e. it is subgame perfect, i.e. its payoff vector as of date 0 is in $\mathbf{V}$. Therefore $\mathbf{B}(\mathbf{W}) \subseteq V$.

3.5 Imperfect Monitoring

In the above analysis, an implicit assumption was that actions in any period were publicly observable at its end, so punishment strategies based on the history of such actions could be implemented in the continuation game starting with the next period. But the framework generalizes quite easily to the case where only some consequence of actions, perhaps also affected by some chance element, is publicly observable. Thus suppose a vector of actions $\mathbf{a}$ gives rise to a random variable y with possible values y_j and corresponding probabilities $p_j(\mathbf{a})$; the realization of y is publicly observable. Note that the possible values y can take (the support of the distribution) is independent of $\mathbf{a}$; otherwise some realizations would convey additional information. Each player's payoff in the stage game should depend only on the realization of y and its own action, and not separately on the actions of others; otherwise further information could be deduced from the payoff. Similarly, continuation payoffs can depend only on the realization of y. (There is recent work on repeated games with private monitoring, but the research is still ongoing and would take us too far afield.) Thus let the stage game payoff functions be $\pi_i(y, a_i)$ for player i; then the stage game expected payoffs are

$$\Pi_i(\mathbf{a}) = \sum_j \ \pi_i(y_j, a_i) \ p_j(\mathbf{a}) \, .$$

The continuation payoff functions $u_i(y)$ can only depend on the publicly observable outcome y; they imply expected continuation payoffs

$$U_i(\mathbf{a}) = \sum_j \ u_i(y_j) \ p_j(\mathbf{a}) \, .$$

With these redefinitions, the theory of self-generation can proceed as before, needing only some further technical care in the rigorous proofs.

This brief account is at best a bare beginning of this theory. Further developments include: [1] An algorithm to find $\mathbf{V}$: start with an arbitrary large $\mathbf{W}$ and successively apply $\mathbf{B}(.)$. Then

$$\mathbf{V} = \bigcap_{k=0}^{\infty} \ \mathbf{B}^k(\mathbf{W}) \, .$$

[2] If a public randomization device is available, a self-generating set can be "convexified," and continuation payoffs can be taken to be extreme points of $\mathbf{V}$. For these and further results I must refer you to the original literature cited in the readings at the end of the section.

3.6 Application: Green-Porter Model

Green and Porter (1984) produced an important model of cartels. Their firms are in a repeated interaction. Each period, each firm chooses its quantity; then the market price is determined from an inverse demand function that has a random component which is independently and identically distributed over time. The firms' quantity choices are not observed by other firms, but the price is publicly observed by all. This is exactly the framework of the above model.

Green and Porter examined how much cooperation could be sustained using particular punishment strategies of the form: when the price falls below a specified threshold $\tilde{p}$, all firms will produce the non-cooperative Cournot-Nash quantities Q for a specified number T

of periods. The Abreu-Pearce-Stacchetti framework permits more general analysis. Here is an example, a much simplified version of Abreu, Milgrom and Pearce (1991).

Two firms can each produce either a low quantity (the cooperative action C) or a high quantity (the competitive or defecting action D). Start with a simple prisoners' dilemma stage game shown in Table 6. Thus π is the cooperative payoff, g is the gain from a unilateral cheating, and z is a negative payoff from being cheated, whose exact value is immaterial except that we need $\pi + g + z << 2\,\pi$ to rule out the case where alternating cheating may be better than constant cooperation.

Table 6: Simple PD game

		Firm 2	
		C	D
Firm 1	C	π , π	z , $\pi + g$
	D	$\pi + g$, z	0 , 0

Recall that in this game the Nash equilibrium serves to minimax both players, and therefore Nash reversion achieves as much cooperation as one ever can. It requires the one-period gain from cheating to be no better than the present value of future losses from the collapse of cooperation, or

$$g \leq \pi\,\left(\,\delta + \delta^2 + \ldots\right) = \pi\,\delta/(1-\delta), \quad \text{or} \quad \delta > g/(g+\pi)\,.$$

This can be expressed in the language of self-generation: the set of payoffs $\mathbf{W} = \{\,(0,0)\,(\pi,\pi),\,\}$ is self-generating if $\delta \geq g/(g+\pi)$. The argument goes as follows. To achieve (π,π), use the continuation payoff function

$$U(C,C) = (\pi,\pi), \quad \text{and} \quad U(C,D) = U(D,C) = U(D,D) = (0,0)\,.$$

This requires, for firm 1: for its conforming,

$$\pi = (1-\delta)\;\pi + \delta\;\pi\,,$$

which is true, and to rule out deviation

$$\pi \geq (1-\delta)\,(g+\pi) + \delta\; 0\,, \quad \text{or} \quad 1-\delta \leq \pi/(g+\pi)\,, \quad \text{or} \quad \delta \geq g/(g+\pi)\,.$$

To achieve (0,0), the continuation payoff should be (0,0) for any action profile and this will work for any δ.

Now add imperfect monitoring. Suppose that actions cannot be observed, but at the end of the stage game the players observe a public signal that can be good (G) or bad (B). In the Green-Porter framework, this may be the price that results from the players' actions and a random shock. The bad signal becomes more likely if there is any cheating:

$$\text{Prob}\{\,B\,\} = \begin{cases} \lambda & \text{if (C,C) is played} \\ \mu & \text{if (C,D) or (D,C) is played} \\ 1 & \text{if (D,D) is played} \end{cases}$$

where $0 < \lambda < \mu < 1$, so the signal is not perfectly informative of the actions.

The stage game payoffs to each player depend on his own action and the public signals. Denote the payoff from action a when signal y is observed by $f(y,a)$. When Firm 1 is playing C, the signal may be G or B, and the probability of B is λ if the other firm is also playing C; therefore Firm 1's expected payoff from the action pair (C,C) is $(1-\lambda)\; f(C,G) + \lambda\; f(C,B)$. The other three expected payoffs can be similarly expressed.

To facilitate comparisons between the cases of perfect and imperfect monitoring, it is convenient to rig the stage game payoffs of the latter in such a way that when expectations are taken, the expected payoffs from action combinations will reproduce the stage game payoff matrix of Table 6. In other words, we want

$$\begin{aligned} (1-\lambda)\; f(C,G) + \lambda\; f(C,B) &= \pi \\ (1-\mu)\; f(C,G) + \mu\; f(C,B) &= z \\ (1-\mu)\; f(D,G) + \mu\; f(D,B) &= \pi + g \\ f(D,B) &= 0 \end{aligned}$$

Solving these equations, we get the appropriate definition of the function f as

$$f(y,a) = \begin{cases} (\mu\,\pi - \lambda\,z)/(\mu-\lambda) & \text{if } a = C \text{ and } y = G \\ ((1-\lambda)\,z - (1-\mu)\,\pi)/(\mu-\lambda) & \text{if } a = C \text{ and } y = B \\ (g+\pi)/(1-\mu) & \text{if } a = D \text{ and } y = G \\ 0 & \text{if } a = D \text{ and } y = B \end{cases}$$

To achieve a (flow-equivalent) payoff pair (v, v) in the repeated game, punishment where if a bad signal is observed, a public randomization device generating an outcome P with probability α is played, and if P happens, the play switches to (D,D).[4] Since the play (C,C) may generate a bad signal with probability λ, we have

$$v = (1 - \delta)\,\pi + \delta\,[\,(1 - \alpha\lambda)\,v + \alpha\lambda\,0] = (1 - \delta)\,\pi + \delta\,(1 - \alpha\lambda)\,v$$

or

$$v = \frac{1 - \delta}{1 - \delta + \delta\,\lambda\,\alpha}\,\pi\,. \tag{7}$$

Since a unilateral deviation by a player will raise the probability of the bad signal to μ, the no-cheating condition is

$$(1 - \delta)\,\pi + \delta\,(1 - \alpha\lambda)\,v \geq (1 - \delta)\,(g + \pi) + \delta\,(1 - \alpha\mu)\,v, \quad \text{or} \quad \delta\,(\mu - \lambda)\,\alpha\,v \geq (1 - \delta)\;g\,,$$

or

$$\alpha\,v \geq \frac{1 - \delta}{\delta\,(\mu - \lambda)}\;g\,. \tag{8}$$

In Figure 2 the feasible values v and the corresponding punishment probabilities α must lie along the thicker, flatter curve representing (7), and on or above the thinner, steeper curve, which is a rectangular hyperbola, representing (8) when it holds as an equality.[5] The best value compatible with this requirement is at the intersection point, provided this yields $\alpha \leq 1$; otherwise the repeated game cannot yield any expected payoffs higher than (0,0).

The intersection is at $v = \overline{v}$ where

$$\overline{v} = \pi - \frac{\lambda}{\mu - \lambda}\;g = \pi - \frac{g}{(\mu/\lambda) - 1}\,, \tag{9}$$

and then the condition for $\alpha \leq 1$ becomes

$$\frac{\delta}{1 - \delta} \geq \frac{g}{(\mu - \lambda)\;\overline{v}} = \frac{g}{(\mu - \lambda)\,\pi - \lambda\,g} \tag{10}$$

Many applied models of cooperation assume grim trigger strategies, namely immediate and certain collapse of cooperation, to sustain good behavior. With imperfect monitoring, a bad signal may be observed even when behavior has been good, it is preferable to use a

[4]In this example (D,D) always generates a bad signal; otherwise we would have to consider (and rule out) the possibility of switching back to (C,C) if a good signal is observed during a punishment phase.

[5]The elasticity of (7) is everywhere < 1 numerically, while that of the hyperbola is equal to 1. Incidentally, this proves that there cannot be multiple intersections.

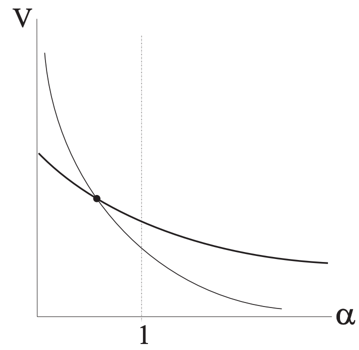

Figure 2: Value of Cooperation and Punishments with Imperfect Monitoring

lower probability of punishment $\alpha < 1$ when this suffices to meet the no-cheating condition. Thus the Green-Porter assumption of sure start of punishment is unduly restrictive (even though they allow the punishment to last only a finite number of periods).

The value expression has an interesting interpretation. Imperfect monitoring reduces the value of the best sustainable cooperation below π; the shortfall depends on (μ/λ), which is a likelihood ratio, being the ratio of probabilities of getting the bad signal with and without cheating. The condition (10) gives a lower bound on δ (or equivalently an upper bound on the benefit of cheating g). The bound is lower (cooperation is easier to sustain) when $(\mu - \lambda)$ is larger, that is, when good and bad behavior have more widely different probabilities of getting the bad signal.

3.7 Multimarket Contact

Often the same firms interact repeatedly in many markets; for example two firms from different countries may compete in both countries' markets. Such repeated games where the stage game consists of two or more games going on simultaneously is of some interest. In the oligopoly context, some early discussion suggested that such multimarket contact could be conducive to greater collusion, because cheating in any one market would be punished by retaliation in all markets. However, this intuition is flawed because the potential cheater can look ahead to this and cheat in all markets simultaneously. If all markets are identical, this merely multiplies the immediate benefits and future costs of cheating by the same factor (the number of markets), leaving their balance, and therefore the sustainability of tacit collusion, unchanged. Any new or added possibility of collusion must be based on some asymmetry, and requires a more careful analysis of how the dynamic incentive constraints in multiple markets can be pooled together. Bernheim and Whinston (1990) have done this.

For definiteness, consider the two-country example. Suppose there are two countries, A and B. Each country's market has the same demand curve $Q = D(p)$. There are two firms, 1 and 2, whose production is located in A and B respectively. Each firm has a constant marginal cost $\underline{c}$ of supplying to its own market, and a higher constant marginal cost $\overline{c}$ of supplying to the other market. The stage game in each market consists of both firms choosing a price. If they choose unequal prices, all demand goes to the firm that names the lower price; if they choose equal prices, the demand can split in any assigned fractions between the two. The duopoly is repeated infinitely, and each firm has the discount factor δ.

For expository convenience, I assume that for any c, the profit function $(p - c)\ Q(p)$ is concave in p and therefore has a unique maximum at $p^m(c)$; this would be the monopoly price when the marginal cost is c. For ease of notation, define

$$\phi(p|c) = \max_{s \leq p}\ (s - c)\, Q(s)\,.$$

This is the best profit one firm can get by undercutting when the other firm's price is p. If $p > p^m(c)$, then $\phi(p|c)$ is maximized at $p^m(c)$; $\phi(p|c) = (p^m - c)\, Q(p^m) > (p - c)\, Q(p)$. If $p \leq p^m(c)$ then $\phi(p|c) = (p - c)\, Q(p)$ as a limiting case of undercutting slightly. So in either case,

$$\phi(p|c) \geq (p - c)\, Q(p)\,. \tag{11}$$

First consider the markets in isolation. In this context, that means any deviation from the tacitly collusive strategy in one market cannot be punished by future actions in the other market. However, in each market there is an optimal punishment strategy yielding zero payoff for both firms: both firms choose price $\underline{c}$, and only the home firm makes all sales.[6] Consider the collusive strategy where in market k both firms set price $p_k > \overline{c}$, and firm i is given a market share λ_{ik}. With zero continuation payoffs, the no-cheating conditions in market A (where firm 1 is the home firm) are

$$\frac{1}{1-\delta}\ \lambda_{1A}\, (p_A - \underline{c})\, Q(p_A) \quad \geq \quad \phi(p_A\,|\,\underline{c})$$

[6] This is a Nash equilibrium; the home firm thinks that if it raises the price slightly above $\underline{c}$, all the demand will go to the foreign firm. It does not take into account the further ramification that the foreign firm, with its delivered marginal cost $\overline{c}$, will refuse to meet this demand, and some customers will then return, yielding positive profit to the home firm. This is outside the rules of Nash equilibrium. But the problem can be resolved by devising other more complicated stick-and-carrot punishment strategies that do not have this defect; see Bernheim and Whinston.

$$\frac{1}{1-\delta}\ \lambda_{2A}\,(p_A - \overline{c})\,Q(p_A) \quad \geq \quad \phi(p_A\,|\,\overline{c})\,.$$

Using (11) on the right hand sides of both lines, canceling common factors, adding the two inequalities, and noting that $\lambda_{1A} + \lambda_{2A} = 1$, we get

$$\frac{1}{1-\delta} \geq 2, \qquad \text{or} \qquad \delta \geq \tfrac{1}{2}\,.$$

This is a necessary condition for self-sustaining collusion; therefore if $\delta < \frac{1}{2}$, no collusion is feasible with isolated markets. When $\delta > \frac{1}{2}$ some collusion is possible, but it is inefficient because the higher-cost firm has to be given a positive share of the market to prevent it from cheating.

Now consider the two markets together. There is still a punishment strategy that gives zero payoff to both firms; it is to set price $\underline{c}$ in each market and let the home firm make all the sales. (The remark in the footnote applies.) Using symmetry, suppose the candidate cooperative strategy is to set price p^* in each market and to let the foreign firm have share λ^*. Consider firm 1; it calculates the benefits and costs of optimally deviating in both markets. Its no-cheating condition is

$$\frac{1}{1-\delta}\ [\ (1-\lambda^*)\ (p^* - \underline{c})\,Q(p^*) + \lambda^*\ (p^* - \overline{c})\,Q(p^*)\ ] \geq \phi(p^*\,|\,\underline{c}) + \phi(p^*\,|\,\overline{c})\,. \tag{12}$$

Since $p^* - \underline{c} > p^* - \overline{c}$, we can most easily meet the no-cheating constraint by choosing $\lambda^* = 0$. That is, the firms should agree to stay away from each other's home market. (This also avoids the inefficiency of allowing the higher-cost firm in each market to supply a positive quantity.) But to prevent them from cheating on such an agreement, it may be necessary to keep the price below the monopoly level. Using $\lambda^* = 0$ and the expression for $\phi(p|c)$ when $p < p^m$, we can rewrite the no-cheating condition (12) as

$$\frac{1}{1-\delta}(p^* - \underline{c})\,Q(p^*) \geq (p^* - \underline{c})\,Q(p^*) + (p^* - \overline{c})\,Q(p^*)\,,$$

or

$$\frac{\delta}{1-\delta}\,(p^* - \underline{c}) \geq (p^* - \overline{c})\,.$$

This can yield some feasible range for p^* even when $\delta < \frac{1}{2}$, specifically,

$$p^* \leq \frac{(1-\delta)\ \overline{c} - \delta\ \underline{c}}{1-2\,\delta}, \quad \text{or} \quad p^* - \overline{c} \leq \frac{\delta}{1-2\,\delta}\ (\,\overline{c} - \underline{c}\,)$$

works. If $\delta > \frac{1}{2}$, full monopoly for each firm in its home market ($p^* = p^m(\underline{c})$) is self-sustaining. So we see how multimarket contact with suitable asymmetry can allow pooling of incentive constraints and expand the possibilities for self-sustaining cooperation or collusion.

3.8 Other Examples and Applications of Interest

Suppose the payoffs of each stage game are subject to a random shock that are independent and identically distributed over time. The shock each period is realized before the actions of that period are taken. Then the temptation to cheat in any stage game will depend on the current realization of the shock, while the expected future payoffs of the continuation game will depend on the whole distribution of the shock. For example, a firm in an oligopoly has a higher temptation to cheat on a tacit collusive agreement in a period when there is a boom in demand, and a country has a higher temptation to cheat on a free trade agreement in a period of high unemployment. The dynamic incentive compatibility condition may fail in such a period.

But if the realization is publicly observable, an optimal cartel arrangement will recognize this temptation, and rather than let it cause a collapse of the whole agreement, allow for flexibility in the specified equilibrium strategies to accommodate the extreme cases of temptation. The constrained optimal subgame perfect collusive arrangement will be better than none at all. Rotemberg and Saloner (1986) and Bagwell and Staiger (1990) have modeled this for the two contexts mentioned above.

If the shocks are player-specific and/or the realization is not publicly observable, matters are more complex. Flexibility cannot be allowed on the say-so of a player about his immediate temptation; that is open to abuse. For example, a self-enforcing mutual insurance arrangement against random consumption needs cannot survive if any member is allowed to claim such need at any time. However, a strategy that offers the immediate help asked, but then reduces the person's future claims on the common pool by a suitably calculated amount, can serve as a screening mechanism that limits claims to situations of truly great need (Atkeson and Lucas, 1992). Related models of repeated oligopoly with individual shocks to cost functions are studied by Athey, Bagwell and Sanchirico (2004).

Readings

Dilip Abreu. 1988. "On the theory of infinitely repeated games with discounting." *Econometrica* 56(2), March, 383–396.

Dilip Abreu, Paul Milgrom, and David Pearce. 1991. "Information and timing in repeated partnerships." *Econometrica* 59(6), November, 1713–1733.

Dilip Abreu, David Pearce, and Ennio Stacchetti. 1990. "Toward a theory of discounted repeated games with imperfect monitoring." *Econometrica* 58(5), September, 1041–1063.

Susan Athey, Kyle Bagwell and Chris Sanchirico. 2004. *Review of Economic Studies* 71(2), April, 317–349.

Andrew Atkeson and Robert E. Lucas, Jr. 1992. "On efficient distribution with private information." *Review of Economic Studies* 59(3), July, 427–453.

Kyle Bagwell and Robert Staiger. 1990. "A theory of managed trade." *American Economic Review* 80(4), September, 779–795.

B. Douglas Bernheim and Michael D. Whinston. 1990. "Multimarket contact and collusive behavior." *RAND Journal of Economics* 21(1), Spring, 1–26.

David Pearce. "Repeated games: Cooperation and rationality." Chapter 4, pp. 132–174, in *Advances in Economic Theory: Sixth World Congress of the Econometric Society,* ed. Jean-Jacques Laffont, Cambridge, UK: Cambridge University Press, 1992.

Julio Rotemberg and Garth Saloner. 1986. "A supergame-theoretic model of price wars during booms." *American Economic Review* 76(3), June, 390–407.

4 Quantal Response Equilibrium

This notion, introduced by McKelvey and Palfrey (1995, 1998) allows for the possibility that players make, and recognize that other players make, some errors in their choices of optimal strategies. It is also open to other interpretations mentioned later. But to explain and motivate the notion in the simplest context, consider the bimatrix game (each player has a finite number of pure strategies) where Aumann's two players, Rowena and Colin, have respective pure strategies indexed by i and j, with resulting payoffs a_{ij} and b_{ij} for the two. Suppose they are choosing mixed strategies, with probabilities p_i for Rowena and q_j for Colin. Table 7 shows all this schematically.

Then Rowena's expected payoffs from her pure strategies i will be

$$u_i = \sum_j \; a_{ij} \; q_j \; ,$$

and Colin's expected payoffs from his pure strategies j will be

$$v_j = \sum_i \; p_i \; b_{ij} \; .$$

Table 7: Schematic Bimatrix Game

		Colin
		q_j
Rowena	p_i	a_{ij} , b_{ij}

In the strictly rational logic of usual game theory, each will choose the strategy that maximizes his or her expected payoff given the strategy (mixture probabilities) of the other; if there are ties at the top, that player will be willing to mix among those pure strategies that achieve that top payoff. The resulting mutually consistent optimal choices (or mixtures) constitute the Nash equilibrium of the game.

But now suppose each player makes small errors. Choices with higher expected payoffs are more likely, but there is some probability that a strategy with a lower expected payoff will be chosen. The theory can be developed in a much more general way, but for expository purposes a useful simple specification is the logit: for a given $\lambda \geq 0$, Rowena chooses her pure strategy i with probability

$$p_i = \exp(\lambda\, u_i) \Big/ \sum_I \exp(\lambda\, u_I) \;,$$

and Colin chooses his pure strategy j with the probability

$$q_j = \exp(\lambda\, v_j) \Big/ \sum_J \exp(\lambda\, v_J) \;.$$

If $\lambda = 0$, each player is making a very naive choice: mixing all available pure strategies with equal probability. In the limit as $\lambda \to \infty$, the choice tends to the fully rational, where only the pure strategies with the highest expected payoffs are chosen. Thus λ is an inverse measure of the departures from rationality, or errors in calculating the optimal choices, that the players make.

The solution of this system of equations for the p_i, q_j, u_i and v_j is called the *quantal response equilibrium (QRE)* of the game.

An alternative interpretation consistent with full rationality is that each player receives some random shocks to his or her payoffs, and these shocks are private observation. For a suitable distribution of these shocks, to an outside observer the outcome of a player's choice appears random with the logistic distribution. In this perspective, this is the model of discrete choice or quantal response, well known in economics from the work of McFadden; see his Nobel Prize lecture (2001). In the game, the distribution of the shocks is common knowledge between the players, and the concept becomes a form of a rational expectations equilibrium of the game.

This equilibrium concept has several useful and attractive properties. [1] Almost all bimatrix games have a QRE. [2] In the logit model, we can select a QRE as a function of λ, which converges to a Nash equilibrium as $\lambda \to \infty$. In this sense, QRE can be regarded as a way of selecting among multiple Nash equilibria, or a *refinement* of Nash equilibrium. It can be related to other refinements such as trembling hand perfectness. [3] Its innate probabilistic structure allows us to formulate its implications in the form of statistical hypotheses to be tested, and to estimate λ. [4] The model fits the data from many experimental and natural games better than fully rational and error-free Nash equilibrium, and much better than a totally naive form of behavior where players pick strategies at random. [5] It appears to involve some ad hockery, in that it introduces an extra parameter λ that can be adjusted to fit data. But the saving grace is that the same range of values of λ seems to fit many different games, suggesting that there is some underlying behavioral feature being captured here. David Reiley's session on experimental games will have more to say on this point.

To illustrate how QRE does a good job of explaining the observed outcomes in some games, consider the centipede game. Figure 2 shows the tree for a 4-move game of this genre; this is Figure 10 from McKelvey and Palfrey (1998). The players move alternately, and at each move can either Take or Pass. Take ends the game and divides the total available at that point in the stated way; Pass causes the total to grow but gives the move to the other player. It is easy to see that the subgame perfect Nash equilibrium strategies are for every player to Take (play T) at every node where he has the move. But in actual play of this game, especially when there are many steps, the players Pass for many early steps until the total grows quite big. The actual outcomes are often Pareto superior to the one predicted by the fully rational subgame perfect Nash equilibrium. This behavior can be explained as a perfect Bayesian Nash equilibrium of a game with a little incomplete information: the possibility that the other player may be irrationally "nice." But QRE provides an attractive alternative explanation.

This is an extensive form game, and needs some reformulation of the QRE concept. The one McKelvey and Palfrey (1998) use is the agent normal form, where the same player when

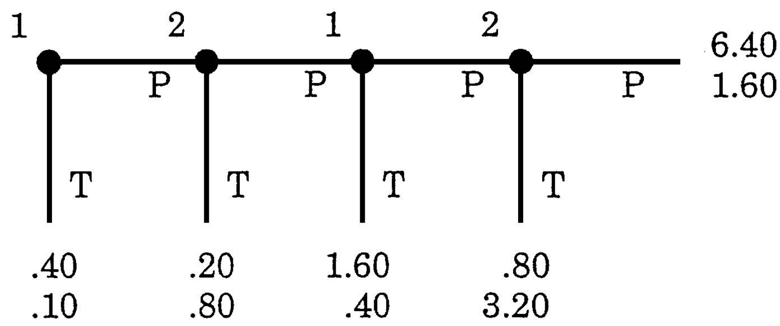

Figure 3: 4-Move Centipede Game

moving at two different nodes (more rigorously, two different information sets is regarded as two separate players. Of course the multiple "agents" or avatars of one physical player all have the same objective function, but they cannot coordinate their actions so each regards the randomness in the other's actions as beyond his control. In the example in the figure, there are two avatars of player 1, moving at the first and the third nodes moving from left to right along the tree, and two avatars of player 2, moving at the second and the fourth nodes. To find the QRE in this agent-normal form (abbreviated AQRE), stipulate probabilities

$$\begin{aligned} p_1 &= \text{Prob}\{\text{ 1 chooses T at the first node }\} \\ p_3 &= \text{Prob}\{\text{ 1 chooses T at the third node (if reached) }\} \\ p_2 &= \text{Prob}\{\text{ 2 chooses T at the second node (if reached) }\} \\ p_4 &= \text{Prob}\{\text{ 2 chooses T at the fourth node (if reached) }\} \end{aligned}$$

The expected payoffs of player 1 at node 1 are

$$\begin{aligned} u_1(T) &= 0.4 \\ u_1(P) &= 0.2\, p_2 + (1-p_2)\ \{1.6\, p_3 + (1-p_3)\,[0.8\, p_4 + 6.4\,(1-p_4)\,]\,\} \end{aligned}$$

Doing similar calculations for the other nodes, we can write down the logistic equations for the probabilities p_i and solve them in terms of λ.

The results of McKelvey and Palfrey's experiments are shown in Table 8 The AQRE "predictions" are based on the point estimate of λ, which was 1.70 with a confidence interval [1.54,1.87]. The agreement is not bad, although the number of subjects who chose to Pass at node 4 is surprisingly high.

Table 8: Results from 4-Move Centipede Experiment

Node	Starters	T-Choosers	Fraction	AQRE Prediction
1	281	20	0.071	0.246
2	261	100	0.383	0.315
3	161	104	0.646	0.683
4	57	42	0.754	0.938

Readings

McFadden, Daniel. 2001. "Economic Choices." *American Economic Review* 91(3), June, 351–378.

McKelvey, Richard D. and Palfrey, Thomas R. 1995. "Quantal response equilibria for normal form games." *Games and Economic Behavior* 10(1): 6–38.

McKelvey, Richard D. and Palfrey, Thomas R. 1998. "Quantal response equilibria for extensive form games." *Experimental Economics* 1(1): 9–41.