First-period profile	Continuation	First-period profile	Continuation
A	B		
BNN, BYY	C	DNN, DYY	D
BYN	E	DYN	E
BNY	B	DNY	B
CNN, CYY	C	ENN, EYY	D
CYN	E	EYN	E
CNY	B	ENY	B

Figure 5.4.9 The continuation play of the equilibrium for the repeated game.

or

$$v_2 \geq \tfrac{11}{2}.$$

Player 1's deviation to B must then be followed by a continuation path giving player 2 a payoff of at least 11/2. A symmetric argument shows that the same must be true for player 3 after 1's deviation to E. However, there is no stage-game action profile that gives players 2 and 3 both at least 11/2. Hence, if player 1 is ever to play A, then deviations to B and E must lead to different continuation paths—the equilibrium strategy cannot be agent simple. Finally, player 1's choice of A gives payoffs of $(1, 5, 5)$, a feat that is impossible if 1 confines himself to choices in $\{B, C, D, E\}$. The payoffs provided by this equilibrium can thus be achieved only via strategies that are not agent simple.

5.5 Dynamic Games: Introduction

This section allows the possibility that the stage game changes from period to period for a fixed set of players, possibly randomly and possibly as a function of the history of play. Such games are referred to as *dynamic games* or, when stressing that the stage game may be a random function of the game's history, *stochastic games*.

The analysis of a dynamic game typically revolves around a set of *game states* that describe how the stage game varies from period to period. Unless we need to distinguish between game states and states of an automaton (*automaton states*), we refer to game states simply as states (see remark 5.5.2). Each state determines a stage game, captured by writing payoffs as a function of states and actions. The specification of the game is completed by a rule for how the state changes over the course of play.

In many applications, the context in which the game arises suggests what appears to be a natural candidate for the set of states. It is accordingly common to treat

the set of states as an exogenously specified feature of the environment. This section proceeds in this way. However, the appropriate formulation of the set of states is not always obvious. Moreover, the notion of a state is an intermediate convention that is not required for the analysis of dynamic games. Instead, we can define payoffs directly as functions of current and past actions, viewing states as tools for describing this function. This suggests that instead of inspecting the environment and asking which of its features appear to define states, we begin with the payoff function and identify the states that implicitly lie behind its structure. Section 5.6 pursues this approach. With these tools in hand, section 5.7 examines equilibria in dynamic games.

5.5.1 The Game

There are n players, numbered $1, \ldots, n$. There is a set of states S, with typical state s. Player i has the compact set of actions $A_i \subset \mathbb{R}^k$, for some k. Player i's payoffs are given by the continuous function $u_i : S \times A \to \mathbb{R}$. Because payoffs are state dependent, the assumption that A_i is state independent is without loss of generality: If state s has action set A_i^s, define $A_i \equiv \prod_s A_i^s$ and set $\tilde{u}_i(s, a) = u_i(s, a^s)$. Players discount at the common rate δ.

The evolution of the state is given by a continuous transition function $q : S \times A \cup \{\varnothing\} \to \Delta(S)$, associating with each current state and action profile a probability distribution from which the next state is drawn; $q(\varnothing)$ is the distribution over initial states. This formulation captures a number of possibilities. If S is a singleton, then we are back to the case of a repeated game. If $q(s, a)$ is nondegenerate but constant in a, then we have a game in which payoffs are random variables whose distribution is constant across periods. If $S = \{0, 1, 2, \ldots\}$ and $q(s_\ell, a)$ puts probability one on $s_{\ell+1}$, we have a game in which the payoff function varies deterministically across periods, independently of behavior.

We focus on two common cases. In one, the set of states S is finite. We then let $q(s' \mid s, a)$ denote the probability that the state s' is realized, given that the previous state was s and the players chose action profile a (the initial state can be random in this case). We allow equilibria to be either pure or, if the action sets are finite, mixed. In the other case, A_i and S are infinite, in which case $S \subset \mathbb{R}^m$ for some m. We then take the transition function to be deterministic, so that for every s and a, there exists s' with $q(s' \mid s, a) = 1$ and $q(s'' \mid s, a) = 0$ for all $s'' \neq s'$ (the initial state is deterministic and given by s^0 in this case). As is common, we then restrict attention to pure-strategy equilibria.

In each period of the game, the state is first drawn and revealed to the players, who then simultaneously choose their actions. The set of period t ex ante histories $\mathscr{H}^t$ is the set $(S \times A)^t$, identifying the state and the action profile in each period.[17] The set of period t ex post histories is the set $\tilde{\mathscr{H}}^t = (S \times A)^t \times S$, giving state and

17. Under the state transition rule q, many of the histories in this set may be impossible. If so, the specification of behavior at these histories will have no effect on payoffs.

action realizations for each previous period and identifying the current state. Let $\tilde{\mathscr{H}} = \cup_{t=0}^{\infty} \tilde{\mathscr{H}}^t$; we set $\mathscr{H}^0 = \{\varnothing\}$, so that $\tilde{\mathscr{H}}^0 = S$. Let $\mathscr{H}^\infty$ be the set of outcomes.

A pure strategy for player i is a mapping $\sigma_i : \tilde{\mathscr{H}} \to A_i$, associating an action with each ex post history. A pure strategy profile σ, together with the transition function q, induces a probability measure over $\mathscr{H}^\infty$. Player i's expected payoff is then

$$U_i(\sigma) = E^\sigma \left\{ (1-\delta) \sum_{t=0}^{\infty} \delta^t u_i(s^t, a^t) \right\},$$

where the expectation is taken with respect to the measure over $\mathscr{H}^\infty$ induced by $q(\varnothing)$ and σ. Note that the expectation may be nontrivial, even for a pure strategy profile, because the transition function q may be random. As usual, this formulation has a straightforward extension to behavior strategies. For histories other than the null history, we let $U_i(\sigma \mid \tilde{h}^t)$ denote i's expected payoffs induced by the strategy profile σ in the continuation game that follows the ex post history $\tilde{h}^t$.

An ex ante history ending in (s, a) gives rise to a continuation game matching the original game but with $q(\varnothing)$ replaced by $q(\cdot \mid s, a)$ and the transition function otherwise unchanged. An ex post history $\tilde{h}^t$, ending with state s, gives rise to the continuation game again matching the original game but with the initial distribution over states now attaching probability 1 to state s. It will be convenient to denote this game by $G(s)$.

Definition 5.5.1 *A strategy profile σ is a* Nash equilibrium *if $U_i(\sigma) \geq U_i(\sigma_i', \sigma_{-i})$ for all σ_i' and for all players i. A strategy profile σ is a* subgame-perfect equilibrium *if, for any ex post history $\tilde{h}^t \in \tilde{\mathscr{H}}$ ending in state s, the continuation strategy $\sigma|_{\tilde{h}^t}$ is a Nash equilibrium of the continuation game $G(s)$.*

Example 5.5.1 Suppose there are two equally likely states, independently drawn in each period, with payoffs given in figure 5.5.1. It is a quick calculation that strategies specifying effort after every ex post history in which there has been no previous shirking (and specifying shirking otherwise) are a subgame-perfect equilibrium if and only if $\delta \geq 2/7$. Strategies specifying effort after any ex post history ending in state 2 while calling for shirking in state 1 (as long as there have been no deviations from this prescription, and shirking otherwise) are an equilibrium if and only if $\delta \geq 2/5$. The deviation incentives are the same in both games, whereas exerting

State 1:

	E	S
E	$2, 2$	$-1, 3$
S	$3, -1$	$0, 0$

State 2:

	E	S
E	$3, 3$	$-1, 4$
S	$4, -1$	$0, 0$

Figure 5.5.1 Payoff functions for states 1 and 2 of a dynamic game.

effort in only one state reduces the equilibrium continuation value, and hence requires more patience to sustain effort in the other state.[18]

●

Remark 5.5.1 **Repeated games with random states** The previous example illustrates the special case in which the probability of the current state is independent of the previous state and the players' actions, that is, $q(s \mid s', a') = q(s \mid s'', a'') \equiv q(s)$. We refer to such dynamic games as *repeated games with random states*. These games are a particularly simple repeated game with imperfect public monitoring (section 7.1.1). Player i's pure action set in the stage game is given by the set of functions from S into A_i, and players simultaneously choose such actions. The pure-action profile σ then gives rise to the signal $(s, \sigma_1(s), \ldots, \sigma_n(s))$ with probability $q(s)$, for each $s \in S$. Player i's payoff $u_i(s, a)$ can then be written as $u_i((s, \sigma_{-i}(s)), a_i(s))$, giving i's payoff as a function of i's action and the public signal. We describe a simpler approach in remark 5.7.1.

◆

5.5.2 Markov Equilibrium

In principle, strategies in a dynamic game could specify each period's action as a complicated function of the preceding history. It is common, though by no means universal, to restrict attention to Markov strategies:

Definition 5.5.2

1. *The strategy profile σ is a* Markov strategy *if for any two ex post histories $\tilde{h}^t$ and $\tilde{h}^\tau$ of the same length and terminating in the same state, $\sigma(\tilde{h}^t) = \sigma(\tilde{h}^\tau)$. The strategy profile σ is a* Markov equilibrium *if σ is a Markov strategy profile and a subgame-perfect equilibrium.*
2. *The strategy profile σ is a* stationary Markov strategy *if for any two ex post histories $\tilde{h}^t$ and $\tilde{h}^\tau$ (of equal or different lengths) terminating in the same state, $\sigma(\tilde{h}^t) = \sigma(\tilde{h}^\tau)$. The strategy profile σ is a* stationary Markov equilibrium *if σ is a stationary Markov strategy profile and a subgame-perfect equilibrium.*

It is sometimes useful to reinforce the requirement of subgame perfection by referring to a Markov equilibrium as a *Markov perfect equilibrium*. Some researchers also refer to game states as *Markov states* when using Markov equilibrium (but see remark 5.5.2).

Markov strategies ignore all of the details of a history except its length and the current state. Stationary Markov strategies ignore all details except the current state.

Three advantages for such equilibria are variously cited. First, Markov equilibria appear to be simple, in the sense that behavior depends on a relatively small set of variables, often being the simplest strategies consistent with rationality. To some, this simplicity is appealing for its own sake, whereas for others it is an analytical or computational advantage. Markov equilibria are especially common in applied work.[19]

18. In contrast to section 5.3, we face here one randomly drawn game in each period, instead of both games.

19. It is not true, however, that Markov equilibria are always simpler than non-Markov equilibria. The proof of proposition 18.4.4 goes to great lengths to construct a Markov equilibrium featuring high effort, in a version of the product choice game, that would be a straightforward calculation in non-Markov strategies.

Second, the set of Markov equilibrium outcomes is often considerably smaller than the set of all equilibrium outcomes. This is a virtue for some and a vice for others, but again contributes to the popularity of Markov equilibria in applied work.

Third, Markov equilibria are often viewed as having some intuitive appeal for their own sake. The source of this appeal is the idea that only things that are "payoff relevant" should matter in determining behavior. Because the only aspect of a history that affects current payoff functions is the current state, then a first step in imposing payoff relevance is to assume that current behavior should depend only on the current state.[20] Notice, however, that there is no reason to limit this logic to dynamic games. It could just as well be applied in a repeated game, where it is unreasonably restrictive, because *nothing* is payoff relevant in the sense typically used when discussing dynamic games. Insisting on Markov equilibria in the repeated prisoners' dilemma, for example, dooms the players to perpetual shirking. More generally, a Markov equilibrium in a repeated game must play a stage-game Nash equilibrium in every period, and a stationary Markov equilibrium must play the same one in every period.

Remark 5.5.2 **Three types of state** We now have three notions of a state to juggle. One is *Markov state*, an equivalence class of histories in which distinctions are payoff irrelevant. The second is *game state*, an element of the set S of states determining the stage game in a dynamic game. Though it is often taken for granted that the set of Markov states can be identified with the set of game states, as we will see in section 5.6, these are distinct concepts. Game states may not always be payoff relevant and, more important, we can identify Markov states without any a priori specification of a game state. Finally, we have *automaton states*, states in an automaton representing a strategy profile. Because continuation payoffs in a repeated game depend on the current automaton state, and only on this state, some researchers take the set of automata states as the set of Markov states. This practice unfortunately robs the Markov notion and payoff relevance of any independent meaning. The particular notion of payoff relevance inherent in labeling automaton states Markov is much less restrictive than that often intended to be captured by Markov perfection. For example, Markov perfection then imposes *no* restrictions beyond subgame perfection in repeated games, because any subgame-perfect equilibrium profile has an automaton representation, in contrast to the trivial equilibria that appear if we at least equate Markov states with game states.

◆

5.5.3 Examples

Example 5.5.2 Suppose that players 1 and 2 draw fish from a common pool. In each period t, the pool contains a stock of fish of size $s^t \in \mathbb{R}_+$. In period t, player i extracts $a_i^t \geq 0$ units of fish, and derives payoff $\ln(a_i^t)$ from extracting a_i^t.[21] The remaining (depleted) stock of fish doubles before the next period. This gives a dynamic game with actions and states drawn from the infinite set $[0, \infty)$, identifying the current quantity of fish extracted (actions) and the current stock of fish (states), and with the deterministic transition function $s^{t+1} = 2(s^t - a_1^t - a_2^t)$. The initial stock is fixed at some value s^0.

We first calculate a stationary Markov equilibrium of this game, in which the players choose identical strategies. That is, although we assume that the players choose Markov strategies in equilibrium, the result is a strategy profile that is optimal in the full strategy set—there are no superior strategies, Markov or otherwise. We are thus calculating not an equilibrium in the game in which players are restricted to Markov strategies but Markov strategies that are an equilibrium of the full game.

The restriction to Markov strategies allows us to introduce a function $V(s)$ identifying the equilibrium value (conditional on the equilibrium strategy profile) in any continuation game induced by an ex post history ending in state s. Let $g^t(s^0)$ be the amount of fish extracted by each player at time t, given the period 0 stock s^0 and the (suppressed, in the notation) equilibrium strategies. The function $V(s^0)$ identifies equilibrium utilities, and hence must satisfy

$$V(s^0) = (1-\delta)\sum_{t=0}^{\infty}\delta^t \ln(g^t(s^0)). \tag{5.5.1}$$

Imposing the Markov restriction that current actions depend only on the current state, let each player's strategy be given by a function $a(s)$ identifying the amount of fish to extract given that the current stock is s. We then solve jointly for the function $V(s)$ and the equilibrium strategy $a(s)$. First, the one-shot deviation principle (which we describe in section 5.7.1) allows us to characterize the function $a(s)$ as solving, for any $s \in S$ and for each player i, the Bellman equation,

$$a(s) \in \arg\max_{\tilde{a} \in A_i} (1-\delta)\ln(\tilde{a}) + \delta V(2(s - \tilde{a} - a(s))),$$

where $\tilde{a}$ is player i's consumption and the $a(s)$ in the final term captures the assumption that player j adheres to the candidate equilibrium strategy. If the value function V is differentiable, the implied first-order condition is

$$\frac{(1-\delta)}{a(s)} = 2\delta V'(2(s - 2a(s))).$$

To find an equilibrium, suppose that $a(s)$ is given by a linear function, so that $a(s) = ks$. Then we have $s^{t+1} = 2(s^t - 2ks^t) = 2(1-2k)s^t$. Using this and $a(s) = ks$ to recursively replace $g^t(s)$ in (5.5.1), we have

$$V(s) = (1-\delta)\sum_{t=0}^{\infty}\delta^t \ln[k(2(1-2k))^t s].$$

20. If the function $u(s, a)$ is constant over some values of s, then we could impose yet further restrictions.

21. We must have $a_1^t + a_2^t \leq s^t$. We can model this by allowing the players to choose extraction levels $\tilde{a}_1^t$ and $\tilde{a}_2^t$, with these levels realized if feasible and with a rationing rule otherwise determining realized extraction levels. This constraint will not play a role in the equilibrium, and so we leave the rationing rule unspecified and treat a_i^t and $\tilde{a}_i^t$ as identical.

and so V is differentiable with $V'(s) = 1/s$. Solving the first-order condition, $k = (1 - \delta)/(2 - \delta)$, and so

$$a(s) = \frac{1-\delta}{2-\delta}s$$

and

$$V(s) = (1-\delta)\sum_{t=0}^{\infty}\delta^t \ln\left(\frac{1-\delta}{2-\delta}s\left(\frac{2\delta}{2-\delta}\right)^t\right).$$

We interpret this expression by noting that in each period, proportion

$$1 - 2a(s) = 1 - 2\frac{1-\delta}{2-\delta} = \frac{\delta}{2-\delta}$$

of the stock is preserved until the next period, where it is doubled, so that the stock grows at rate $2\delta/(2-\delta)$. In each period, each player consumes fraction $(1-\delta)/(2-\delta)$ of this stock.

Notice that in this solution, the stock of resource grows without bound if the players are sufficiently patient ($\delta > 2/3$), though payoffs remain bounded, and declines to extinction if $\delta < 2/3$. As is expected from these types of common pool resource problems, this equilibrium is inefficient. Failing to take into account the externality that their extraction imposes on their partner's future consumption, each player extracts too much (from an efficiency point of view) in each period.

This stationary Markov equilibrium is not the only equilibrium of this game. To construct another equilibrium, we first calculate the largest symmetric payoff profile that can be achieved when the firms choose identical Markov strategies. Again representing the solution as a linear function $a = ks$, we can write the appropriate Bellman equation as

$$a(s) = \operatorname*{argmax}_{\tilde{a}\in A_i} 2(1-\delta)\ln(\tilde{a}) + \delta(1-\delta)\sum_{t=0}^{\infty}\delta^t 2\ln(k(2(1-2k))^t(s-2\tilde{a})).$$

Taking a derivative with respect to $\tilde{a}$ and simplifying, we find that the efficient solution is given by

$$a(s) = \frac{1-\delta}{2}s.$$

As expected, the efficient solution extracts less than does the Markov equilibrium. The efficient solution internalizes the externality that player i's extraction imposes on player j, through its effect on future stocks of fish.

Under the efficient solution, we have

$$s^{t+1} = 2s^t\left(1 - 2\frac{1-\delta}{2}\right) = 2\delta s^t,$$

and hence

$$s^t = (2\delta)^t s^0.$$

The stock of the resource grows without bound if $\delta > 1/2$. Notice also that just as we earlier solved for Markov strategies that are an equilibrium in the complete strategy set, we have now found Markov strategies that maximize total expected payoffs over the set of all strategies.

We can support the efficient solution as a (non-Markov) equilibrium of the repeated game, if the players are sufficiently patient. Let strategies prescribe the efficient extraction after every history in which the quantity extracted has been efficient in each previous period, and prescribe the Markov equilibrium extraction $a(s) = [(1-\delta)/(2-\delta)]s$ otherwise. Then for sufficiently large δ, we have an equilibrium.

●

Example 5.5.3 Consider a market with a single good, produced by a monopoly firm facing a continuum of small, anonymous consumers.[22] We think of the firm as a long-lived player and interpret the consumers as short-lived players.

The good produced by the firm is durable. The good lasts forever, subject to continuous depreciation at rate η, so that 1 unit of the good purchased at time 0 depreciates to $e^{-\eta t}$ units of the good at time t. This durability makes this a dynamic rather than repeated game.

Time is divided into discrete periods of length Δ, with the firm making a new production choice at the beginning of each period. We will subsequently be interested in the limiting case as Δ becomes very short. The players in the model discount at the continuously compounded rate r. For a period of length Δ, the discount factor is thus $e^{-r\Delta}$.

The stock of the good in period t is denoted $x(t)$. The stock includes the quantity that the firm has newly produced in period t, as well as the depreciated remnants of past production. Though the firm's period t action is the period t quantity of production, it is more convenient to treat the firm as choosing the stock $x(t)$. The firm thus chooses a sequence of stocks $\{x(0), x(1), \ldots\}$, subject to $x(t) \geq e^{-\eta\Delta}x(t-1)$.[23] Producing a unit of the good incurs a constant marginal cost of c.

Consumers take the price path as given, believing that their own consumption decisions cannot influence future prices. Rather than modeling consumers' maximization behavior directly, we represent it with the inverse demand curve $f(x) = 1 - x$. We interpret $f(x)$ as the *instantaneous* valuation consumers attach to x units of the durable good. We must now translate this into our setting with periods of length Δ. The value *per unit* a consumer assigns to acquiring a quantity x at the beginning of a period and used *only* throughout that period, with no previous or further purchases, is

$$\begin{aligned} F(x) &= \int_0^{\Delta} f(xe^{-\eta s})e^{-(r+\eta)s}ds \\ &= \frac{1}{r+\eta}(1-e^{-(r+\eta)\Delta}) - x\frac{1}{r+2\eta}(1-e^{-(r+2\eta)\Delta}) \\ &\equiv \theta - \beta x. \end{aligned}$$

22. For a discussion of durable goods monopoly problems, see Ausubel, Cramton, and Deneckere (2002). The example in this section is taken from Bond and Samuelson (1984, 1987). The introduction of depreciation simplifies the example, but is not essential to the results (Ausubel and Deneckere 1989).

23. Because this constraint does not bind in the equilibrium we construct, we can ignore it.

Because the good does not disappear at the end of the period, the period t price (reflecting current and future values) given the sequence $\mathbf{x}(t) \equiv \{x(t), x(t+1), \ldots\}$ of period t and future stocks of the good, is given by

$$p(t, \mathbf{x}(t)) = \sum_{s=0}^{\infty} e^{-(r+\eta)\Delta s} F(x(t+s)).$$

The firm's expected payoff in period t, given the sequence of actions $x(t-1)$ and $\mathbf{x}(t)$, is given by[24]

$$\sum_{\tau=t}^{\infty} (x(\tau) - x(\tau-1)e^{-\eta\Delta})(p(\tau, \mathbf{x}(\tau)) - c)e^{-r\Delta(\tau-t)}.$$

If the good were perishable, this would be a relatively straightforward intertemporal price discrimination problem. The durability of the good complicates the relationship between current prices and future actions. We begin by seeking a stationary Markov equilibrium. The state variable in period t is the stock $x(t-1)$ chosen in the previous period. The firm's strategy is described by a function $x(t) = g(x(t-1))$, giving the period t stock as a function of the previous period's stock. However, a more flexible description of the firm's strategy is more helpful. We consider a function

$$g(s, t, x),$$

identifying the period s stock, given that the stock in period $t \le s$ is x. Hence, we build into our description of the Markov strategy the observation that if period t's stock is a function of period $t-1$'s, then so is period $t+2$'s stock a (different) function of $x(t-1)$, and so is $x(t+3)$, and so on. To ensure this representation of the firm's strategy is coherent, we impose the consistency condition that $g(s', t, x) = g(s', s, g(s, t, x))$ for $s' \ge s \ge t$. In addition, $g(s+\tau, s, x)$ must equal $g(t+\tau, t, x)$ for $s \ne t$, so that the same state variable produces identical continuation behavior, regardless of how it is reached and regardless of when it is reached.

The firm's profit maximization problem, in any period t, is to choose the sequence of stocks $\{x(t), x(t+1), \ldots, \}$ to maximize

$$\begin{aligned} V(\mathbf{x}(t), t \mid x(t-1)) &= \sum_{\tau=t}^{\infty} [x(\tau) - x(\tau-1)e^{-\eta\Delta}](p(\tau, \mathbf{x}(\tau)) - c)e^{-r\Delta(\tau-t)} \qquad (5.5.2)\\ &= \sum_{\tau=t}^{\infty} [x(\tau) - x(\tau-1)e^{-\eta\Delta}] \\ &\quad \times \left(\sum_{s=0}^{\infty} e^{-(r+\eta)\Delta s}(\theta - \beta g(s+\tau, \tau, x(\tau))) - c\right) e^{-r\Delta(\tau-t)}. \qquad (5.5.3)\end{aligned}$$

In making the substitution for $p(\tau, x(\tau))$ that brings us from (5.5.2) to (5.5.3), $g(s+\tau, \tau, x(\tau))$ describes consumers' expectations of the firm's future stocks and

24. Notice that we must specify the stock in period $t-1$ because this combines with $x(t)$ to determine the quantity produced and sold in period t.

so their own future valuations. To find an optimal strategy for the firm, we differentiate $V(\mathbf{x}(t), t \mid x(t-1))$ with respect to $x(t')$ for $t' \ge t$ to obtain a first-order condition for the latter. In doing so, we hold the values $x(\tau)$ for $\tau \ne t'$ fixed, so that the firm chooses $x(\tau)$ and $x(t')$ independently. However, consumer expectations are given by the function $g(s+\tau, \tau, x(\tau))$, which builds in a relationship between the current stock and anticipated future stocks that determines current prices.

Fix $t' \ge t \ge 0$. The first-order condition $dV(\mathbf{x}(t), t \mid x(t-1))/dx(t') = 0$ is, from (5.5.3),

$$\begin{aligned}&[\theta - \beta x(t') - c(1 - e^{-(r+\eta)\Delta})] \\ &\quad - \beta[x(t') - e^{-\eta\Delta}x(t'-1)] \sum_{s=0}^{\infty} e^{-(r+\eta)\Delta s} \left.\frac{dg(t'+s, t', x)}{dx}\right|_{x=x(t')} = 0.\end{aligned}$$

Using $x(t') = g(t', t, x(t))$, we rewrite this as

$$\begin{aligned}&[\theta - \beta g(t', t, x(t)) - c(1 - e^{-(r+\eta)\Delta})] \\ &\quad - \beta[g(t', t, x(t)) - e^{-\eta\Delta} g(t'-1, t, x(t))] \\ &\quad \times \left[\sum_{s=0}^{\infty} e^{-(r+\eta)\Delta s} \frac{dg(t'+s, t', g(t', t, x(t)))}{dx}\right] = 0.\end{aligned}$$

As is typically the case, this difference equation is solved with the help of some informed guesswork. We posit $g(s, t, x)$ takes the form

$$g(s, t, x) = \bar{x} + \mu^{s-t}(x - \bar{x}),$$

where we interpret $\bar{x}$ as a limiting stock of the good and μ as identifying the rate at which the stock adjusts to this limit. With this form for g, it is immediate that the first-order conditions characterize the optimal value of $x(t')$.

Substituting this expression into our first-order condition gives,

$$\begin{aligned}&\left\{\theta - c(1 - e^{-(r+\eta)\Delta}) - \beta\bar{x}\left(1 + \frac{1 - e^{-\eta\Delta}}{1 - e^{-(r+\eta)\Delta}\mu}\right)\right\} \\ &\quad - \beta\mu^{t'-t-1}(x - \bar{x})\left\{\mu + \frac{\mu - e^{-\eta\Delta}}{1 - e^{-(r+\eta)\Delta}\mu}\right\} = 0.\end{aligned}$$

Because this equation must hold for all t', we conclude that each expression in braces must be 0. We can solve the second for μ and then insert in the first to solve for $\bar{x}$, yielding

$$\mu = \frac{1 - \sqrt{1 - e^{-(r+2\eta)\Delta}}}{e^{-(r+\eta)\Delta}}$$

and

$$\bar{x} = \frac{[\theta - c(1 - e^{-(r+\eta)\Delta})]}{\beta} \frac{\sqrt{1 - e^{-(r+2\eta)\Delta}}}{(\sqrt{1 - e^{-(r+2\eta)\Delta}} + 1 - e^{-\eta\Delta})}.$$

Notice first that $\mu < 1$. The stock of good produced by the monopoly thus converges monotonically to the limiting stock $\bar{x}$. In the expression for the limit $\bar{x}$,

$$\frac{[\theta - c(1 - e^{-(r+\eta)\Delta})]}{\beta}$$

is the competitive stock. Maintaining the stock at this level in every period gives $p(t, \mathbf{x}) = c$, and hence equality of price and marginal cost. The term

$$\frac{\sqrt{1 - e^{-(r+2\eta)\Delta}}}{(\sqrt{1 - e^{-(r+2\eta)\Delta}} + 1 - e^{-\eta\Delta})} \equiv \gamma(\Delta, \eta) \tag{5.5.4}$$

then gives the ratio between the monopoly's limiting stock of good and the competitive stock. The following properties follow immediately from (5.5.4):

$$\gamma(\Delta, \eta) < 1, \tag{5.5.5}$$

$$\lim_{\eta \to \infty} \gamma(\Delta, \eta) = \tfrac{1}{2}, \tag{5.5.6}$$

$$\gamma(\Delta, 0) = 1, \tag{5.5.7}$$

and

$$\lim_{\Delta \to 0} \gamma(\Delta, \eta) = 1. \tag{5.5.8}$$

Condition (5.5.5) indicates that the monopoly's limiting stock is less than that of a competitive market. Condition (5.5.6) shows that as the depreciation rate becomes arbitrarily large, the limiting monopoly quantity is half that of the competitive market, recovering the familiar result for perishable goods.

Condition (5.5.7) indicates that if the good is perfectly durable, then the limiting monopoly quantity approaches that of the competitive market. As the competitive stock is approached, the price-cost margin collapses to 0. With a positive depreciation rate, it pays to keep this margin permanently away from 0, so that positive profits can be earned on selling replacement goods to compensate for the continual deprecation. As the rate of depreciation goes to 0, however, this source of profits evaporates and profits are made only on new sales. It is then optimal to extract these profits, to the extent that the price-cost margin is pushed to 0.

Condition (5.5.8) shows that as the period length goes to 0, the stock again approaches the competitive stock. If the stock stops short of the competitive stock, every period brings the monopoly a choice between simply satisfying the replacement demand, for a profit that is proportional to the length of the period, or pushing the price lower to sell new units to additional consumers. The latter profit is proportional to the price and hence must overwhelm the replacement demand for short time periods, leading to the competitive quantity as the length of a period becomes arbitrarily short.

More important, because the adjustment factor μ is also a function of Δ, by applying l'Hôpital's rule to $\Delta^{-1} \ln \mu$, one can show that

$$\lim_{\Delta \to 0} \mu^{\frac{1}{\Delta}} = 0.$$

Hence, as the period length shrinks to 0, the monopoly's output path comes arbitrarily close to an instantaneous jump to the competitive quantity.[25] Consumers

25. At calender time T, T/Δ periods have elapsed, and so, as $\Delta \to 0$, $x(T/\Delta) \to \bar{x}$.

build this behavior into their pricing behavior, ensuring that prices collapse to marginal cost and the firm's profits collapse to 0. This is the Coase conjecture in action.[26]

How should we think about the Markov restriction that lies behind this equilibrium? The key question facing a consumer, when evaluating a price, concerns how rapidly the firm is likely to expand the stock and depress the price in the future. The firm will firmly insist that there is nary a price reduction in sight, a claim that the consumer would do well to treat with some skepticism. One obvious place for the consumer to look in assessing this claim is the firm's past behavior. Has the price been sitting at nearly its current level for a long time? Or has the firm been racing down the demand curve, having charged ten times as much only periods ago? Markov strategies insist that consumers ignore such information. If consumers find the information relevant, then we have moved beyond Markov equilibria.

To construct an alternative equilibrium with quite different properties, let

$$x_R = \frac{[\theta - c(1 - e^{-(r+\eta)\Delta})]}{2\beta}.$$

This is half the quantity produced in a competitive market. Choosing this stock in every period maximizes the firm's profits over the set of Nash equilibria of the repeated game.[27] Now let $\mathscr{H}^*$ be the set of histories in which the stock x_R has been produced in every previous period. Notice that this includes the null history. Then consider the firm's strategy $x(h^t)$, giving the current stock as a function of the history h^t, given by

$$x(h^t) = \begin{cases} x_R, & \text{if } h_t \in \mathscr{H}^*, \\ g(t, t-1, x(t-1)), & \text{otherwise}, \end{cases}$$

where $g(\cdot)$ is the Markov equilibrium strategy calculated earlier. In effect, the firm "commits" to produce the profit-maximizing quantity x_R, with any misstep prompting a switch to continuing with the Markov equilibrium. Let σ^R denote this strategy and the attendant best response for consumers.

It is now straightforward that σ^R is a subgame-perfect equilibrium, as long as the length of a period Δ is sufficiently short. To see this, let $U(\sigma^R \mid x)$ be the monopoly's continuation payoff from this strategy, given that the current stock is $x \leq x_R$. We are interested in the continuation payoffs given stock x_R, or

$$U(\sigma^R \mid x_R) = \sum_{\tau=0}^{\infty} e^{-r\Delta\tau}(1 - e^{-\eta\Delta})x_R\left(-c + \sum_{s=0}^{\infty} e^{-(r+\eta)\Delta s} F(x_R)\right)$$

$$= \frac{(1 - e^{-\eta\Delta})[F(x_R) - c(1 - e^{-(r+\eta)\Delta})]}{(1 - e^{-r\Delta})(1 - e^{-(r+\eta)\Delta})} x_R.$$

26. Notice that in examining the limiting case of short time periods, depreciation has added nothing other than extra terms to the model.

27. This quantity maximizes $[\theta - \beta x - c(1 - e^{-(r+\eta)\Delta})]x$, and hence is the quantity that would be produced in each period by a firm that retained ownership of the good and rented its services to the customers. A firm who sells the good can earn no higher profits.

The key observation now is that

$$\lim_{\Delta\to\infty} U(\sigma^R \mid x_R) > 0.$$

Even as time periods become arbitrarily short, there are positive payoffs to be made by continually replacing the depreciated portion of the profit-maximizing quantity x_R. In contrast, as we have seen, as time periods shorten, the continuation payoff of the Markov equilibrium, from any initial stock, approaches 0. This immediately yields:

Proposition 5.5.1 *There exists Δ^* such that, if $\Delta < \Delta^*$, then strategies σ^R are a subgame-perfect equilibrium.*

This example illustrates that a Markov restriction can make a great difference in equilibrium outcomes. One may or may not be convinced that a focus on Markov equilibria is appropriate, but one cannot rationalize the restriction simply as an analytical convenience.

●

5.6 Dynamic Games: Foundations

What determines the set S of states for a dynamic game? At first the answer seems obvious—states are things that affect payoffs, such as the stock of fish in example 5.5.2 or the stock of durable good in example 5.5.3. However, matters are not always so clear.

For example, our formulation of repeated games in chapter 2 allows players to condition their actions on a public random variable. Are these realizations states, in the sense of a dynamic game, and does Markov equilibrium allow behavior to be conditioned on such realizations? Alternatively, consider the infinitely repeated prisoners' dilemma. It initially appears as if there are no payoff-relevant states, so that Markov equilibria must feature identical behavior after every history and hence must feature perpetual shirking. Suppose, however, that we defined two states, an effort state and a shirk state. Let the game begin in the effort state, and remain there as along as there has been no shirking, being otherwise in the shirk state. Now let players' strategies prescribe effort in the effort state and shirking in the shirk state. We now have a Markov equilibrium (for sufficiently patient players) featuring effort. Are these states real, or are they a sleight of hand?

In general, we can define payoffs for a dynamic game as functions of current and past actions, without resorting to the idea of a state. As a result, it can be misleading to think of the set of states as being exogenously given. Instead, if we would like to work with the notions of payoff relevance and Markov equilibrium, we must endogenously infer the appropriate set of states from the structure of payoffs.

This section pursues this notion of a state, following Maskin and Tirole (2001). We examine a game with players $1, \ldots, n$. Each player i has the set A_i of stage-game actions available in each period. Hence, the set of feasible actions is again independent

of history.[28] Player i's payoff is a function of the outcome path $\mathbf{a} \in A^\infty$. This formulation is sufficiently general to cover the dynamic games of section 5.6. Deterministic transitions are immediately covered because the history of actions h^t determines the state s reached in period t. For stochastic transitions, such as example 5.5.1, introduce an artificial player 0, nature, with action space $A_0 = S$ and constant payoffs; random state transitions correspond to the appropriate fixed behavior strategy for player 0. In what follows, the term *players* refers to players $i \geq 1$, and *histories* include nature's moves.

Let σ be a pure strategy profile and h^t a history. Then we write $U_i^*(\sigma \mid h^t)$ for player i's payoffs given history h^t and the subsequent continuation strategy profile $\sigma|_{h^t}$. This is in general an expected value because future utilities may depend randomly on past play, for the same reason that the current state in a model with exogenously specified states may depend randomly on past actions.

Note that $U_i^*(\sigma \mid h^t)$ is *not*, in general, the continuation payoff from $\sigma|_{h^t}$. For example, for a repeated game in the class of chapter 2,

$$U_i^*(\sigma \mid (a^0, a^1, \ldots, a^{t-1})) = \sum_{\tau=0}^{t-1} \delta^\tau u_i(a^\tau) + \delta^t U_i(\sigma|_{(a^0,a^1,\ldots,a^{t-1})}),$$

where u_i is the stage game payoff and U_i is given by (2.1.2). In this case, $U_i^*(\sigma \mid h^t)$ and $U_i^*(\sigma' \mid \hat{h}^t)$ differ by only a constant if the continuation strategies $\sigma|_{h^t}$ and $\sigma'|_{\hat{h}^t}$ are identical, and history is important only for its role in coordinating future behavior.

This dual role of histories in a dynamic game gives rise to ambiguity in defining states. Suppose two histories induce different continuation payoffs. Do these differences arise because differences in future play are induced, in which case the histories would not satisfy the usual notion of being payoff relevant (though it can still be critical to take note of the difference), or because identical continuation play gives rise to different payoffs? Can we always tell the difference?

5.6.1 Consistent Partitions

Let $\mathscr{H}^t$ be the set of period t histories. Notice that we have no notion of a state in this context, and hence no distinction between ex ante and ex post histories. A period t history is an element of A^t. A partition of $\mathscr{H}^t$ is denoted $\mathbb{H}^t$, and $\mathbb{H}^t(h^t)$ is the partition element containing history h^t.

A sequence of partitions $\{\mathbb{H}^t\}_{t=0}^\infty$ is denoted $\mathbb{H}$; viewed as $\cup_t \mathbb{H}^t$, $\mathbb{H}$ is a partition of the set of all histories $\mathscr{H} = \cup_t \mathscr{H}^t$. We often find it convenient to work with several such sequences, one associated with each player, denoting them by $\mathbb{H}_1, \ldots, \mathbb{H}_n$. Given such a collection of partitions, we say that two histories h^t and $\hat{h}^t$ are i-equivalent if $h^t \in \mathbb{H}_i^t(\hat{h}^t)$.

A strategy σ_i is *measurable* with respect to $\mathbb{H}_i$ if, for every pair of histories h^t and $\hat{h}^t$ with $h^t \in \mathbb{H}_i^t(\hat{h}^t)$, the continuation strategy $\sigma_i|_{h^t}$ equals $\sigma_i|_{\hat{h}^t}$. Let $\Sigma_i(\mathbb{H})$ denote the set of pure strategies for player i that are measurable with respect to the partition $\mathbb{H}$.

28. If this were not the case, then we would first partition the set of period t histories $\mathscr{H}^t$ into subsets that feature the same feasible choices for each player i in period t and then work throughout with refinements of this partition, to ensure that our subsequent measurability requirements were feasible.

A collection of partitions $\{\mathbb{H}_1^t, \ldots, \mathbb{H}_n^t\}_{t=0}^{\infty}$ is *consistent* if for every player i, whenever other players' strategies σ_{-i} are measurable with respect to their partition, then for any pair of i-equivalent period t histories h^t and $\hat{h}^t$, player i has the same preferences over i's continuation strategies. Hence, consistency requires that for any player i, pure strategies $\sigma_j \in \Sigma_j(\mathbb{H}_j)$ for all $j \neq i$, and i-equivalent histories h^t and $\hat{h}^t$, there exist constants θ and $\beta > 0$ such that

$$U_i^*((\sigma_i, \sigma_{-i}) \mid h^t) = \theta + \beta U_i^*((\sigma_i, \sigma_{-i}) \mid \hat{h}^t).$$

If this relationship holds, conditional on the (measurable) strategies of the other players, i's utilities after histories h^t and $\hat{h}^t$ are affine transformations of one another. We represent this by writing

$$U_i^*((\cdot, \sigma_{-i}) \mid h^t) \sim U_i^*((\cdot, \sigma_{-i}) \mid \hat{h}^t). \tag{5.6.1}$$

We say that two histories h^t and $\hat{h}^t$, with the property that player i has the same preferences over continuation payoffs given these histories (as just defined) are i-payoff equivalent. Consistency of a partition is thus the condition that equivalence (under the partition) implies payoff equivalence.

The idea now is to define a Markov equilibrium as a subgame-perfect equilibrium that is measurable with respect to a consistent collection of partitions. To follow this program through, two additional steps are required. First, we establish conditions under which consistent partitions have some intuitively appealing properties. Second, there may be many consistent partitions, some of them more interesting than others. We show that a maximal consistent partition exists, and use this one to define Markov equilibria.

5.6.2 Coherent Consistency

One might expect a consistent partition to have two properties. First, we might expect players to share the same partition. Second, we might expect the elements of the period t partition to be subsets of partition in period $t-1$, so that the partition is continually refined. Without some additional mild conditions, both of these properties can fail.

Lemma 5.6.1 *Suppose that for any players i and j, any period t, and any i-equivalent histories h^t and $\hat{h}^t$, there exists a repeated-game strategy profile σ and stage-game actions a_j and a'_j such that*

$$U_i^*((\cdot, \sigma_{-i}) \mid (h^t, a_j)) \not\sim U_i^*((\cdot, \sigma_{-i}) \mid (\hat{h}^t, a'_j)). \tag{5.6.2}$$

Then if $(\mathbb{H}_1, \ldots, \mathbb{H}_n)$ is a consistent collection of partitions, then in every period t and for all players i and j, $\mathbb{H}_i^t = \mathbb{H}_j^t$.

The expression $U_i^*((\sigma_i, \sigma_{-i}) \mid (h^t, a_j))$ gives player i's payoffs, given that h^t has occurred and given that player j chooses a_j in period t, with behavior otherwise specified by (σ_i, σ_{-i}). Condition (5.6.2) then requires that player i's preferences, given (h^t, a_j) and $(\hat{h}^t, a'_j)$, *not* be affine transformations of one another.

Proof Let h^t and $\hat{h}^t$ be i-equivalent. Choose a player j and suppose the strategy profile σ and actions a_j and a'_j satisfy (5.6.2). We suppose h^t and $\hat{h}^t$ are not j-equivalent (and derive a contradiction). Then player j's strategy of playing as in σ, except playing a_j after histories in $\mathbb{H}_j(h^t)$ and a'_j after histories in $\mathbb{H}_j(\hat{h}^t)$ is measurable with respect to $\mathbb{H}_j$. But then (5.6.2) contradicts (5.6.1): Player i's partition is not consistent (condition (5.6.2)), as assumed (condition (5.6.1)).

■

To see the argument behind this proof, suppose that a period t arrives in which player i and j partition their histories differently. We exploit this difference to construct a measurable strategy for player j that differs across histories within a single element of player i's partition, in a way that affects player i's preferences over continuation play. This contradicts the consistency of player i's partition. There are two circumstances under which such a contradiction may not arise. One is that all players have the same partition, precluding the construction of such a strategy. This leads to the conclusion of the theorem. The other possibility is that we may not be able to find the required actions on the part of player j that affect i's preferences. In this case, we have reached a point at which, given a player i history $h^t \in \mathbb{H}_i(h^t)$, there is nothing player j can do in period t that can have any effect on how player i evaluates continuation play. Such degeneracies are possible (Maskin and Tirole, 2001, provide an example), but we hereafter exclude them, assuming that the sufficient conditions of lemma 5.6.1 hold throughout.

We can thus work with a single consistent partition $\mathbb{H}$ and can refer to histories as being "equivalent" and "payoff equivalent" rather than i-equivalent and i-payoff equivalent.

We are now interested in a similar link between periods.

Lemma 5.6.2 *Suppose that for any players i and j, and period t, any equivalent histories h^t and $\hat{h}^t$, and any stage-game action profile a^t, there exists a repeated-game strategy profile σ and player j actions a_j^{t+1} and $\tilde{a}_j^{t+1}$ such that*

$$U_i^*((\cdot, \sigma_{-i}) \mid (h^t, a_{-j}^t, a_j^{t+1})) \not\sim U_i^*((\cdot, \sigma_{-i}) \mid (\hat{h}^t, a_{-j}^t, \tilde{a}_j^{t+1})). \tag{5.6.3}$$

If $(\mathbb{H}_1, \ldots, \mathbb{H}_n)$ is a consistent collection of partitions under which h^t and $\hat{h}^t$ are equivalent histories, then for any action profile a^t, (h^t, a^t) and $(\hat{h}^t, a^t)$ are equivalent.

Proof Fix a consistent collection $\mathbb{H} = (\mathbb{H}_1, \ldots, \mathbb{H}_n)$ and an action profile a^t. Suppose h^t and $\hat{h}^t$ are equivalent, and the action profile a^t, strategy σ, and player j actions a_j^{t+1} and $\tilde{a}_j^{t+1}$ satisfy (5.6.3). The strategy for every player k other than i and j that plays according to σ_k, except for playing a_k^t in period t, is measurable with respect to $\mathbb{H}$. We now suppose that (h^t, a^t) and $(\hat{h}^t, a^t)$ are not equivalent and derive a contradiction. In particular, the player j strategy of playing a_j^{t+1} and then playing according to σ_j, after any history $h^{t+1} \in \mathbb{H}((h^t, a^t))$, and otherwise playing $\tilde{a}_j^{t+1}$ (followed by σ_j) is then measurable with respect to $\mathbb{H}$ and from (5.6.3), allows us to conclude that h^t and $\hat{h}^t$ are not equivalent (recall (5.6.1)), the contradiction.

■

The conditions of this lemma preclude cases in which player j's behavior in period $t+1$ has no effect on player i's period t continuation payoffs. If the absence of such an effect, the set of payoff-relevant states in period $t+1$ can be coarser than the set in period t.[29]

We say that games satisfying the conditions of lemmas 5.6.1 and 5.6.2 are *nondegenerate* and hereafter restrict attention to such games.

5.6.3 Markov Equilibrium

There are typically many consistent partitions. The trivial partition, in which every history constitutes an element, is automatically consistent. There is clearly nothing to be gained in defining a Markov equilibrium to be measurable with respect to this collection of partitions, because every strategy would then be Markov. Even if we restrict attention to nontrivial partitions, how do we know which one to pick?

The obvious response is to examine the maximally coarse consistent partition, meaning a consistent partition that is coarser than any other consistent partition.[30] This will impose the strictest version of the condition that payoff-irrelevant events should not matter. Does such a partition exist?

Proposition 5.6.1 *Suppose the game is nondegenerate (i.e., satisfies the hypotheses of lemmas 5.6.1 and 5.6.2). A maximally coarse consistent partition exists. If the stage game is finite, then this maximally coarse consistent partition is unique.*

Proof Let Φ be the set of all consistent partitions of histories. Endow this set with the partial order $\prec$ defined by $\hat{\mathbb{H}} \prec \mathbb{H}$ if $\mathbb{H}$ is a coarsening of $\hat{\mathbb{H}}$. We show that there exists a maximal element under this partial order, unique for finite games.

This argument proceeds in two steps. The first is to show that there exist maximal elements. This in turn follows from Zorn's lemma (Hrbacek and Jech 1984, p. 171), if we can show that every chain (i.e., totally ordered subset) $\mathscr{C} = \{\mathbb{H}_{(m)}\}_{m=1}^{\infty}$ admits an upper bound. Let $\mathbb{H}_{(\infty)}$ denote the finest common coarsening (or meet) of $\mathscr{C}$, that is, for each element $h \in \mathscr{H}$, $\mathbb{H}_{(\infty)}(h) = \cup_{m=1}^{\infty} \mathbb{H}_{(m)}(h)$. Because $\mathscr{C}$ is a chain, $\mathbb{H}_{(m)}(h) \subset \mathbb{H}_{(m+1)}(h)$, and so $\mathbb{H}_{(\infty)}$ is a partition that is coarser than every partition in $\mathscr{C}$. It remains to show that $\mathbb{H}_{(\infty)}$ is consistent. To do this, suppose that two histories h^t and $\hat{h}^t$ are contained in a common element of $\mathbb{H}_{(\infty)}$. Then they must be contained in some common element of $\mathbb{H}_{(m)}$ for some m, and hence must satisfy (5.6.1). This ensures that $\mathbb{H}_{(\infty)}$ is consistent, and so is an upper bound for the chain. Hence by Zorn's lemma, there is a maximally coarse consistent partition.

The second step is to show that there is a unique maximal element for finite stage games. To do this, it suffices to show that for any two consistent partitions, their meet (i.e., finest common coarsening) is consistent. Let $\mathbb{H}$ and $\hat{\mathbb{H}}$ be consistent partitions, and let $\bar{\mathbb{H}}$ be their meet. Suppose h^t and $\hat{h}^t$ are contained in a single element of $\bar{\mathbb{H}}$. Because the stage game is finite, $\mathbb{H}^t$ and $\hat{\mathbb{H}}^t$ are both finite partitions. Then, by the definition of meet, there is a finite sequence of histories

29. Again, Maskin and Tirole (2001) provide an example.

30. A partition $\mathbb{H}'$ is a *coarsening* of another partition $\mathbb{H}$ if for all $H \in \mathbb{H}$ there exists $H' \in \mathbb{H}'$ such that $H \subset H'$.

$\{h^t, h^t(1), \ldots, h^t(n), \hat{h}^t\}$ such that each adjacent pair is contained in either the same element of $\mathbb{H}^t$ or the same element of $\hat{\mathbb{H}}^t$, and hence satisfy payoff equivalence. But then h^t and $\hat{h}^t$ must be payoff equivalent, which suffices to conclude that $\bar{\mathbb{H}}$ is consistent. ■

Denote the maximally coarse consistent partition by $\mathbb{H}^*$.

Definition 5.6.1 *A strategy profile σ is a* Markov strategy *profile if it is measurable with respect to the maximally coarse consistent partition $\mathbb{H}^*$. A strategy profile is a* Markov equilibrium *if it is a subgame-perfect equilibrium and it is Markov. Elements of the partition $\mathbb{H}^*$ are called* Markov states *or* payoff-relevant histories.

No difficulty arises in finding a Markov equilibrium in a repeated game, because one can always simply repeat the Nash equilibrium of the stage game, making history completely irrelevant. This is a reflection of the fact that in a repeated game, the maximally coarse partition is the set of all histories. Indeed, all Markov equilibria feature a Nash equilibrium of the stage game in every period.

Repeated games have the additional property that every history gives rise to an identical continuation game. As we noted in section 5.5.2, the Markov condition on strategies is commonly supplemented with the additional requirement that identical continuation games feature identical continuation strategies. Such strategies are said to be *stationary*. A stationary Markov equilibria in a repeated game must feature the same stage-game Nash equilibrium in every period.

More generally, it is straightforward to establish the existence of Markov equilibria in dynamic games with finite stage games and without private information. A backward induction argument ensures that finite horizon versions of the game have Markov equilibria, and discounting ensures that the limit of such equilibria, as the horizon approaches infinity, is a Markov equilibrium of the infinitely repeated game (Fudenberg and Levine 1983).

Now consider dynamic games G in the class described in section 5.5. The set S induces a partition on the set of ex post histories in a natural manner, with two ex post histories being equivalent under this partition if they are histories of the same length and end with the same state $s \in S$. Refer to this partition as $\mathbb{H}^S$. Because the continuation $G(s)$ is identical, regardless of the history terminating in s, the following is immediate:

Proposition 5.6.2 *Suppose G is a dynamic game in the class described in section 5.5. Suppose $\mathbb{H}^S$ is the partition of $\mathscr{H}$ with $h \in \mathbb{H}^S(h')$ if h and h' are of the same length and both result in the same state $s \in S$. Then, $\mathbb{H}^S$ is finer than $\mathbb{H}^*$. If for every pair of states $s, s' \in S$, there is at least one player i for which $u_i(s, a)$ and $u_i(s', a)$ are not affine transformations of one another, then $\mathbb{H}^* = \mathbb{H}^S$.*

The outcomes ω and ω' of a public correlating device have no effect on players' preferences and hence fail the condition that there exist a player i for whom $u_i(\omega, a)$ and $u_i(\omega', a)$ are not affine transformations of one another. Therefore, the outcomes of a public correlating device in a repeated game do not constitute states.

Markov equilibrium precludes the use of public correlation in repeated games and restricts the players in the prisoners' dilemma to consistent shirking. Alternatively, much of the interest in repeated games focuses on non-Markov equilibria.[31] In our view, the choice of an equilibrium is part of the construction of the model. Different choices, including whether Markov or not, may be appropriate in different circumstances, with the choice of equilibrium to be defended not within the confines of the model but in terms of the strategic interaction being modeled.

Remark 5.6.1 **Games of incomplete information** The notion of Markov strategy also plays an important role in incomplete information games, where the beliefs of uninformed players are often treated as Markov states (see, for example, section 18.4.4). At an intuitive level, this is the appropriate extension of the ideas in this section. However, determining equivalence classes of histories that are payoff-equivalent is now a significantly more subtle question. For example, because the inferences that players draw from histories depend on the beliefs that players have about past play, the equivalence classes now must satisfy a complicated fixed point property. Moreover, Markov equilibria (as just defined) need not exist, and this has led to the notion of a weak Markov equilibrium in the literature on bargaining under incomplete information (see, for example, Fudenberg, Levine, and Tirole, 1985). We provide a simple example of a similar phenomenon in section 17.3.

◆

5.7 Dynamic Games: Equilibrium

5.7.1 The Structure of Equilibria

This section explores some of the common ground between ordinary repeated games and dynamic games. Recall that we assume either that the set of states S is finite, or that the transition function is deterministic. The proofs of the various propositions we offer are straightforward rewritings of their counterparts for repeated games in chapter 2 and hence are omitted.

We say that strategy $\hat{\sigma}_i$ is a one-shot deviation for player i from strategy σ_i if there is a unique ex post history $\tilde{h}^t$ such that

$$\hat{\sigma}_i(\tilde{h}^t) \neq \sigma_i(\tilde{h}^t).$$

It is then a straightforward modification of proposition 2.2.1, substituting ex post histories for histories and replacing payoffs with expected payoffs to account for the potential randomness of the state transition function, to establish a one-shot deviation principle for dynamic games:

Proposition 5.7.1 *A strategy profile σ is subgame perfect in a dynamic game if and only if there are no profitable one-shot deviations.*

Given a dynamic game, an automaton is $(\mathscr{W}, \mathbf{w}^0, \tau, f)$, where $\mathscr{W}$ is the set of automaton states, $\mathbf{w}^0 : S \to \mathscr{W}$ gives the initial automaton state as a function of the initial game state, $\tau : \mathscr{W} \times A \times S \to \mathscr{W}$ is the transition function giving the automaton state in the next period as a function of the current automaton state, the current action profile, and the next draw of the game state. Finally, $f : \mathscr{W} \to \prod_i \Delta(A_i)$ is the output function. (Note that this description agrees with remark 2.3.3 when S is the space of realizations of the public correlating device.)

The initial automaton state is determined by the initial game state, through the function $\mathbf{w}^0$. We will often be interested in the strategy induced by the automaton beginning with an automaton state w. As in the case of repeated games, we write this as $(\mathscr{W}, w, \tau, f)$.[32]

Let $\tau(\tilde{h}^t)$ denote the automaton state reached under the ex post history $\tilde{h}^t \in \tilde{\mathscr{H}}^t = (S \times A)^t \times S$. Hence, for a history $\{s\}$ that identifies the initial game-state s, we have

$$\tau(\{s\}) = \mathbf{w}^0(s)$$

and for any ex post history $\tilde{h}^t = (\tilde{h}^{t-1}, a, s)$,

$$\tau(\tilde{h}^t) = \tau(\tau(\tilde{h}^{t-1}), a, s).$$

Given an ex post history $\tilde{h}^t \in \tilde{\mathscr{H}}$, let $s(\tilde{h}^t)$ denote the current game state in $\tilde{h}^t$. Given a game state $s \in S$, the set of automaton states *accessible in game state* s is $\mathscr{W}(s) = \{w \in \mathscr{W} : \exists \tilde{h}^t \in \tilde{\mathscr{H}}, w = \tau(\tilde{h}^t), s = s(\tilde{h}^t)\}$.[33]

Proposition 5.7.2 *Suppose the strategy profile σ is described by the automaton $\{\mathscr{W}, \mathbf{w}^0, \tau, f\}$. Then σ is a subgame-perfect equilibrium of the dynamic game if and only if for any game state $s \in S$ and automaton state w accessible in game state s, the strategy profile induced by $\{\mathscr{W}, w, \tau, f\}$, is a Nash equilibrium of the dynamic game $G(s)$.*

Our next task is to develop the counterpart for dynamic games of the recursive methods for generating equilibria introduced in section 2.5 for repeated games. Restrict attention to finite sets of signals (with $|S| = m$) and pure strategies. For each game state $s \in S$, and each state $w \in \mathscr{W}(s)$, associate the profile of values $V_s(w)$, defined by

$$V_s(w) = (1-\delta)u(s, f(w)) + \delta \sum_{s' \in S} V_{s'}(\tau(w, f(w), s'))q(s' \mid s, f(w)).$$

As is the case with repeated games, $V_s(w)$ is the profile of expected payoffs when beginning the game in game state s and automaton state w. Associate with each game state $s \in S$, and each state $w \in \mathscr{W}(s)$, the function $g^{(s,w)}(a) : A \to \mathbb{R}^n$, where

$$g^{(s,w)}(a) = (1-\delta)u(s, a) + \delta \sum_{s' \in S} V_{s'}(\tau(w, a, s'))q(s' \mid s, a).$$

Proposition 5.7.3 *Suppose the strategy profile σ is described by the automaton $(\mathscr{W}, \mathbf{w}^0, \tau, f)$. Then σ is a subgame-perfect equilibrium if and only if for all game states $s \in S$ and all $w \in \mathscr{W}(s)$, $f(w)$ is a Nash equilibrium of the normal-form game with payoff function $g^{(s,w)}$.*

31. This contrast may not be so stark. Maskin and Tirole (2001) show that most non-Markov equilibria of repeated games are limits of Markov equilibria in nearby dynamic games.

32. Hence, $(\mathscr{W}, \mathbf{w}^0, \tau, f)$ is an automaton whose initial state is specified as a function of the game state, $(\mathscr{W}, \mathbf{w}^0(s), \tau, f)$ is the automaton whose initial state is given by $\mathbf{w}^0(s)$, and $(\mathscr{W}, w, \tau, f)$ is the automaton whose initial state is fixed at an arbitrary $w \in \mathscr{W}$.

33. Note that $\mathbf{w}^0(s)$ is thus accessible in game state s, even if game state s does not have positive probability under $q(\cdot \mid \varnothing)$.

Let $\mathscr{W}^s$ be a subset of $\mathbb{R}^n$, for $s = 1, \ldots, m$. We interpret this set as a set of feasible payoffs in dynamic game $G(s)$. We say that the pure action profile a^* is *pure-action enforceable* on $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$ given $s \in S$ if there exists a function $\gamma : A \times S \to \mathscr{W}^1 \cup \ldots \cup \mathscr{W}^m$ with $\gamma(a, s') \in \mathscr{W}^{s'}$ such that for all players i and all $a_i \in A_i$,

$$(1-\delta)u_i(s, a^*) + \delta \sum_{s' \in S} \gamma_i(a^*, s')q(s' \mid s, a^*)$$
$$\geq (1-\delta)u_i(s, a_i, a^*_{-i}) + \delta \sum_{s' \in S} \gamma_i(a_i, a^*_{-i}, s')q(s' \mid s, a_i, a^*_{-i}).$$

We say that the payoff profile $v \in \mathbb{R}^n$ is *pure-action decomposable* on $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$ given $s \in S$ if there exists a pure action profile a^* that is pure-action enforceable on $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$ given s, with the enforcing function γ satisfying, for all players i,

$$v_i = (1-\delta)u_i(s, a^*) + \delta \sum_{s' \in S} \gamma_i(a^*, s')q(s' \mid s, a^*).$$

A vector of payoff profiles $\mathbf{v} \equiv (v(1), \ldots, v(m)) \in \mathbb{R}^{nm}$, with $v(s)$ interpreted as a payoff profile in game $G(s)$, is *pure-action decomposable* on $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$ if for all $s \in S$, $v(s)$ is pure-action decomposable on $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$ given s. Finally, $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$ is *pure-action self-generating* if every vector of payoff profiles in $\prod_{s \in S} \mathscr{W}^s$ is pure-action decomposable on $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$. We then have:[34]

Proposition 5.7.4 *Any self-generating set of payoffs $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$ is a set of pure-strategy subgame-perfect equilibrium payoffs.*

As before, we have the corollary:

Corollary 5.7.1 *The set $(\mathscr{W}^{1*}, \ldots, \mathscr{W}^{m*})$ of pure-strategy subgame-perfect equilibrium payoff profiles is the largest pure-action self-generating collection $(\mathscr{W}^1, \ldots, \mathscr{W}^m)$.*

Remark 5.7.1 **Repeated games with random states** For these games (see remark 5.5.1), the set of ex ante feasible payoffs is independent of last period's state and action profile. Consequently, it is simpler to work with ex ante continuations in the notions of enforceability, decomposability, and pure-action self-generation. A pure action profile a^* is *pure-action enforceable* in state $s \in S$ on $\mathscr{W} \subset \mathbb{R}^n$ if there exists a function $\gamma : A \to \mathscr{W}$ with such that for all players i and all $a_i \in A_i$,

$$(1-\delta)u_i(s, a^*) + \delta\gamma_i(a^*) \geq (1-\delta)u_i(s, a_i, a^*_{-i}) + \delta\gamma_i(a_i, a^*_{-i}).$$

The ex post payoff profile $v^s \in \mathbb{R}^n$ is *pure-action decomposable* in state s on $\mathscr{W}$ if there exists a pure-action profile a^* that is pure-action enforceable in s on $\mathscr{W}$ with the enforcing function γ satisfying $v^s = (1-\delta)u(s, a^*) + \delta\gamma(a^*)$. An ex ante payoff profile $v \in \mathbb{R}^n$ is *pure-action decomposable* on $\mathscr{W}$ if there exist ex post payoffs $\{v^s : s \in S\}$, v^s pure-action decomposable in s on $\mathscr{W}$, such that $v = \sum_s v^s q(s)$. Finally, $\mathscr{W}$ is *pure-action self-generating* if every payoff profile in $\mathscr{W}$ is pure-action decomposable on $\mathscr{W}$. As usual, any self-generating set of ex ante payoffs is a set of subgame-perfect equilibrium ex ante payoffs, with the set of subgame-perfect equilibrium ex ante payoffs being the largest such set.

Section 6.3 analyzes a repeated game with random states in which players have an opportunity to insure one another against endowment shocks. Interestingly, this game is an example in which the efficient symmetric equilibrium outcome is sometimes necessarily nonstationary. It is no surprise that the equilibrium itself might not be stationary, that is, that efficiency might call for nontrivial intertemporal incentives and their attendant punishments. However, we have the stronger result that efficiency cannot be obtained with a stationary-outcome equilibrium.

◆

34. See section 9.7 on games of symmetric incomplete information (in particular, proposition 9.7.1 and lemma 9.7.1) for an application.

5.7.2 A Folk Theorem

This section presents a folk theorem for dynamic games. We assume that the action spaces and the set of states are finite.

Let Σ be the set of pure strategies in the dynamic game and let Σ^M be the set of pure Markov strategies. For any $\sigma \in \Sigma$, let

$$U^\delta(\sigma) = E^\sigma\left\{(1-\delta)\sum_{t=0}^{\infty} \delta^t u(s^t, a^t)\right\}$$

be the expected payoff profile under strategy σ, given discount factor δ. The expectation accounts not only for the possibility of private randomization and public correlation but also for randomness in state transitions, including the determination of the initial state. We let

$$U^\delta(\sigma \mid h^t) = E^{\sigma, h^t}\left\{(1-\delta)\sum_{\tau=t}^{\infty} \delta^{\tau-t} u(a^\tau, s^\tau)\right\}$$

be the analogous expectation conditioned on having observed the history h^t, where continuation play is given by $\sigma|_{h^t}$. As a special case of this, we have the expected payoff $U^\delta(\sigma \mid s)$, which conditions on the initial state realization s.

Let

$$\mathscr{F}(\delta) = \{\mathbf{v} \in \mathbb{R}^{nm} : \exists \sigma \in \Sigma^M \text{ s.t. } U^\delta(\sigma \mid s) = v(s)\ \forall s\}.$$

This is our counterpart of the set of payoff profiles produced by pure stage-game actions in a repeated game. There are two differences here. First, we identify functions that map from initial states to expected payoff profiles. Second, we now work directly with the collection of repeated-game payoffs rather than with stage-game payoffs because we have no single underlying stage game. In doing so, we have restricted attention only to payoffs produced by pure Markov strategies. We comment shortly on the reasons for doing so, and in the penultimate paragraph of this section on the sense in which this assumption is not restrictive.

We let

$$\mathscr{F} = \lim_{\delta \to 1} \mathscr{F}(\delta). \tag{5.7.1}$$

This is our candidate for the set of feasible pure-strategy payoffs available to patient players. We similarly require a notion of minmax payoffs, which we define contingent on the current state,

$$\underline{v}_i^\delta(s) = \inf_{\sigma_{-i} \in \Sigma_{-i}} \sup_{\sigma_i \in \Sigma_i} U_i^\delta(\sigma \mid s)$$

with

$$\underline{v}_i(s) = \lim_{\delta \to 1} \underline{v}_i^\delta(s). \tag{5.7.2}$$

Dutta (1995, lemma 2, lemma 4) shows that the limits in (5.7.1) and (5.7.2) exist. The restriction to Markov strategies in defining $\mathscr{F}(\delta)$ is useful here, as it is relatively easy to show that the payoffs to a pure Markov strategy, in the presence of finite sets of actions and states, are continuous as $\delta \to 1$.

Finally, we say that the collection of pure strategies $\{\sigma^1, \ldots, \sigma^n\}$ is a *player-specific punishment* for $\mathbf{v}$ if the limits $U(\sigma^i \mid s) = \lim_{\delta \to 1} U^\delta(\sigma^i \mid s)$ exist and the following hold for all s, s', and s'' in S:

$$U_i(\sigma^i \mid s') < v_i(s) \tag{5.7.3}$$

and

$$\underline{v}_i(s'') < U_i(\sigma^i \mid s') < U_i(\sigma^j \mid s). \tag{5.7.4}$$

We do not require the player-specific punishments to be Markov. The inequalities in conditions for player-specific punishments are required to hold uniformly across states. This imposes a tremendous amount of structure on the payoffs involved in these punishments. We comment in the final paragraph of this section on sufficient conditions for the existence of such punishments.

We then have the pure-strategy folk theorem.

Proposition 5.7.5 *Let $\mathbf{v} \in \mathscr{F}$ be strictly individually rational, in the sense that for all players i and pairs of states s and s' we have*

$$v_i(s) > \underline{v}_i(s'),$$

let $\mathbf{v}$ admit a player specific punishment, and suppose that the players have access to a public correlating device. Then, for any $\varepsilon > 0$, there exists $\underline{\delta}$ such that for all $\delta \in (\underline{\delta}, 1)$, there exists a subgame-perfect equilibrium σ whose payoffs $U(\sigma \mid s)$ are within ε of $v(s)$ for all $s \in S$.

The proof of this proposition follows lines that are familiar from proposition 3.4.1 for repeated games. The additional complication introduced by the dynamic game is that there may now be two reasons to deviate from an equilibrium strategy. One is to obtain a higher current payoff. The other, not found in repeated games, is to affect the transitions of the state. In addition, this latter incentive potentially becomes more powerful as the players become more patient, and hence the benefits of affecting the future state become more important. We describe the basic structure of the proof. Dutta (1995) can be consulted for more details.

Proof We fix a sequence of values of δ approaching 1 and corresponding strategy profiles $\sigma(\delta)$ with the properties that

$$\lim_{\delta \to 1} U^\delta(\sigma(\delta) \mid s) = v(s).$$

This sequence allows us to approach the desired equilibrium payoffs. Moreover, Dutta (1995, proposition 3) shows that there exists a strategy profile $\hat{\sigma}$ such that for any $\eta > 0$, there exists an integer $L(\eta)$ such that for all $T \geq L(\eta)$ and $s \in S$,

$$E^{\hat{\sigma},s} \frac{1}{T} \sum_{t=0}^{T-1} u_i(s^t, a^t) \leq \underline{v}_i^\delta(s) + \frac{\eta}{2},$$

and hence a value $\delta_1 < 1$ such that for $\delta \in (\delta_1, 1)$,

$$E^{\hat{\sigma},s} \frac{1}{T} \sum_{t=0}^{T-1} u_i(s^t, a^t) \leq \underline{v}_i(s) + \eta.$$

Let $\{\sigma^1, \ldots, \sigma^n\}$ be the player-specific punishment for $\mathbf{v}$. Conditions (5.7.3) and (5.7.4) ensure that we can fix $\delta_2 \in (\delta_1, 1)$, $\underline{v}_i \equiv \max_{s \in S} \underline{v}_i(s)$ and η sufficiently small that, for all $\delta \in (\delta_2, 1)$, all i, and for any states s, s' and s'',

$$\underline{v}_i(s) + \eta \leq \underline{v}_i + \eta < U_i^\delta(\sigma^i \mid s') < v_i(s'')$$

and

$$U_i^\delta(\sigma^i \mid s) < U_i^\delta(\sigma^j \mid s').$$

We now note that we can assume, for each player i, that the player-specific punishment σ^i has the property that there exists a length of time T_i such that for each $t, t' = 0, T_i, 2T_i, \ldots$, and for all ex ante histories h^t and $h^{t'}$ and states s, $\sigma^i|_{\{h^t, s\}} = \sigma^i|_{\{h^{t'}, s\}}$. Hence, σ^i has a cyclical structure, erasing its history and starting from the beginning every T_i periods.[35] We can also assume that σ^i provides a payoff to player i that is independent of its initial state.[36] We hereafter retain these properties for the player specific punishments.

The strategy profile now mimics that used to prove proposition 3.4.1, the corresponding result for repeated games. It begins with play following the strategy profile $\sigma(\delta)$; any deviation by player i from $\sigma(\delta)$, and indeed any deviation from any subsequent equilibrium prescription other than deviations from being

35. Suppose σ^i does not have this property. Because each inequality in (5.7.3) and (5.7.4) holds by at least ε, for some ε, we need only choose T_i sufficiently large that the average payoff from any strategy profile over its first T_i periods is within at least $\varepsilon/3$ of its payoff. Now construct a new strategy by repeatedly playing the first T_i periods of σ^i, beginning each time with the null history. This new strategy has the desired cyclic structure and satisfies (5.7.3) and (5.7.4). Dutta (1995, section 6.2) provides details.

36. Suppose this is not the case. Let s maximize $U_i(\sigma^i \mid s)$. Define a new strategy as follows: In each period $0, T_i, 2T_i, \ldots$, conduct a state-contingent public correlation that mixes between σ^i and a strategy that maximizes player i's repeated-game payoff, with the correlation set so as to equate the continuation payoff for each state s' with $U_i(\sigma^i \mid s)$. The uniform inequalities of the player-specific punishments ensure this is possible.

minmaxed, which are ignored, prompts the first L periods of the corresponding minmax strategy $\hat{\sigma}^i(\delta)$, followed by the play of the player-specific punishment σ^i.

Our task now is to show that these strategies constitute a subgame-perfect equilibrium, given the freedom to restrict attention to large discount factors and choose $L \geq L(\eta)$. As usual, let M and m be the maximum and minimum stage-game payoffs.

The condition for deviations from the equilibrium path to be unprofitable is that, for any state $s \in S$ (suppressing the dependence of strategies on δ)

$$(1-\delta)M + \delta(1-\delta^L)(\underline{v}_i + \eta) + \delta^{L+1}U_i^\delta(\sigma^i) \leq U_i^\delta(\sigma \mid s),$$

which, as δ converges to one for fixed L, becomes $U_i(\sigma^i) \leq v_i(s)$, which holds with strict inequality by virtue of our assumption that $\mathbf{v}$ admits player-specific punishments. There is then a value $\delta_3 \in [\delta_2, 1)$ such that this constraint holds for any $\delta \in (\delta_3, 1)$ and $L \geq L(\eta)$.

For player j to be unwilling to deviate while minmaxing i, the condition is

$$(1-\delta)M + \delta(1-\delta^L)(\underline{v}_j + \eta) + \delta^{L+1}U_j^\delta(\sigma^j) \leq (1-\delta^L)m + \delta^L U_j^\delta(\sigma^i).$$

Rewrite this condition as

$$(1-\delta)M + (1-\delta^L)(\delta(\underline{v}_j + \eta) - m) + \delta^L[\delta U_j^\delta(\sigma^j) - U_j^\delta(\sigma^i)] \leq 0.$$

The term $[\delta U_j^\delta(\sigma^j) - U_j^\delta(\sigma^i)]$ converges to $U_j(\sigma^j) - U_j(\sigma^i) < 0$ as $\delta \to 1$. We can then find a value $\delta_4 \in [\delta_3, 1)$ and an increasing function $L(\delta)$ ($\geq L(\eta)$) such that this constraint holds for any $\delta \in (\delta_4, 1)$ and the associated $L(\delta)$, and such that $\delta^{L(\delta)} < 1 - \gamma$, for some $\gamma > 0$. We hereafter take L to be given by $L(\delta)$.

Now consider the postminmaxing rewards. For player i to be willing to play σ^i, a sufficient condition is that for any ex post history $\tilde{h}^t$ under which current play is governed by σ^i,

$$(1-\delta)M + \delta(1-\delta^{L(\delta)})(\underline{v}_i + \eta) + \delta^{L(\delta)+1}U_i^\delta(\sigma^i) \leq U_i^\delta(\sigma^i \mid \tilde{h}^t).$$

This inequality is not obvious. The difficulty here is that we cannot exclude the possibility that $U_i^\delta(\sigma^i) > U_i^\delta(\sigma^i \mid \tilde{h}^t)$. There is no reason to believe that player i's payoff from strategy σ^i is constant across time or states. Should player i find himself at an ex post history (h^t, s) in which this strategy profile gives a particularly low payoff, i may find it optimal to deviate, enduring the resulting minmaxing to return to the relatively high payoff of beginning σ^i from the beginning. This is the incentive to deviate to affect the state that does not appear in an ordinary repeated game. In addition, this incentive seemingly only becomes stronger as the player gets more patient, and hence the intervening minmaxing becomes less costly.

A similar issue arises in the proof of proposition 3.8.1, the folk theorem for repeated games without public correlation, where we faced the fact that the deterministic sequences of payoffs designed to converge to a target payoff may feature continuation values that differ from the target. In the case of proposition 3.8.1,

the response involved a careful balancing of the relative sizes of δ and L. Here, we can use the cyclical nature of the strategy σ^i to rewrite this constraint as

$$(1-\delta)M + \delta(1-\delta^{L(\delta)})(\underline{v}_i + \eta) + \delta^{L(\delta)+1}U_i^\delta(\sigma^i) \leq (1-\delta^{T_i})m + \delta^{T_i}U_i^\delta(\sigma^i),$$

where $(1-\delta^{T_i})m$ is a lower bound on player i's payoff from σ^i over T_i periods, and then the strategy reverts to payoff $U_i^\delta(\sigma^i)$. The key to ensuring this inequality is satisfied is to note that T_i is fixed as part of the specification of σ^i. As a result, $\lim_{\delta\to 1}\delta^{T_i} = 1$ while $\delta^{L(\delta)}$ remains bounded away from 1.

A similar argument establishes that a sufficiently patient player i has no incentive to deviate when in the middle of strategy σ^j. This argument benefits from the fact that a deviation trades a single-period gain for L periods of being minmaxed followed by a return to the less attractive payoff $U_i^\delta(\sigma^i)$. Letting $\delta_5 \in (\delta_4, 1)$ be the bound on the discount factor to emerge from these two arguments, these strategies constitute a subgame-perfect equilibrium for all $\delta \in (\delta_5, 1)$.

■

We have worked throughout with pure strategies and with pure Markov strategies when defining feasible payoffs. Notice first that these are pure strategies in the dynamic game. We are thus not restricting ourselves to the set of payoffs that can be achieved in pure stage-game actions. This makes the pure strategy restriction less severe than it may first appear. In addition, any feasible payoff can be achieved by publicly mixing over pure Markov strategies (Dutta 1995, lemma 1), so that the Markov restriction is also not restrictive in the presence of public correlation (which we used in modifying the player-specific punishments in the proof).

Another aspect of this proposition can be more directly traced to the dynamic structure of the game. We have worked with a function $\mathbf{v}$ that specifies a payoff profile for each state. Suppose instead we defined, for each $s \in S$,

$$\mathscr{F}(\delta, s) = \{v \in \mathbb{R}^n : \exists \sigma \in \Sigma^M \text{ s.t. } U^\delta(\sigma \mid s) = v\},$$

the set of feasible payoffs (in pure Markov strategies) given initial state s, with $\mathscr{F}(s) = \lim_{\delta\to 1}\mathscr{F}(\delta, s)$. Let us say that a stochastic game is *communicating* if for any pair of states s and s', there is a strategy σ and a time t such that if the game begins in state s, there is positive probability under strategy σ that the game is in state s' in period t. Dutta (1995, lemma 12) shows that in communicating games, $\mathscr{F}(s)$ is independent of s. If the game communicates independently of the actions of player i, for each i, then minmax values will also be independent of the state. In this case, we can formulate the folk theorem for stochastic games in terms of payoff profiles $v \in \mathbb{R}^n$ and minmax profiles $\underline{v}_i$ that do not depend on the initial state. In addition, full dimensionality of the convex hull of $\mathscr{F}$ then suffices for the existence of player-specific punishments for interior v. We can establish a similar result in games in which $\mathscr{F}(\delta, s)$ depends on s (and hence which are not communicating), in terms of payoffs that do not depend on the initial state by concentrating on those payoffs in the set $\cap_{s\in S}\mathscr{F}(s)$.