GAMES AND ECONOMIC BEHAVIOR 1, 5–39 (1989)

Cooperation and Bounded Recall*

ROBERT J. AUMANN

Institute of Mathematics, The Hebrew University, 91904 Jerusalem, Israel

AND

SYLVAIN SORIN

Département de Mathématiques, Université Louis Pasteur, 67084 Strasbourg, France

A two-person game has *common interests* if there is a single payoff pair z that strongly Pareto dominates all other payoff pairs. Suppose such a game is repeated many times, and that each player attaches a small but positive probability to the other playing some fixed strategy with bounded recall, rather than playing to maximize his payoff. Then the resulting supergame has an equilibrium in pure strategies, and the payoffs to all such equilibria are close to optimal (i.e., to z).

1. INTRODUCTION

Is cooperation rational? Can the principle of cooperation be derived from basic principles of rational behavior on the part of individuals, such as strategic ("Nash") equilibrium?

* Aumann's research was partially supported by NSF Grant IST-85-21838 at IMSSS, Stanford, and by MSRI, Berkeley. Sorin's research was partially supported by MSRI, Berkeley. It is the extraordinary research atmosphere at these institutions that made this work possible.

0899-8256/89 $3.00

2,2	0,0
0,0	1,1

FIGURE 1

Without further qualification, this question must be answered in the negative. Nothing seems more obvious than that in the game of Fig. 1, rational, self-seeking players should play for (2,2). Yet the strategy pair yielding (1,1) is a perfectly good equilibrium.

Call an outcome of a strategic ("normal form") game *cooperative* if it is efficient ("Pareto optimal") in the set of all feasible outcomes. The folk theorem tells us that in a repeated game, cooperation is possible in equilibrium; that the cooperative outcomes are among the equilibrium outcomes. But equilibrium does not *assure* cooperation, even in a repeated game; in almost all cases there are also equilibrium outcomes that are not efficient. Figure 1 shows that even in games with *common payoffs*—defined as games in which all players get the same payoff at each outcome—there are equilibrium outcomes that are not cooperative; and this remains so when the game is repeated. Thus when the game of Fig. 1 is repeated, (1,1) remains an equilibrium payoff.

It is the purpose of the research reported here to identify a model in which *all* equilibrium outcomes are efficient, in which utility-maximizing behavior on the part of each separate individual *necessarily* leads to cooperation.

2. VERBAL DESCRIPTION OF THE MAIN RESULT

We start with a two-person game with *common interests*, defined as a game in which there is one payoff pair that strongly Pareto dominates all other payoff pairs. The games with common payoffs to which we referred above have common interests, but they are not the only ones. For example, the game of Fig. 2 has common interests but not common payoffs.

Let G^∞ be a repetition of G, which may be either a long finite repetition

9,9	0,8
8,0	7,7

FIGURE 2

with average payoff or an infinite repetition where the payoff is the normalized present value of the slightly discounted stage payoffs. A pure strategy of Player i in the repetition may be viewed as a function from strings of actions[1] of the other player in stages up to the present, to present actions of i. For a given positive integer ℓ, we say that such a strategy has *recall at most* ℓ if this function depends only on the last ℓ moves of the other player.

Consider now a *perturbation* of the repetition G^∞ by the strategies of recall $\leq \ell$. By this we mean a situation where each player in G^∞ is fairly convinced that he is facing a rational, utility-maximizing player on the other side; but he ascribes a small positive probability to the possibility that the player on the other side is an irrational automaton who plays according to some fixed strategy. In that case, it is assumed that the fixed strategy of the other player has recall $\leq \ell$; also, that there is considerable uncertainty about *which* fixed strategy it is: specifically, that the strategies assigned positive probability include at least all those of recall zero.[2]

Our main result says that *this perturbation of the repeated game possesses pure strategy equilibria, and all such equilibria are close (in payoff) to the unique cooperative outcome.* (By an *outcome* we mean a pair of payoffs.)

3. Discussion

It is worthwhile to recapitulate the conditions under which this result holds:

(i) The original game must be a two-person game with common interests.

(ii) The game must be repeated.

(iii) The repeated game must be perturbed.

(iv) The perturbation must consist *only* of strategies with uniformly bounded recall.

(v) All strategies of recall zero must occur with positive probability in the perturbation.

(vi) The equilibria must be pure.

Each of these conditions is essential. We discuss them one by one.

[1] Pure strategies in G are called *actions* (see Section 6).

[2] These strategies prescribe the same action at each period.

(i) It is conceivable that the result can be extended to more than two players, but before this is done, there are many obstacles to overcome, of both a conceptual and a technical nature. But the condition that the game have common interests is indispensable; if there are several efficient outcomes, there will in general be no pure strategy equilibria at all.

Intuitively, if the interests are not common, the players have a conflict of interest in agreeing on which efficient outcome to reach. In a game with common interests, there is no such conflict of interest; there is a *unique* point—the single efficient outcome—that is best for *all*. Nevertheless, even in those games, with no conflict of interest at all, the conditions that *necessarily* lead to cooperation are quite circumscribed. In a game without common interests, one might prefer to call an efficient point a "compromise." But in a game with common interests there is no issue of compromise, it presents the quest for cooperation in its purest form; and even there, doubts and mistrust and suspicion get in the way and make cooperation difficult to achieve. Indeed, this may happen even when the preferences are common—when the payoffs are identical at each square of the payoff matrix.

(ii) The example of Fig. 1 shows that without repetition there is no hope of getting the kind of result we are looking for. Repetition represents interacting, teaching, learning. Under the right conditions, people may perhaps *learn* to cooperate; but they cannot be expected to do so in a one-time, static situation.

(iii) In the Introduction, we discussed the fact that repetition by itself will not ensure cooperation either, even in a game with common payoffs. To achieve cooperation, a certain amount of irrationality must be built into the system, in the form of a perturbation.

Perturbations have played a very important role in game theory, starting with Selten's trembling hand perfect equilibria (1975).[3] In a sense, full rationality cannot feed on itself only; it must have a broader base. Perfection, however, does not lead to cooperation, as the "perfect folk theorem"[4] shows. One needs to perturb more selectively.

(iv) At first, it was conjectured that it might be sufficient to perturb with strategies that can be played by automata of bounded complexity. This, however, is not correct; bounded *recall* is essential. People must be willing to forget past grievances; remembering the distant past is not a good means for fostering cooperation. More accurately, in a culture in which irrational people have long memories, rational people are less likely to cooperate.

[3] See, for example, Myerson (1978), Kreps and Wilson (1982), Kalai and Samet (1984), and Kohlberg and Mertens (1986) for alternative formulations of the idea of perfection in equilibrium (what has come to be known as "refining" Nash equilibrium). All these definitions, which are playing an increasingly important role in the applications, are based on some kind of perturbation from pure rationality.

[4] As in Rubinstein (1979).

1,1	0,0
0,0	0,0

FIGURE 3

(v) The result is false unless the set of possible automata is sufficiently rich, in that it contains at least all the recall zero strategies. In particular, this is the case if each player's *only* solid information about the other's irrational type is that it cannot remember more than a specified time back; i.e., if subject to that caveat, nothing can be completely ruled out. This is to be contrasted with the assumptions of the "gang of four," which we discuss below.

(vi) Even when all the above conditions are met, and indeed the game G even has common payoffs, the result is still false when the equilibria are mixed rather than pure. Counterexamples are the rule rather than the exception; a game as simple as that of Fig. 3 is a counterexample. Mixed strategy equilibria imply that each player is uncertain about what the other will do; the uncertainty breeds mistrust and suspicion—which is, in the event, justified!—and stands in the way of cooperation.

On the other hand, perturbed repeated games with common interests always do possess pure strategy equilibria, indicating that these games have a certain innate stability, which may be an important factor in achieving cooperation.

4. HISTORICAL BACKGROUND

The first hint that bounded recall might have something to do with cooperation came in the summer of 1978. Aumann and Kurz, with the help of Jonathan Cave (see Aumann, 1981), worked out a version of the infinitely repeated Prisoner's Dilemma with memory one; this means that each player can base his action *only* on what his opponent did at the previous stage—he has "forgotten" everything else. This results in an 8 × 8 bimatrix game; iterated removal of weakly dominated strategies yields a unique strategy pair, in which both players start by playing "friendly" and continue with "tit-for-tat" thereafter. The outcome is cooperative, both players always playing "friendly."[5] The result was

[5] We prefer "friendly" to describe what is sometimes called the "cooperative" action in the one-shot Prisoner's Dilemma, and "greedy" for what is sometimes called "defect" or "double-cross." The term "cooperative" has other meanings in game theory; the fact that they are related—but not identical—to the one under discussion only makes matters worse. "Defect" and "double-cross" have the connotation of a person who has agreed to something and then reneges, which need not at all be the case in the Prisoner's Dilemma.

purely computational—there was no discernible theoretical explanation. However, domination arguments are often associated with trembling hand perfection; this suggested that perhaps some kind of perturbation is at work (see Section 3(iii) above).

At the same time that this bounded recall model—leading to the cooperative tit-for-tat outcome—was being investigated, a remarkable parallel development was taking place on a completely different front. Robert Axelrod (1984) conducted an experiment in which he asked scientists in various fields to write computer programs that would play the Prisoner's Dilemma. He then matched the programs against each other, and tit-for-tat turned out to be the most successful.

Suppose one had a reproducing population whose members were playing the Prisoner's Dilemma against each other, survival depended on the size of one's payoff, and different strategies were represented by different genes. Axelrod's experiment suggests that once introduced, the gene for tit-for-tat might be rather successful, gradually replacing other genes (i.e., strategies). This might explain the "evolution of cooperation"—which is the title of Axelrod's book.

Yet a theoretical model that would account for Axelrod's experimental result, and for the Aumann–Kurz–Cave computational observation, remained missing.

The first real step toward such a theoretical explanation was made by the "gang of four" (Kreps, Milgrom, Roberts, and Wilson, 1982). They showed that if one perturbs the finitely repeated Prisoner's Dilemma by assuming that with an arbitrarily small but positive exogenous probability, one of the players—say $P1$—is playing tit-for-tat rather than maximizing, then with a sufficiently long repetition, all sequential equilibrium outcomes are close to cooperative. The idea of the ingenious proof is that $P1$ (Player 1) may then *pretend* that he is actually in the perturbed mode that plays tit-for-tat in any case. Since there is a positive exogenous probability that in fact he is, $P2$ will gradually become convinced that this is the case. But then it is worthwhile for $P2$ to play "friendly" herself, and this leads to the cooperative outcome.[6]

[6] This oversimplifies their argument considerably. (Indeed, since it is asserted that "pretending" to be in the perturbed mode is an equilibrium strategy, it must be presumed that $P2$ knows this fact and therefore will *not* conclude that $P1$ is actually in the perturbed mode.) More accurately, the argument is by contradiction. If the equilibrium outcome is not near the friendly one, then $P1$'s "main" strategy (i.e., without the perturbation) cannot be close to tit-for-tat; because if it were, $P2$ would be motivated to play friendly herself, and this would lead to an outcome that is, after all, close to friendly. Since $P1$'s main strategy is not anything like tit-for-tat, $P2$ can easily distinguish between it and the perturbation; and against the perturbation, she will be motivated to play "friendly." But then $P1$ could, after all, pretend that he is in the perturbed mode, and this would elicit a friendly response from $P2$.

Important as it is, the Kreps–Milgrom–Roberts–Wilson result is conceptually not quite as strong as one might have liked. Rather than perturbing the game by a mixture of all possible strategies, as in the definition of trembling hand perfection, the gang of four perturbs it with tit-for-tat *only*. In a sense, therefore, tit-for-tat is the input to the theorem as well as the output. It is put in with small probability, and comes out with probability 1; gets sown as a tiny seed, and grows to a mature plant. While the result is very suggestive and important, one would have liked a stronger result, in which one is led to a cooperative outcome entirely endogenously. For example, this would be the case if the perturbation were a mixture of *all* alternative strategies, and one could then prove that the equilibrium strategy must be close to tit-for-tat; that tit-for-tat is, so to speak, endogenously picked out from all possible strategies, that it is *the* seed that sprouts from among all those that were sown.

Several years ago, the notion of a general finite automaton was introduced into the study of repeated games. Perhaps the most relevant here is the work of A. Neyman (1985), who investigated what happens when fully rational players are replaced by automata in finitely repeated games. It is well known that the folk theorem does not always apply in the case of finite repetitions; in the Prisoner's Dilemma, for example, the only equilibrium outcome in any finite repetition, no matter how long, is for both players always to play "greedy." Neyman showed that when one restricts the players to finite automata, even though they may be large compared with the number of repetitions (e.g., 1,000,000 states for 100 repetitions), there are equilibria with payoffs that are on average close to the friendly payoff. This means that automata *enable* cooperation, when it is impossible with full rationality; but it still does not achieve our aim of a model that *forces* the players into cooperation.

5. About the Proof

To prove that every perturbed repeated game with common interests has a pure strategy equilibrium, let us call a pair of actions *cooperative* if it leads to the unique cooperative outcome. The underlying idea is to choose some specific cooperative action pair and have the players always play that. This is all right if there is only one cooperative action pair. But if there is more than one, a player might be motivated in certain circumstances to use a deviant strategy, because it (the deviant strategy) might do as well against the "main" strategy of the other player, and better against the perturbation. The existence proof is directed at resolving this issue.

The proof of optimality of the equilibrium takes off from the basic idea

of the "gang of four" proof. If the equilibrium were not optimal, one of the players, say $P1$, could deviate from his equilibrium strategy and pretend to be an irrational automaton who always plays $P1$'s component s^1 of some cooperative action pair $s = (s^1, s^2)$. Because of the deviation, $P2$ would think that $P1$ is an irrational automaton; she would then seek to identify that automaton, in order to be able to maximize against it. Since $P1$ is pretending to be a "benevolent" automaton who always plays s^1, and s leads to the maximum possible payoff for each player, it follows that $P2$, in turn, would be motivated always to play s^2, and this would lead to the cooperative outcome. But this is a contradiction, and we conclude that the assumption of a nonoptimal equilibrium was untenable.

In this proof it is necessary to assume bounded recall for the following reason. For the proof to work, $P2$ must be willing to "explore," in order to identify which automaton $P1$ is. If the automata in the perturbation of $P1$ are not limited to being of bounded recall, the automaton that actually occurs might, for all $P2$ knows, be very vindictive; "exploration" by $P2$ might lead this automaton to "punish" $P2$ forever (see Section 11(ii)). Therefore $P2$ might not dare to explore, with the result that she is cowed into maintaining her suboptimal strategy forever.

The above gives only a rough idea of the basic issues in the proofs of existence and optimality.

6. The Formal Model

Let G be a two-person strategic game with finite pure strategy spaces S^1, S^2, and payoff function f. Set $S := S^1 \times S^2$. We will consider two kinds of *supergames: θ-discounted* supergames G^θ, and *k-stage* supergames G^k. Each play of each of these supergames consists of an infinite sequence of plays of G, called *stages*. After each stage, each player is informed of what the other did at the previous stage, and he remembers what he himself did and what he knew at previous stages. In the case of G^θ, the payoff is

$$(1 - \theta) \sum_{m=1}^{\infty} \theta^{m-1} f_m,$$

where f_m denotes the mth stage payoff. In the case of G^k, it is

$$\frac{1}{k} \sum_{m=1}^{k} f_m.$$

Intuitively, G^k has k stages only, and the payoff is the average; but formally, it is convenient that G^θ and G^k have the same strategic structure ("extensive game form") and differ in their payoff only.

In the sequel it will be convenient to have a uniform notation for all these supergames. Abusing our notation somewhat, we write G^α to denote either kind, where α stands for either k or $1/(1-\theta)$, as the case may be. We call α the *effective length* of the supergame G^α. It is the coefficient sum before normalization: the total payoff, or its present value, when all stage payoffs are 1 (*not* the per-period figure). Henceforth every statement about G^α refers to both the discounted and the finite-stage supergames.

Pure strategies in G are called *actions* or *moves*, to distinguish them from pure strategies in G^α. Mixed strategies in G are called *mixed actions*. If i is one of the players ($i = 1,2$), denote the other one by j (i.e., set $j := 3 - i$).

If σ^i is a pure strategy for i in G^α, then each finite sequence $(s^j_1, \ldots, s^j_n)$ of actions of the other player determines an action of i at stage $n+1$. Denote this action $\sigma^i(s^j_1, \ldots, s^j_n)$. We say that σ^i has *recall* $\leq \ell$ if

$$\sigma^i(s^j_1, \ldots, s^j_n) = \sigma^i(t^j_1, \ldots, t^j_m)$$

whenever $m, n \geq \ell$ and $(s^j_{n-\ell+1}, \ldots, s^j_n) = (t^j_{m-\ell+1}, \ldots, t^j_m)$—in words, if i's choice depends only on the last ℓ choices of the other player.

For each i, let μ^i be a mixed strategy in G^α whose support is included in the set $BR^i(\ell)$ of all pure strategies of recall $\leq \ell$ and includes $BR^i(0)$ (which consists of those strategies that prescribe the same action s^i at each stage, no matter what j does). That is, μ^i assigns probability 1 to $BR^i(\ell)$, and positive probability to each strategy in $BR^i(0)$. Let $\mu := (\mu^1, \mu^2)$.

Let $\varepsilon := (\varepsilon^1, \varepsilon^2)$ be a pair of positive real numbers that are ≤ 1. Define the (ε, μ)-*perturbation of* G^α, denoted $G^\alpha_{\varepsilon,\mu}$, as the following game: First, each player i chooses a pure strategy σ^i in G^α, unbeknownst to the other player. Then with probability $1 - \varepsilon^i$, player i actually plays σ^i in G^α, whereas with probability ε^i, nature chooses a pure strategy ξ^i in accordance with the distribution μ^i and forces i to play ξ^i in G^α. The other player, j, is never directly informed as to which of these alternatives occurred, though he may eventually be able to deduce this information from what happens during the course of play.

The (ε, μ)-perturbation is related to the "trembling hand" concept; intuitively, each player i wishes to play σ^i, but with probability ε^i, he will, unvolitionally, play μ^i instead.

Denote by $N(G^\alpha_{\varepsilon,\mu})$ the set of pure strategy Nash equilibrium outcomes of the perturbed supergame $G^\alpha_{\varepsilon,\mu}$. Call G a game with *common interests* if

there is an outcome (i.e., payoff pair) $z \in f(S)$ that strictly Pareto dominates all other outcomes, i.e., such that

$$z^1 > w^1 \qquad \text{and} \qquad z^2 > w^2 \tag{6.1}$$

for all other w in $f(S)$. Clearly, z is unique in such games.

We come now to our result.

MAIN THEOREM. *Let G be a game with common interests, and let z denote its unique Pareto optimal outcome. Then for each ℓ and μ,*

$$N(G^{\alpha}_{\varepsilon,\mu}) \to \{z\} \tag{6.2}$$

(in the Hausdorff metric[7]), as $\varepsilon \to 0$ and $\alpha \to \infty$.

The content of (6.2) is that for ε sufficiently small, and α sufficiently large, $G^{\alpha}_{\varepsilon,\mu}$ has an equilibrium payoff, and every equilibrium payoff is close to z. Thus the theorem has two components: existence and optimality. Existence says that every sufficiently small perturbation of a sufficiently long (or sufficiently slightly discounted) supergame has a pure strategy Nash equilibrium. Optimality says that every such equilibrium is nearly Pareto optimal. (As usual, $\varepsilon \to 0$ means $\varepsilon^1 \to 0$ and $\varepsilon^2 \to 0$.)

Note that there is no requirement for the maximum recall ℓ to go to ∞; it can be any fixed finite number, even 0.

7. TERMINOLOGY AND NOTATION FOR THE PROOFS AND EXAMPLES

A smiley (☺) indicates the end of a proof.

It is convenient to assume that the action spaces S^1 and S^2 are disjoint. This simplifies the notation because it enables one to identify a player i by his actions t^i.

Pairs $\sigma = (\sigma^1, \sigma^2)$ of strategies are, by a slight abuse of notation, sometimes denoted (σ^i, σ^j). Also, if σ^i and σ^j have been separately defined, σ denotes the pair whose components are σ^i and σ^j. Similar notations are used for action pairs, etc.

The universal quantifier is to be understood for variables that are not explicitly quantified. For example, the statement that σ^i is a best reply to

[7] The Hausdorff distance $d(A, B)$ between two sets A and B is the supremum of the numbers d such that for each a in A there is a point in B whose distance from a is at most d, and for each b in B there is a point in A whose distance from b is at most d. Thus (6.2) says that for each δ there are ε_0 and α_0 such that $d(N(G^{\alpha}_{\varepsilon}), \{z\}) \le \delta$ whenever $0 < \varepsilon^1 \le \varepsilon_0$, $0 < \varepsilon^2 \le \varepsilon_0$, and $\alpha \ge \alpha_0$.

σ^j, with i unquantified, means that each of σ^1 and σ^2 is a best reply to the other—in brief, that σ is an equilibrium.

Assume w.l.o.g. (without loss of generality) that all the payoffs are positive. Set

$$\rho = \min\{z^i - f^i(s): s \in S, f(s) \neq z, \quad i = 1,2\}; \tag{7.1}$$

ρ is a lower bound on the losses of any player if the payoff is not the unique Pareto optimal outcome z; it is the *least* that any player can suffer in that case.

Denote by a_m the coefficient of the mth stage payoff in the expression that defines the payoff to G^α; that is,

$$a_m = \begin{cases} \theta^{m-1}, & \text{if } G^\alpha \text{ is the } \theta\text{-discounted supergame;} \\ 1, & \text{if } G^\alpha \text{ is the } k\text{-stage supergame and } m \leq k; \\ 0, & \text{if } G^\alpha \text{ is the } k\text{-stage supergame and } m > k; \end{cases} \tag{7.21}$$

note that

$$\alpha = \sum_{m=1}^{\infty} a_m. \tag{7.22}$$

Denote by $f_m(\sigma)$ the outcome for the mth stage of G^α when the strategy pair σ is used, and set

$$\varphi(\sigma) = \frac{1}{\alpha} \sum_{m=1}^{\infty} a_m f_m(\sigma). \tag{7.23}$$

Thus $\varphi(\sigma)$ $(= (\varphi^1(\sigma), \varphi^2(\sigma))$ is the average (per stage) payoff when the strategy pair σ is used in the unperturbed supergame G^α.

The term "strategy" is henceforth reserved for pure strategies in G^α; when discussing mixed strategies, we say so explicitly. Denote by Σ^i the space of i's strategies in G^α; note that Σ^i, unlike φ^i, is independent of the coefficients a_m (i.e., whether G^α is finite stage or discounted, and what the value of α is). On Σ^i we impose the smallest (coarsest, with fewest open sets) topology such that for each finite action sequence $(s^j_1, \ldots, s^j_n)$ of j, the function $\sigma^i \to \sigma^i(s^j_1, \ldots, s^j_n)$ from Σ^i to S^i is continuous, where S^i is endowed with the discrete topology (all sets open). In this topology, Σ^i is compact, and φ is continuous on $\Sigma^1 \times \Sigma^2$. Note that Σ^i and its topology are independent of α and of whether G^α is discounted or finite stage. Note also that the players' strategy spaces in $G^\alpha_{\varepsilon,\mu}$ are the same as in G^α.

If μ is a mixed strategy pair, define $\varphi(\mu)$ as the expectation of $\varphi(\xi)$ when the strategy pair ξ is distributed according to μ.

When no confusion can result, we sometimes abbreviate $G^{\alpha}_{\varepsilon,\mu}$ by Γ. The payoff in Γ is denoted Φ $(= (\Phi^1, \Phi^2))$ and is given by

$$\Phi(\sigma) = \varphi((1 - \varepsilon^1)\sigma^1 + \varepsilon^1\mu^1, (1 - \varepsilon^2)\sigma^2 + \varepsilon^2\mu^2). \tag{7.31}$$

When we wish to be more explicit, we use $\Phi^{\alpha}_{\varepsilon,\mu}$, $\Phi^{\alpha}_{\varepsilon}$, or Φ_{ε} for Φ.

Now define

$$\varphi^i_{\varepsilon}(\tau) := \varphi^i(\tau^i, (1 - \varepsilon^j)\tau^j + \varepsilon^j\mu^j). \tag{7.32}$$

In words, $\varphi^i_{\varepsilon}(\tau)$ is the expected payoff to i when τ is played in Γ, given that i is "rational" (actually plays τ^i, rather than one of the strategies comprising the perturbation μ^i). It follows that

$$\begin{aligned} &\sigma \text{ is an equilibrium of } \Gamma \text{ if and only if} \\ &\varphi^i_{\varepsilon}(\xi^i, \sigma^j) \leq \varphi^i_{\varepsilon}(\sigma) \text{ for all strategies } \xi^i \text{ of } i \end{aligned} \tag{7.33}$$

(i.e., when φ_{ε} is substituted for Φ_{ε} as the payoff function in Γ). Formally, (7.33) follows from noting that the expected payoff to i when τ is played in Γ is given by

$$\Phi^i(\tau) = (1 - \varepsilon^i)\varphi^i_{\varepsilon}(\tau) + q^j,$$

where q^j is independent of τ^i.

For $m = 0, 1, 2, \ldots, \infty$, define a *history* h_m as a sequence of m action pairs. Call h_m *infinite* or *finite* according as $m = \infty$ or $m < \infty$. Intuitively, a history describes the sequence of actions used in a specific play of the supergame. If $h_m = (s_1, s_2, \ldots)$, write $h^i_m := (s^i_1, s^i_2, \ldots)$; call it i's *part* of h_m. If $n \leq m$, write $h_n := (s_1, s_2, \ldots, s_n)$, and call h_m an *extension* of h_n (in any one context, use of the notations h_m and h_n for finite histories of different lengths will mean that they coincide on their overlap).

A strategy pair σ is said to *generate* (or *induce*) a history $h_m = (s_1, s_2, \ldots)$, if $\sigma^i(h^j_n) = s^i_{n+1}$ whenever $0 \leq n < m$. This terminology is used also for the perturbed game Γ; that is, when we say that σ "generates" a certain history, we take into account only the history that is generated when both players play their "main" strategies σ^i, ignoring the perturbations μ^i.

If σ is a pair of strategies and h_m a finite history, then $\sigma(h_m)$ denotes $(\sigma^1(h^2_m), \sigma^2(h^1_m))$. Thus $\sigma(h_m)$ is the action pair played at stage $m + 1$ if the players are playing according to σ and the history up to stage m was h_m.

Call a strategy σ^i in Σ^i *history-independent* if $\sigma^i(h^j_n)$ is independent of h^j_n for all n. Thus, a history-independent strategy calls for i to play the same sequence $s^i_1, s^i_2, \ldots$, no matter what j does. Note that

every pure strategy σ^j has a history-independent best reply. (7.4)

Indeed, that σ^j has *some* best reply σ^i follows from the compactness of the strategy space and the continuity of the payoff function; together, σ^i and σ^j generate a sequence $s^i_1, s^i_2, \ldots$ of actions on the part of i, which define a history-independent strategy $\hat{\sigma}^i$ that yields the same history, and therefore the same payoff, as σ^i. Therefore, also $\hat{\sigma}^i$ is a best reply to σ^j. ☺

Next, suppose the finite history $h = h_n = (s_1, s_2, \ldots, s_n)$ has occurred. We need some notation for what then happens starting with stage $n + 1$—how the remaining supergame appears, and how specific pure and mixed strategies appear when viewed as applying to this remainder only.

Consider first the remainder $G^\alpha|h$ of the supergame G^α. Though the history h may well affect how this remainder is played, the remainder itself technically depends at most on n: for the k-stage supergame G^k it is G^{k-n}, and for the θ-discounted supergame G^θ it is simply G^θ itself. Thus $G^\alpha|h$ is defined as $G^\alpha|h$, where $\alpha|h = \alpha|h_n$ is defined as $k - n$ for the k-stage supergame and as α for the discounted supergame.

Suppose next that σ^i is a pure strategy for i in G^α. We say that σ^i is *compatible*[8] with h if there exists a strategy σ^j for j such that (σ^i, σ^j) generates h. If σ^i is compatible with h, define the strategy $\sigma^i|h$ of i in $G^\alpha|h$ by $(\sigma^i|h)(t^j_1, \ldots, t^j_p) := \sigma^i(s^j_1, s^j_2, \ldots, s^j_n, t^j_1, \ldots, t^j_p)$; in words, if i plays σ in G^α, and the history h takes place, then $\sigma^i|h$ is the induced strategy in the remaining game. If σ^i is not compatible with h, then $\sigma^i|h$ is not defined.

We come now to mixed strategies ν^i in G^α, restricting ourselves to those with denumerable support. Write $\nu^i(h)$ for the total probability that ν^i assigns to pure strategies σ^i that are compatible with h; in words, $\nu^i(h)$ is the probability under ν^i that h will occur, if j plays his part of it.[9] Call ν^i *compatible* with h if $\nu^i(h) > 0$; in that case, write $\nu^i|h$ for the conditional probability distribution of $\sigma^i|h$ when σ^i is distributed according to ν^i, given that σ^i is compatible with h. In words, if at the beginning of the supergame G^α, player j thinks that i's pure strategy is distributed according to ν^i, and if the first m stages resulted in the history h, then for the

[8] The reverse terminology—that the history is compatible with the strategy—will be used to mean the same thing.

[9] Note that if ν^i happens to be pure, then $\nu^i(h) = 1$ if ν^i is compatible with h, and $= 0$ if not.

remaining supergame $G^{\alpha|h}$, player j will think that i's pure strategy is distributed according to $\nu^i|h$. If $\nu^i(h) = 0$, then $\nu^i|h$ is not defined.

Note that as far as the definition of the perturbed supergame $G^{\alpha}_{\varepsilon,\mu}$ is concerned, there is nothing to prevent either or both of the ε^i from vanishing; we required this in Section 6 only because the optimality part of the main theorem is false without it. As we shall see in the next section, either or both of the ε^i may indeed vanish in the existence theorem. To prevent confusion, we will henceforth call the vector $\varepsilon = (\varepsilon^1, \varepsilon^2)$ *positive* if both its components are positive, *nonnegative* if both its components are nonnegative; if nothing is said, positivity is to be understood. In either case, set $\|\varepsilon\| := \max\{\varepsilon^1, \varepsilon^2\}$. Note that if $\varepsilon^i = 0$, then $G^{\alpha}_{\varepsilon,\mu}$ is independent of μ^i.

Now define

$$G^{\alpha}_{\varepsilon,\mu}|h := G^{\alpha|h}_{\varepsilon\mu(h),\mu|h}, \tag{7.5}$$

where $\varepsilon\mu(h) := (\varepsilon^1\mu^1(h), \varepsilon^2\mu^2(h))$ and $\mu|h := (\mu^1|h, \mu^2|h)$. Intuitively, $\Gamma|h := G^{\alpha}_{\varepsilon,\mu}|h$ is the remainder of the perturbed supergame Γ, after the history h has taken place. By the last sentence of the previous paragraph, this definition remains meaningful even if one or both of the $\mu^i(h)$ vanish.

We end this section with a discussion of what happens when a player using a strategy σ^i in Γ changes his mind if a certain history $h = h_n$ happens, and continues with a different strategy. For convenience, use primes ($'$) for objects related to the game $\Gamma|h$ remaining after h has occurred; thus $\Gamma' := \Gamma|h$, $\nu' := \nu|h$, $\alpha' := \alpha|h$, stage m of Γ corresponds to stage $m' := m - n$ of Γ', the coefficient of the payoff in stage m' of Γ' is denoted $a'_{m'}$ (and given by

$$a_{n+1}a'_{m'} = a_m), \tag{7.51}$$

and φ' is the payoff function in the remaining unperturbed supergame $G^{\alpha'}$.

Suppose now that η'^i is a strategy in Γ'. Define a strategy $\sigma^i \lhd \eta'^i$ as follows: $\sigma^i \lhd \eta'^i$ coincides with σ^i unless the history h has occurred.[10] If h occurs, define $\sigma^i \lhd \eta'^i$ from then on to coincide with η'^i in Γ'; that is, if h_m is an extension of h, define h'_{m-n} by $h_m = (h, h'_{m-n})$, and $(\sigma^i \lhd \eta'^i)(h^j_m) := \eta'^i(h'^j_{m-n})$. Then[11]

$$\begin{aligned}\alpha\varphi(\sigma^i \lhd \eta'^i, \mu^j) - \alpha\varphi(\sigma^i, \mu^j) \\ = \sigma^i(h)\mu^j(h)a_{n+1}(\alpha'\varphi'(\eta'^i, \mu'^j) - \alpha'\varphi'(\sigma'^i, \mu'^j)),\end{aligned} \tag{7.6}$$

Indeed, since $\sigma^i \lhd \eta'^i$ differs from σ^i only when h has occurred, the

[10] That is $(\sigma^i \lhd \eta'^i)(g^j_m) = \sigma^i(g^j_m)$ when g^j_m is not an extension of h^j_n.

[11] Regarding the meaning of $\sigma^i(h)$, see footnote 9.

difference between the total expected payoffs is the same as the difference after h has occurred (in which case $\sigma^i \lhd \eta'^i$ coincides with η'^i), multiplied by the probability that h occurs.

8. The Existence Proof

In this section we prove the existence component of the main theorem, in the following formulation:

Existence Theorem. *Let G be a game with common interests. Then for each ℓ, there is a positive number ε_0, such that for all ε with $0 \leq \|\varepsilon\| \leq \varepsilon_0$, all α, and all μ, each of the perturbed supergames $G^{\alpha}_{\varepsilon,\mu}$ has a pure strategy equilibrium.*

This is stronger than needed for the main theorem, in four (related) respects: First, ε may be nonnegative here, whereas in the main theorem it is restricted to be positive. Second, there is no requirement here that the support of μ^i contain any particular set (though it still must be contained in $BR^i(\ell)$). Third, ε_0 is independent of μ in the current formulation, whereas in the main theorem, it implicitly depends on μ. (In the optimality part of the main theorem, proved in Section 9, strong use will be made of the dependence of ε_0 on μ.) Fourth, the current formulation makes no requirement on the effective length α, whereas in the main theorem, the existence is only asserted for all sufficiently large α.

The basic idea of the proof is to pick an action pair s_* yielding the unique cooperative outcome z and have each player i play s^i_* at each stage, no matter what the other did previously. If the players were rational with probability 1, nothing more would have to be said, as neither player can do better than at z. The difficulty is that they are not; the simplistic strategies just described ignore the fact that with a small but positive probability, the players are irrational automata. Two implications of this are as follows:

(i) If i observes that j has deviated from s_*, then he should conclude that j is in fact an irrational automaton, and should therefore maximize against his estimate of what this automaton might be doing, rather than continuing to play s^i_*.

(ii) Even if i does not observe any deviations on the part of j, he himself may wish to deviate. For example, this could happen if s^i_* is not the *only* best reply to s^j_*—if the unique cooperative payoff pair z occurs more than once in the prescribed row or column of G. In that case, the possibility arises that though the prescribed strategy for i is a best reply to the prescribed strategy of j, it need not be a best reply to i's conception of what j is really doing, because it ignores the small but positive probability

that j is an irrational automation. Such an automaton might conform with the prescribed strategy at all stages previous to some given stage, and at that given stage might deviate. If s_*^i is the *only* best reply to s_*^j, then for sufficiently small ε^j, it is also a best reply to $(1-\varepsilon^j)s_*^j + \varepsilon^j x^j$, where x^j is the mixed action of j induced by μ^j; the possibility that j is an irrational automaton creates only a second-order effect, which i may in this case safely ignore. But if there are other best replies to s_*^j, then s_*^i need no longer be a best reply to $(1-\varepsilon^j)s_*^j + \varepsilon^j x^j$. Thus i can no longer ignore the second-order effect; he must consider what j would do if he were an irrational automaton, and this might dictate something other than s_*^i. This is especially true in view of the fact that the prescribed strategy for j does not provide for any reprisals in the case of deviations by i, so that i may maximize on a stage-by-stage basis, without worrying about the future.

These difficulties necessitate a somewhat roundabout proof, which we now outline. Let φ_ε^i denote i's expected payoff in the perturbed supergame $\Gamma = G_{\varepsilon,\mu}^{\alpha}$ when he plays rationally, but is not sure whether j is playing rationally or is an irrational automaton (see (7.32)). Let σ be a strategy pair in Γ that maximizes $\varphi_\varepsilon^1(\sigma) + \varphi_\varepsilon^2(\sigma)$; we claim that σ is then an equilibrium of Γ. If not, then there is a strategy, say τ^1 of $P1$, that does better in Γ against σ^2 than σ^1 does. This implies that it yields a higher value for the function φ_ε^1 (which differs from $P1$'s payoff in Γ only by a term that does not depend on $P1$'s "main" strategy; see (7.33)). Since σ maximizes $\varphi_\varepsilon^1 + \varphi_\varepsilon^2$, it follows that (τ^1, σ^2) must yield a lower value than σ for φ_ε^2. Now φ_ε^2 is composed of two terms, reflecting $P2$'s payoff against a rational and an irrational $P1$, respectively. The second term is not changed when σ^1 is replaced by τ^1, so it cannot account for φ_ε^2 getting smaller thereby. So it must be the first term that gets smaller. Hence the stream of outcomes that (τ^1, σ^2) yields cannot consist of z's only, since these are the best possible outcomes. So there must be a stage at which the "main" strategies τ^1 and σ^2 (as distinguished from the perturbations μ^i) yield less than z to both players.

Starting with the first such stage, which we dub $n+1$, suppose that both players switch from (τ^1, σ^2) to another strategy pair η' that does yield a constant stream of z's (when both players are rational); there is no difficulty in finding such a pair. If either player i deviates, the other player j responds by assuming that i is an irrational automaton of recall $\leq \ell$, and embarks on a systematic exploration to identify precisely *which* automaton. During the exploration, i may be playing in a suboptimal manner; but after a fixed finite number of stages, he will indeed have identified precisely which automaton j is. From then on, he can play to optimize against the automaton that he now *knows* j to be. Therefore his "losses"—the amount by which his total[12] payoff is less than what it might have been had

[12] *Not* per-stage average!

he known all along how j is playing—do not exceed some fixed constant C; that is, their order of magnitude is at most that of the payoff to one play (Lemma 8.2).

Denote by ζ the overall result of this maneuver (replacing the continuation of (τ^1, σ^2) after n by η'), and suppose the players play ζ. If both are rational, then 1 will make a considerable one-time gain vis-a-vis (τ^1, σ^2), since, at least once, he will be getting the maximal payoff z^1 rather than a smaller amount. If $P2$ is irrational, then $P1$ may suffer a loss by going to ζ^1; but this loss is bounded by C, and since the probability of irrationality is small, there will be a net gain for $P1$ vis-a-vis (τ^1, σ^2), whose order of magnitude is that of the payoff to one stage.[13] By assumption, (τ^1, σ^2) yields $P1$ more than σ, so we conclude that $\varphi_\varepsilon^1(\zeta) > \varphi_\varepsilon^1(\sigma)$.

As for $P2$, since ζ yields a constant stream of z's, it must, when both players are rational, lead to a payoff for $P2$ that is $\geq$ that yielded by σ. When $P1$ is irrational, then ζ^2 can lead to a smaller total payoff for $P2$ than σ^2; but it cannot be smaller by more than C. Since the probability of irrationality is small, this effect will be smaller than the difference between the total payoffs $\alpha\varphi_\varepsilon^1(\zeta)$ and $\alpha\varphi_\varepsilon^1(\sigma)$, whose order of magnitude is at least that of the payoff to one stage, as we just saw. Thus $\varphi_\varepsilon^1(\zeta) + \varphi_\varepsilon^2(\zeta)$ exceeds $\varphi_\varepsilon^1(\sigma) + \varphi_\varepsilon^2(\sigma)$, contrary to σ maximizing $\varphi_\varepsilon^1 + \varphi_\varepsilon^2$.

So much for the outline[14]; we proceed now to the formal proof.

For each strategy ξ^j of j, define $\gamma^i(\xi^j)$ as the payoff to i of a best reply to ξ^j; in symbols,

$$\gamma^i(\xi^j) := \max\{\varphi^i(\xi^i, \xi^j): \xi^i \in \Sigma^i\}. \tag{8.1}$$

As usual, $\gamma^i(\mu^j)$ is defined as the expectation of $\gamma^i(\xi^j)$ when ξ^j is distributed according to μ^j.

LEMMA 8.2. *For each game G and length ℓ of recall, there is a positive number C, such that for each supergame G^α of G, there is a strategy pair η with $\varphi(\eta) = z$ and*

$$\varphi^i(\eta^i, \mu^j) \geq \gamma^i(\mu^j) - C/\alpha. \tag{8.21}$$

[13] There is a subtlety here that should be noted. In the discounted case, $P1$'s net gain from the switch must be discounted, since it occurs only at stage $n + 1$, rather than at stage 1. Formally, its order of magnitude is a_n rather than 1. However, the losses are also $O(a_n)$, since they, too, start only at stage $n + 1$; so since they are multiplied by ε, they are more than offset by the gain. Thus the phrase "payoff to one stage" in the text means "payoff to stage n."

[14] This proof uses Lemma 8.2 on the remainder Γ' of Γ that starts at stage $n + 1$. In the finite-stage case, there is a simpler proof, which uses Lemma 8.2 directly on all of Γ. In the discounted case the simpler proof does not work, since if n is large, the discounted gain may fail to offset the cost (even when multiplied by ε) of an earlier exploration.

Remark. The positive number C depends only on the game G and the length ℓ of recall; it is independent of the parameters of G^α (whether it is discounted or finite stage, and the numerical value of α) and of the perturbation μ. The strategy pair η does depend on the parameters of G^α; but it, too, is independent of μ.

Proof. The lemma implies that the players can do almost as well in Γ as if each were informed before the beginning of play whether the other is rational or a bounded recall automaton, and if the latter, precisely which one. Explicitly, η^i yields the best possible result z^i against j's "main" strategy η^j, and against the perturbation μ^j, it comes within the order of magnitude of the payoff to one stage (or equivalently, a finite number of stages) of the optimum $\gamma^i(\mu^j)$.

Let s_* be an action pair with $f(s_*) = z$. The idea of the proof is that if j sticks to s_*^j, then i can obtain the maximum possible payoff z^i simply by playing s_*^i. If j deviates from s_*^j, the bounded recall enables i to play so that for all practical purposes, he finds out within a bounded number of stages exactly which automaton he is facing, and then maximizes against that automaton.[15]

To make this precise, we distinguish between the "transient" and the "steady-state" behavior of a strategy for j with recall bounded by ℓ. By the *steady-state behavior* of such a strategy we mean its behavior starting with stage $\ell + 1$. This is determined if we know which one-move response the strategy prescribes for each sequence of ℓ moves that i can make. The steady-state behavior of such a strategy may therefore be considered a function from ℓ-move sequences of i to single moves of j. It will be convenient to use the term *steady-state strategy* for the steady-state behavior of a pure bounded recall strategy; it is defined only for fixed ℓ. Thus if i knows only that j is playing some strategy in $BR^j(\ell)$, and if he observes j's actions as he (that is, i) goes through *all* ℓ-move sequences one by one, then he can identify j's steady-state strategy exactly. There are altogether $|S^i|^\ell$ such sequences, each of length ℓ; thus j's last response to the last of these sequences occurs at stage $M^i := \ell|S^i|^\ell + 1$, which implies that i can identify j's steady-state strategy in a period that does not exceed[16] M^i stages.

Now let η^i be the following strategy: Play s_*^i for the first $\ell + 1$ stages. If j plays s_*^j at all these stages, play s_*^i forever. If j deviates from s_*^j in one of them, then starting at stage $\ell + 2$, play all ℓ-move sequences, one by one;

[15] Bounded recall is required for the maximization as well as for the exploration. Even if all automata have a bounded number of states, and i finds out exactly which such automaton he is facing, he may be unable to maximize if the automaton is not of bounded recall. That is because the exploration itself may have put the automaton into a permanently vindictive frame of mind that no action of i can change.

[16] If he takes account of overlapping sequences, he can reduce this period considerably.

this takes $M^i - 1$ stages. At stage $\ell + 1 + M^i$, do something arbitrary (this is needed for i to observe j's response to i's previous actions). The play of j during these M^i stages is either (i) consistent with a unique steady-state strategy ω^j or (ii) consistent with none. In case (i), let $\hat{\xi}^j$ be one of the pure strategies in $BR^j(\ell)$ whose steady-state behavior is ω^j. Suppose j deviated from s_*^j at stage n. Let $\hat{\xi}^i$ be a pure history-independent best reply to $\hat{\xi}^j$ in G^α (see (7.4)), prescribing the actions $(t_1^i, t_2^i, \ldots)$. Define η^i to prescribe $(t_1^i, t_2^i, \ldots)$ for stages $\ell + 1 + M^i + 1, \ell + 1 + M^i + 2, \ldots$, no matter what j does. In case (ii), pick an arbitrary strategy for the remainder of the game, and play in accordance with it.[17]

In words, if j does not deviate during the first $\ell + 1$ stages, then i knows either that j is not an automaton, or that he is, and so will always respond with s_*^j to an ℓ-string of s_*^i's (since he responded in this way at stage $\ell + 1$ and has recall bounded by ℓ). If j *does* deviate, η^i assumes that j is an automaton, and identifies the steady-state behavior of that automaton. It then ignores what the automaton is actually doing, and responds in a way that would have been optimal if the supergame were only now beginning and j were only now starting to play in accordance with some automaton that has the identified steady-state behavior.

That $\varphi(\eta) = z$ follows from its construction.

To prove (8.21), suppose that j uses a strategy ξ^j distributed according to μ^j. If i knew ξ^j, he could play an optimal response ξ^i immediately; the result would be $\gamma^i(\xi^j)$, which is the first term on the right of (8.21). But in fact, i starts responding to ξ^j only after $\ell + 1 + M^i$ stages; and even then his response $\hat{\xi}^i$ is not necessarily optimal against ξ^j, but it is optimal against a strategy $\hat{\xi}^j$ whose steady-state behavior is the same as that of ξ^j. We now show that these departures from optimality against ξ^j do not affect the total payoff by more than a constant C, and therefore the average payoff by more than C/α, which is the second term on the right of (8.21).

Assume w.l.o.g. that ξ^i is history-independent (see (7.4)). Since ξ^j and $\hat{\xi}^j$ have the same steady-state behavior, we deduce

$$f_m(\xi) = f_m(\xi^i, \xi^j) = f_m(\xi^i, \hat{\xi}^j), \qquad \text{whenever } m > \ell. \tag{8.22}$$

Similarly, since $\hat{\xi}^i$ is by definition history-independent,

$$f_m(\hat{\xi}^i, \xi^j) = f_m(\hat{\xi}^i, \hat{\xi}^j) = f_m(\hat{\xi}) \qquad \text{whenever } m > \ell. \tag{8.23}$$

[17] When j plays η^j or μ^j, as we assume here, then (ii) is impossible. But a formal definition of η^i must cover all eventualities, even those that are impossible when it is played against a particular strategy (such as $(1 - \varepsilon^j)\eta^j + \varepsilon^j\mu^j$).

Since $\hat{\xi}^i$ is a best reply to $\hat{\xi}^j$ (and so at least as good as ξ^i), (7.22), $0 \le f^i_m \le z^i$, $0 \le a_m \le 1$ (see (7.21)), and (8.22) yield

$$\alpha\varphi^i(\hat{\xi}) \ge \alpha\varphi^i(\xi^i, \hat{\xi}^j) = \sum_{m=1}^{\infty} a_m f^i_m(\xi^i, \hat{\xi}^j) \ge \sum_{m=\ell+1}^{\infty} a_m f^i_m(\xi^i, \hat{\xi}^j)$$

$$= \sum_{m=\ell+1}^{\infty} a_m f^i_m(\xi) \ge \sum_{m=1}^{\infty} a_m f^i_m(\xi) - \sum_{m=1}^{\ell} a_m z^i \ge \varphi^i(\xi) - \ell z^i. \quad (8.24)$$

Set $n := n^i := \ell + 1 + M^i + \ell + 1$ and $C := \max\{(\ell + n^1)z^1, (\ell + n^2)z^2\}$. Then (7.22), $0 \le a_m \le 1$, $0 \le f^i_m \le z^i$, the definition of η^i, (8.23), the monotonicity of a_m in m, and (8.24) yield

$$\alpha\varphi^i(\eta^i, \xi^j) = \sum_{m=1}^{\infty} a_m f^i_m(\eta^i, \xi^j) \ge \sum_{m=n}^{\infty} = \sum_{m=n}^{\infty} a_m f^i_{m-n+\ell+1}(\hat{\xi}^i, \xi^j)$$

$$\ge \sum_{m=n}^{\infty} a_{m-n+\ell+1} f^i_{m-n+\ell+1}(\hat{\xi}^i, \hat{\xi}^j) - \sum_{m=n}^{\infty} (a_{m-n+\ell+1} - a_m) z^i$$

$$= \sum_{m=\ell+1}^{\infty} a_m f^i_m(\hat{\xi}) - \sum_{m=\ell+1}^{n-1} a_m z^i$$

$$\ge \sum_{m=1}^{\infty} a_m f^i_m(\hat{\xi}) - \sum_{m=1}^{\ell} a_m z^i - \sum_{m=\ell+1}^{n-1} a_m z^i$$

$$\ge \alpha\varphi^i(\hat{\xi}) - (n-1)z^i \ge \alpha\varphi^i(\xi) - \ell z^i - n^i z^i$$

$$\ge \alpha\varphi^i(\xi^i, \xi^j) - C = \alpha\gamma^i(\xi^j) - C.$$

Since ξ^j is distributed according to μ^j, (8.21) follows. ☺

Proceeding with the existence proof, define

$$\psi(\tau) := \psi_\varepsilon(\tau) := \varphi^1_\varepsilon(\tau) + \varphi^2_\varepsilon(\tau), \quad (8.31)$$

where φ_ε is as in (7.32). Since the strategy spaces Σ^i are compact, so is $\Sigma^1 \times \Sigma^2$. Since the payoff function φ is continuous on $\Sigma^1 \times \Sigma^2$, so are φ_ε and ψ, and therefore the maximum of ψ over $\Sigma^1 \times \Sigma^2$ is attained; let σ $(= \sigma_\varepsilon)$ be a strategy pair that attains this maximum. Choose ε_0 so that

$$\varepsilon_0 < \rho/(\rho + 2C), \quad (8.32)$$

where ρ is as in (7.1) and C is as in Lemma 8.2, and let $\varepsilon \leq \varepsilon_0$. We will show that then

$$\sigma \text{ is an equilibrium of } \Gamma. \tag{8.4}$$

Suppose not. Then by (7.33) we may assume that $P1$, say, has a strategy τ^1 with

$$\varphi_\varepsilon^1(\tau^1, \sigma^2) > \varphi_\varepsilon^1(\sigma). \tag{8.5}$$

If $\varphi(\tau^1, \sigma^2) = z$, then

$$\begin{aligned}\varphi_\varepsilon^2(\tau^1, \sigma^2) &= \varphi^2((1 - \varepsilon^1)\tau^1 + \varepsilon^1\mu^1, \sigma^2) = (1 - \varepsilon^1)z^2 + \varepsilon^1\varphi^2(\mu^1, \sigma^2) \\ &\geq (1 - \varepsilon^1)\varphi^2(\sigma) + \varepsilon^1\varphi^2(\mu^1, \sigma^2) = \varphi^2((1 - \varepsilon^1)\sigma^1 + \varepsilon^1\mu^1, \sigma^2) \\ &= \varphi_\varepsilon^2(\sigma);\end{aligned}$$

hence by (8.5) and (8.31), $\psi(\tau^1, \sigma^2) > \psi(\sigma)$, contrary to σ maximizing ψ. Hence $\varphi(\tau^1, \sigma^2) \neq z$, so there must be a stage m with $a_m > 0$ at which (τ^1, σ^2) does not yield z. Let $n + 1$ be the first such stage, $h := h_n$ the history generated by (τ^1, σ^2) during the first n stages, and $\Gamma' := \Gamma|h$ (see 7.5)). Since (τ^1, σ^2) does not yield z at stage $n + 1$ of Γ, it follows that (τ'^1, σ'^2) does not yield z at stage 1 of Γ'. Using (7.1) and that the coefficient of the first stage payoff is 1, this yields

$$\alpha'\varphi'^i(\tau'^1, \sigma'^2) \leq \alpha' z^i - \rho. \tag{8.61}$$

Then Lemma 8.2, applied to Γ' (rather than Γ), yields a strategy pair η' with $\varphi'(\eta') = z$ and

$$\varphi'^i(\eta'^i, \mu'^j) \geq \gamma'^i(\mu'^j) - C/\alpha', \tag{8.62}$$

where γ' is defined as in (8.1), with φ' instead of φ. Now define $\zeta^1 := \tau^1 \lhd \eta'^1$ and $\zeta^2 := \sigma^2 \lhd \eta'^2$, where $\lhd$ is as in (7.6). Since ζ differs from (τ^1, σ^2) only after stage n, (7.51) yields

$$\alpha\varphi^1(\zeta) - \alpha\varphi^1(\tau^1, \sigma^2) = a_{n+1}(\alpha'\varphi'^1(\eta') - \alpha'\varphi'^1(\tau'^1, \sigma'^2)) \geq a_{n+1}\rho. \tag{8.71}$$

Since ζ generates z at each stage, which yields the largest possible payoff for both players, we have

$$\alpha\varphi^2(\zeta) = \alpha z^2 \geq \alpha\varphi^2(\sigma). \tag{8.72}$$

Using (7.6), (8.62), $0 \le \sigma^2(h)\mu^1(h) \le 1$, and (8.1), we get

$$\begin{aligned}
&\alpha\varphi^2(\mu^1, \zeta^2) - \alpha\varphi^2(\mu^1, \sigma^2) \\
&\quad = \sigma^2(h)\mu^1(h)a_{n+1}(\alpha'\varphi'^2(\mu'^1, \eta'^2) - \alpha'\varphi'^2(\mu'^1, \sigma'^2)) \\
&\quad \ge \sigma^2(h)\mu^1(h)a_{n+1}(\alpha'\gamma'^2(\mu'^1) - C - \alpha'\gamma'^2(\mu'^1)) \ge -a_{n+1}C. \qquad (8.81)
\end{aligned}$$

Similarly, using $i = 1$ in (7.6), and substituting τ^1 for σ^i,

$$\alpha\varphi^1(\zeta^1, \mu^2) - \alpha\varphi^1(\tau^1, \mu^2) \ge -a_{n+1}C. \qquad (8.82)$$

Now (8.5), (7.32), (8.71), and (8.82) yield

$$\begin{aligned}
\alpha\varphi^1_\varepsilon(\zeta) - \alpha\varphi^1_\varepsilon(\sigma) &> \alpha\varphi^1_\varepsilon(\zeta) - \alpha\varphi^1_\varepsilon(\tau^1, \sigma^2) \\
&= (1 - \varepsilon^2)(\alpha\varphi^1(\zeta) - \alpha\varphi^1(\tau^1, \sigma^2)) \\
&\quad + \varepsilon^2(\alpha\varphi^1(\zeta^1, \mu^2) - \alpha\varphi^1(\tau^1, \mu^2)) \\
&\ge (1 - \varepsilon^2)a_{n+1}\rho - \varepsilon^2 a_{n+1}C \ge a_{n+1}(\rho - \|\varepsilon\|(\rho + C)). \qquad (8.91)
\end{aligned}$$

Moreover, (7.32), (8.72), and (8.81) yield

$$\begin{aligned}
\alpha\varphi^2_\varepsilon(\zeta) - \alpha\varphi^2_\varepsilon(\sigma) &= (1 - \varepsilon^1)(\alpha\varphi^2(\zeta) - \alpha\varphi^2(\sigma)) \\
&\quad + \varepsilon^1(\alpha\varphi^2(\mu^1, \zeta^2) - \alpha\varphi^1(\mu^1, \sigma^2)) \\
&\ge 0 - \varepsilon^1 a_{n+1}C \ge -a_{n+1}\|\varepsilon\|C. \qquad (8.92)
\end{aligned}$$

Combining (8.31), (8.91), (8.92), $\|\varepsilon\| \le \varepsilon_0$, and (8.32), we obtain

$$\begin{aligned}
\psi(\zeta) - \psi(\sigma) &= (\varphi^1_\varepsilon(\zeta) - \varphi^1_\varepsilon(\sigma)) + (\varphi^2_\varepsilon(\zeta) - \varphi^2_\varepsilon(\sigma)) \\
&> (a_{n+1}/\alpha)(\rho - \|\varepsilon\|(\rho + 2C)) > 0,
\end{aligned}$$

which contradicts the definition of σ as a maximizer of ψ. Thus (8.4) is proved, and with it the existence theorem. ☺☺

9. The Optimality Proof

In this section we prove the optimality component of the main theorem, in the following formulation:

OPTIMALITY THEOREM. *Keep μ fixed, let $\{G_\varepsilon^\alpha\} := \{G_{\varepsilon,\mu}^\alpha\}$ be a sequence*[18] *of perturbed supergames with $\varepsilon \to (0,0)$ and $\alpha \to \infty$, and let τ_ε^α be an equilibrium of G_ε^α. Then $\Phi_\varepsilon^\alpha(\tau_\varepsilon^\alpha) \to z$.*

An equivalent formulation of this is as follows: For each $\delta > 0$ there are ε_0 and α_0 such that if $0 < \|\varepsilon\| \leq \varepsilon_0$, $\alpha \geq \alpha_0$, and τ is an equilibrium of G_ε^α, then $|\Phi_\varepsilon^\alpha(\tau) - z| \leq \delta$. Together with the existence theorem proved in the previous section, this proves the main theorem. Henceforth in this section, we assume given a sequence as in the statement of the optimality theorem; statements involving limits or symbols like $\to$ or $O(1)$ refer to this sequence.

W.l.o.g. we assume that $\Phi_\varepsilon^\alpha(\tau_\varepsilon^\alpha)$ converges. Indeed, the theorem as stated follows from its truth for all convergent subsequences.

Let s_* be an action pair in G whose payoff is z, and let ξ_*^i be the history-independent strategy of i that prescribes s_*^i at each stage. Intuitively, we wish to show that when ε is small and α large, then $\Phi_\varepsilon^\alpha(\tau_\varepsilon^\alpha)$ is close to z. This is clear if $\tau := \tau_\varepsilon^\alpha$ yields s_* at all stages. If not, there is a first stage (say n) at which τ dictates something other than s_* for one of the players (say $P2$). Suppose now that $P2$ plays s_*^2 at each stage—that is, plays ξ_*^2 rather than τ^2. Then $P1$ will, immediately after stage n, deduce that $P2$ is a bounded recall automaton. One of his options is then to explore which automaton she ($P2$) is. Lemma 8.2 implies that he can find this out precisely at a cost that is "bounded" (on the order of magnitude of the payoff to one stage; see (9.2)[19]). Since τ is an equilibrium, τ^1 must yield a result at least as good as that of any option he has. Therefore, once $P1$ has concluded that $P2$ is an automaton, τ^1 must yield him an expected payoff "close" to (i.e., within a bounded amount of) what he could get if he knew *which* automaton she is (see 9.3)). This implies that against any automaton of $P2$ to which μ^2 ascribes positive probability, τ^1 must yield an actual (not just expected!) payoff close to the payoff that a best possible response yields. In particular, that is the case for ξ_*^2, which is assigned positive probability by μ^2, since it is in $BR^2(0)$. A best possible response to ξ_*^2 yields z^1 (on average) to $P1$, so we conclude that against ξ_*^2, the strategy τ^1 must yield $P1$ a payoff close to z^1 (see (9.4)). That is the case starting with stage $n+1$; up to stage $n+1$, the strategy τ^1 yields s_* against ξ_*^2, so $P1$ again[20] gets z^1. All in all, therefore, if (τ^1, ξ_*^2) is played, $P1$ must get close to z^1. But since z strictly Pareto dominates all other outcomes, it follows that $P2$, too, must get close to z^2 (see (9.6)). There-

[18] That is, a sequence $G_{\varepsilon_1,\mu}^{\alpha_1}, G_{\varepsilon_2,\mu}^{\alpha_2}, \ldots$ with $\varepsilon_i \to (0,0)$ and $\alpha_i \to \infty$.

[19] Footnote 13 applies here as well.

[20] Near the end of section 11(i) we discuss a simplification based on strategies that deviate at stage 1 rather than n.

fore also in the perturbed game, ξ_*^2 must bring $P2$ close to z^2 against τ^1. But since τ is an equilibrium of the perturbed game, it follows that τ^2 is at least as good as ξ_*^2 against τ^1, so τ must bring $P2$ close to z^2. Again using the strict Pareto domination of z, we conclude that τ also gets $P1$ close to z^1.

Proceeding to the formal proof,[21] if, for all but finitely many G_ε^α, the strategy pair τ_ε^α generates a constant stream of action pairs s_*, then it follows immediately that $\Phi_\varepsilon^\alpha(\tau_\varepsilon^\alpha) \to z$. So assume that there is a subsequence of the G_ε^α—w.l.o.g.[22] the whole sequence—for each member of which there is a stage at which $\tau := \tau_\varepsilon^\alpha$ generates an action pair other than s_*; let $n := n_\varepsilon^\alpha$ be the first such stage. Let h_{n-1} be the $(n-1)$-stage history consisting only of s_*'s. By the definition of n, either $\tau^1(h_{n-1}^2) \neq s_*^1$ or

$$\tau^2(h_{n-1}^1) \neq s_*^2; \tag{9.1}$$

w.l.o.g., the latter. Let $h := h_n$ denote the n-stage history generated by (τ^1, ξ_*^2), and let $\Gamma' := \Gamma|h$ (see (7.5)). Applying Lemma 8.2 to Γ' yields a strategy η'^1 of $P1$ such that

$$\varphi'^1(\eta'^1, \mu'^2) \geq \gamma'^1(\mu'^2) - C/\alpha'. \tag{9.2}$$

We now assert that

$$\varphi'^1(\tau'^1, \mu'^2) \geq \gamma'^1(\mu'^2) - C/\alpha'. \tag{9.3}$$

Indeed, define $\zeta^1 := \tau^1 \lhd \eta'^1$; thus ζ^1 coincides with τ^1 unless h has occurred, and if it has, ζ^1 coincides with η'^1. If both players play τ, then h does not occur, so

$$\varphi^1(\zeta^1, \tau^2) = \varphi^1(\tau). \tag{9.31}$$

Since τ is an equilibrium of Γ,

$$\alpha\varphi_\varepsilon^1(\tau) \geq \alpha\varphi_\varepsilon^1(\zeta^1, \tau^2) \tag{9.32}$$

(see (7.33)), and so (7.32), (9.31), and (7.6) yield

$$\begin{aligned} 0 &\geq \alpha\varphi^1(\zeta^1, \mu^2) - \alpha\varphi^1(\tau^1, \mu^2) \\ &= \tau^1(h)\mu^2(h)a_{n+1}\alpha'(\varphi'(\eta'^1, \mu'^2) - \varphi'(\tau'^1, \mu'^2)). \end{aligned} \tag{9.33}$$

[21] Remarks similar to those in footnote 14 apply here.

[22] Since $\Phi_\varepsilon^\alpha(\tau_\varepsilon^\alpha)$ converges, it is enough to prove that some subsequence has limit z.

Since h is the n-stage history generated by (τ^1, ξ^2_*), it is compatible with both τ^1 and ξ^2_*, so $\tau^1(h) > 0$ and $\mu^2(h) \geq \mu^2(\xi^2_*) > 0$. Using $a_{n+1} > 0$ (whence also $\alpha' > 0$), (9.33) yields $\varphi'(\tau'^1, \mu'^2) \geq \varphi'(\eta'^1, \mu'^2)$, so (9.3) follows from (9.2). ☺

Denote by $\mu^2(\xi^2)$ the probability that μ^2 assigns to ξ^2. The definition (8.1) of γ^1 yields $\varphi'^1(\tau'^1, \mu'^2) \leq \gamma'^1(\mu'^2)$, and together with (9.3), this yields

$$C/\alpha' \geq \gamma'^1(\mu'^2) - \varphi'^1(\tau'^1, \mu'^2) = \Sigma\mu'^2(\xi^2)(\gamma'^1(\xi^2) - \varphi'^1(\tau'^1, \xi^2)) \geq 0,$$

where the summation extends over all ξ^2 in $BR^2(\ell)$. The definition (8.1) of γ^1 implies that $\gamma'^1(\xi^2) \geq \varphi'^1(\tau'^1, \xi^2)$ for all ξ^2, so each term in the summation is nonnegative, so the summation is $\geq$ each of its terms, in particular that corresponding to ξ^2_*. Hence

$$\begin{aligned} C/\alpha' &\geq \mu'^2(\xi^2_*)(\gamma'^1(\xi^2_*) - \varphi'^1(\tau'^1, \xi^2_*)) \\ &\geq \mu^2(\xi^2_*)(z^1 - \varphi'^1(\tau'^1, \xi^2_*)) \geq 0. \end{aligned} \tag{9.4}$$

The coefficient $\mu^2(\xi^2_*)$, unlike $\mu'^2(\xi^2_*)$, depends only on the fixed, constant μ, not on α or on ε; and because $\xi^2_* \in BR^2(0)$, it is positive. Hence (9.4) and $a_{n+1} \leq 1$ yield[23]

$$a_{n+1}\alpha'\varphi'^1(\tau'^1, \xi^2_*) = a_{n+1}\alpha' z^1 + O(1). \tag{9.5}$$

By the definition of h, the strategy pair (τ'^1, ξ^2_*) generates z up to stage $n - 1$. Hence noting that by (7.51) and $m' = m - n$,

$$\sum_{m=n+1}^{\infty} a_m = \sum_{m'=1}^{\infty} a_m = \sum_{m'=1}^{\infty} a_{n+1}a'_{m'} = a_{n+1}\alpha',$$

we obtain from (9.5) that

$$\begin{aligned} \varphi^1(\tau^1, \xi^2_*) = \frac{1}{\alpha}\Bigg(\Bigg(\sum_{m=1}^{n-1} a_m\Bigg) z^1 + f_n(\tau^1, \xi^2_*) \\ + \Bigg(\sum_{m=n+1}^{\infty} a_m\Bigg)\varphi'^1(\tau'^1, \xi^2_*)\Bigg) \to z^1, \end{aligned}$$

since the nth stage payoff $f_n(\tau^1, \xi^2_*)$ remains bounded. Hence

$$\varphi(\tau^1, \xi^2_*) \to z, \tag{9.6}$$

[23] α and ε are implicit in the expression $\varphi'^1(\tau'^1, \xi^2_0)$, as $\tau = \tau^\alpha_\varepsilon$.

for otherwise $\lim \varphi(\tau^1, \xi_*^2)$ is a "feasible payoff"—a point in the convex hull of the pure outcomes to G—whose first coordinate is z^1 but whose second coordinate is $< z^2$, which contradicts (6.1). Since τ is an equilibrium of Γ, (9.6) and $\varepsilon \to 0$ yield

$$z^2 \geq \Phi^2(\tau) \geq \Phi^2(\tau^1, \xi_*^2) \geq (1 - \varepsilon^1)(1 - \varepsilon^2)\varphi^2(\tau^1, \xi_*^2) \to z^2;$$

thus $\Phi^2(\tau) \to z^2$. Hence (as in the proof of (9.6)), $\Phi^1(\tau) \to z^1$, since $\lim \Phi(\tau)$ is a feasible payoff. ☺ ☺

10. Varying or Different Discount Factors

Both the existence and the optimality theorems—and therefore also the main theorem—continue to hold when the discount factor θ is allowed to vary from stage to stage, as long as it never exceeds 1 or falls below 0. That is, the coefficients a_m of the mth stage payoffs in (7.23) may be any nonincreasing sequence of nonnegative numbers, normalized so that $a_1 = 1$. All three theorems remain literally true, word for word, as they stand. Thus ε_0 in the existence theorem is a fixed constant that works for all such sequences of coefficients; and in the optimality theorem, $\{G_{\varepsilon,\mu}^{\alpha}\}$ may be any sequence of perturbed supergames, as long as $\varepsilon \to (0,0)$ and $\alpha \to \infty$, where α is the effective length of the supergame (defined by (7.22)). The proofs, too, are literally unchanged, word for word.

One may also ask what happens when the discount factors are different for the two players, whether or not they vary from stage to stage. In that case the optimality theorem appears to go through without difficulty. The existence theorem is more delicate; but it, too, appears to go through, though with a longer proof, and perhaps with some modification. We have, however, not checked out these matters in detail and hope to treat them in a subsequent paper.

11. Counterexamples, Conjectures, Further Discussion

(i) *Noncommon interests.* The existence theorem fails for games G in which the interests are not common; that is, the perturbed repetitions Γ of such games may fail to have pure strategy equilibria. This can happen in several ways. For example, if G is "matching pennies" (Fig. 4), then for each pure strategy τ^2 of $P2$ (the column player) there is a pure strategy τ^1 of $P1$ that yields him 1 (and so -1 to $P2$) in the unperturbed repeated game G^{α}, and similarly for each τ^1 there is a τ^2 with $\varphi(\tau) = (-1,1)$. This

1,−1	−1,1
−1,1	1,−1

FIGURE 4

implies that G^α has no equilibrium, and since the payoff in Γ is close to that in G^α, there is no equilibrium in Γ either.

This is clearly due to G not having an equilibrium. But even when G does have an equilibrium, Γ may not. Indeed, if G is the battle of the sexes (Fig. 5), $\ell \geq 1$, and μ^i assigns positive probability to each strategy with recall ≤ 1 (and as always, probability 1 to all of $BR^i(\ell)$), then for *all* sufficiently large α and small ε, the perturbed repeated game $G^\alpha_\varepsilon := G^\alpha_{\varepsilon,\mu}$ has no pure strategy equilibrium.

The proof is based on the idea behind the optimality theorem; we content ourselves with an outline. Briefly, each player, by pretending to be an automaton that in the steady-state always plays the same action (top for $P1$ and right for $P2$), can "force" the other to the equilibrium that is more favorable to him, and this is a contradiction.

More precisely, if the assertion is false, there exists a sequence of perturbed supergames G^α_ε with $\varepsilon \to (0,0)$ and $\alpha \to \infty$, each of which has a pure strategy equilibrium $\tau := \tau^\alpha_\varepsilon$. Statements involving convergence, limits, or symbols like $\to$ or $O(1)$ refer to this sequence, as do the phrases "all" or "almost all" (all but finitely many) G^α_ε. W.l.o.g., the payoffs $\Phi(\tau)$ converge, where $\Phi := \Phi^\alpha_\varepsilon$; denote the limit by y. Since $\varepsilon \to (0,0)$, we have $\lim \varphi(\tau) = \lim \Phi(\tau) = y$; hence $y^1 + y^2 \leq 3$, so w.l.o.g. $y^2 < 2$.

Let t^2_1 be the action prescribed by τ^2 for the first stage, and let ξ^2_* be the history-independent strategy of $P2$ prescribing the other action (that is, not t^2_1) for the first stage, and thereafter "right," no matter what $P1$ has previously done. W.l.o.g. $\Phi(\tau^1, \xi^2_*)$ converges; denote the limit by x. Then $x^2 \leq y^2 < 2$; otherwise, τ would not be an equilibrium, since $P2$ could gain by switching from τ^2 to ξ^2_*. From $\varepsilon \to (0,0)$ it follows that $\lim \varphi(\tau^1, \xi^2_*) = \lim \Phi(\tau^2, \xi^2_*) = x$. The definition of ξ^2_* implies that $x = \lim \varphi(\tau^1, \xi^2_*)$ is a convex combination of (2,1) and (0,0). Since $x^2 < 2$, it follows that $x^1 < 1$.

Suppose now that $P2$ plays ξ^2_* rather than τ^2. Since ξ^2_* differs from τ^1 already at the first stage, $P1$ may conclude immediately after the first stage that she ($P2$) is a bounded recall automaton. One of his options is

2,1	0,0
0,0	1,2

FIGURE 5

then to explore which automaton she is. Lemma 8.2 implies that he can find this out precisely at a cost that is bounded (in units of total payoff). Since τ is an equilibrium, τ^1 must yield a result at least as good as that of any option that $P1$ has. Therefore, once $P1$ has concluded that $P2$ is an automaton, τ^1 must yield him an expected payoff "close" to (i.e., within $O(1)$ of) what he could get if he knew *which* automaton she is. This implies that against any automaton of $P2$ to which μ^1 ascribes positive probability, τ^1 must yield an actual (not expected!) payoff close to the payoff that a best possible response yields. In particular, this is true of ξ^2_*, which has recall 1, and so is assigned positive probability by μ^2. Thus the total payoff $\alpha\varphi^1(\tau^1, \xi^2_*) = \alpha\gamma^1(\xi^2_*) + O(1) = \alpha \cdot 1 + O(1)$, since $\gamma^1(\xi^2_*) = 1 + O(1/\alpha)$ by the definition of ξ^2_*. Hence $x^1 = \lim \varphi^1(\tau^1, \xi^2_*) = 1$, contrary to the conclusion of $x^1 < 1$ reached above. ☺

The reader will have noticed that this proof is based on a direct application of Lemma 8.2 to all of Γ, rather than to the remainder Γ' starting with stage $n + 1$. This was made possible by assuming that all strategies with recall bounded by 1 have positive probability, which enabled the use of a strategy that deviates from the main strategy already at the first stage, and thereafter continues with any desired action (in this case, the one that is more favorable to $P2$). Another way of achieving the same purpose is to use a 4×4 version of the battle of the sexes, with each row and column occurring in two identical copies. Similar simplifications would work in the proof of the optimality theorem; but in a theorem, as opposed to an example, it is desirable to avoid unnecessary assumptions. We mention for the record that for the 2×2 version of the battle of the sexes, Γ has no equilibrium even when $\ell = 0$, but the proof is more complicated.

Though the above proof depends in several places on specific features of the battle of the sexes, nevertheless the underlying idea is quite general. Thus it appears that for games without common interests, the existence theorem *never* holds. More precisely, we venture the following conjecture[24]: Suppose that there is no outcome z in G that weakly Pareto dominates all other outcomes y (i.e., for all z there is an outcome y and a player i such that $y^i > z^i$). Then there are ℓ and ℓ' with $\ell' \leq \ell$ such that if $BR^i(\ell') \subset \text{Support } \mu^i \subset BR^i(\ell)$, then for all sufficiently large α and small ε, the perturbed repeated game $G^\alpha_\varepsilon := G^\alpha_{\varepsilon,\mu}$ has no pure strategy equilibrium.

The intermediate case, in which there is an outcome that Pareto dominates all other outcomes weakly but not strongly, is covered neither by the existence theorem nor by this conjecture.

(ii) *Nonbounded recall.* The optimality theorem fails if the perturbations do not consist of bounded recall automata only; that is, the pure strategy equilibria of Γ may then be far from optimal (but see (vi) below).

[24] With some trepidation, as we have not examined it carefully.

3,3	1,1	0,0
1,1	2,2	0,0
0,0	0,0	0,0

FIGURE 6

For example, let G be the game of Fig. 6, and let ξ_1^* be the following strategy of $P1$ (the row player): At the first stage, play Top. Thereafter, if $P2$ has ever played Left, play Bottom; otherwise, play Top. Let ξ_*^2 be the symmetrically defined strategy of $P2$. Suppose μ^i assigns positive probability to each strategy in $BR^i(0) \cup \{\xi_*^i\}$, and only to those. Consider now the pair σ of history-independent strategies that calls for both $P1$ and $P2$ always to play Middle. $P1$ might consider deviating to the history-independent strategy ξ_T^1 in $BR^1(0)$ that prescribes always playing Top, hoping that $P2$ will respond with Left from stage 2 on. But if μ^1 assigns a sufficiently high probability to ξ_*^1, then $P2$ will be afraid to respond in this way, for fear that $P1$ is actually playing ξ_*^1, in which case she ($P2$) will end up with 0. Therefore $P1$ will not even attempt ξ_T^1, since it will decrease his payoff from 2 to 1. Therefore σ, which yields the nonoptimal payoff (2,2), is an equilibrium.

Note that the strategies ξ_*^i, while not of bounded recall, may be considered finite-state automata (see, e.g., Neyman [1985] or Rubinstein [1986]).

As for existence, while the proof of the existence theorem depends strongly on the bounded recall, we do not have a counterexample to it when the perturbations consist, say, of finite-state automata with a bounded number of states.

(iii) *Mixed strategies.* The optimality theorem fails for mixed strategies; that is, there are Γ satisfying all the requirements of Section 6, with mixed strategy equilibria whose payoffs are far from optimal. For example, if G is the game of Fig. 3, and μ^i assigns probability $\frac{1}{2}$ to each strategy in $BR^i(0)$, then for all large even k and small (scalar) ε, the perturbed k-stage repetition $\Gamma := G^k_{(\varepsilon,\varepsilon),\mu}$ has a mixed strategy equilibrium with payoff close to $\frac{3}{4}$, defined as follows:

At stage 1, $P1$ picks T (Top) and B (Bottom) with probabilities $\frac{1}{2} - \delta$ and $\frac{1}{2} + \delta$, respectively, where δ is a small number to be specified later; similarly, $P2$ plays $(\frac{1}{2} - \delta)L + (\frac{1}{2} + \delta)R$ (with the usual notation), using the same δ.

If TL or BR was played at stage 1, then forever afterward, $P1$ plays T, and $P2$ plays L, whether or not the other deviates.

If TR was played at stage 1, then on the equilibrium path, $P1$ keeps playing T forever; whereas $P2$ plays R up to stage $k/2$, and after that, L

k	$k/2$
$k/2$	$k-1$

FIGURE 7

until the end. $P1$ punishes deviations of $P2$ by playing B until the end; but $P2$ ignores deviations of $P1$.

If BL was played at stage 1, the description is symmetric to that for TR. This completes the description of the equilibrium.

Let $\varepsilon' := \frac{1}{2}\varepsilon/((1-\varepsilon)(\frac{1}{2}-\delta)+\frac{1}{2}\varepsilon)$ be the conditional probability that $P1$ is an automaton, given T at stage 1. Note that if δ and ε are small, then so is ε'. After TR at stage 1, the equilibrium strategy yields $k/2$ to the rational type of $P2$; whereas by taking her chances on $P1$ being an automaton who always plays T—i.e., by deviating to L—she would get an expectation of only $(1-\varepsilon')1+\varepsilon' k$, which for small ε' and large k, is much smaller.

Figure 7 gives the total payoff when both players are rational and play the equilibrium strategies, the rows and columns representing the stage 1 choices. If $P2$ takes into account the possibility that $P1$ may be an automaton, the expected payoffs to her rational type are given in Fig. 8, where the rows and columns represent the stage 1 choices of the players' *rational* types. The number δ must be specified so that she is indifferent between L and R, i.e.,

$$(\tfrac{1}{2}-\delta)a_{TL}+(\tfrac{1}{2}+\delta)a_{BL}=(\tfrac{1}{2}-\delta)a_{TR}+(\tfrac{1}{2}+\delta)a_{BR},$$

where a_{TL}, etc., represent the entries in Fig. 8. For $P1$, we get the same equation. Solving for δ, we find $\delta \to 0+$ as $\varepsilon \to 0$ and $k \to \infty$. Thus total payoff is $\approx(\frac{3}{4})k$, so average payoff $\to\frac{3}{4}$. ☺

The reasoning is very robust; it works also for odd k, for discounted games, general perturbations $(\varepsilon^1, \varepsilon^2)$, general μ^i (with support $BR^i(0)$), and so on. Much the same construction shows that there are equilibrium payoffs close to (0,0), and indeed that the set of mixed strategy equilibrium payoffs of $G^{\alpha}_{\varepsilon,\mu}$ converges (in the sense of Hausdorff) to the entire interval from (0,0) to (1,1) (where ε is now again a vector $(\varepsilon^1, \varepsilon^2)$).

$(1-\frac{1}{2}\varepsilon)k$	$(1-\frac{1}{2}\varepsilon)k/2$
$(1-\frac{1}{2}\varepsilon)k/2+\frac{1}{2}\varepsilon k$	$(1-\frac{1}{2}\varepsilon)(k-1)+\varepsilon k/4$

FIGURE 8

(iv) *Recalling one's own actions.* The definition of bounded recall refers to each player's recall of the *other* player's actions, not of his own. That is because a strategy of i is defined as a function from the past actions of j only—not of i's own past actions—to i's current actions; i's past actions are determined by i's strategy and by j's past actions, and so would be redundant as separate independent variables.

But sometimes—like for the definition of perfect equilibrium—it is convenient that a strategy allow explicitly for deviations from what that strategy itself prescribed at previous stages. In that case, the strategy must prescribe a player's current actions as a function of the entire past history, including his own actions. Adopting this viewpoint, we say that a strategy has recall $\leq \ell$ *in the wide sense* if it calls for the same action after two histories that coincide on the last ℓ stages.

This is substantively different from the definition used in this paper. Indeed, a wide sense bounded recall strategy is essentially the same as a finite-state automaton, since a player can use his own actions to store information he wishes to remember. Holding a grudge is easy with such a strategy; with ordinary bounded recall, it is impossible. Thus the situation for wide sense bounded recall is like at (ii) above: the optimality theorem fails; the existence theorem probably fails, but we do not have a counterexample.

(v) *Three or more players.* This is similar to wide sense bounded recall, because any two players can, in effect, use each other's actions as a memory device to "hold a grudge" against a third. Specifically, the example at (ii) can be modified to yield a counterexample here as well. We omit details.

(vi) *Impurities in the perturbations.* It seems likely that for the proofs to work, the perturbations μ^i need not assign probability 1 to bounded recall automata, but only a sufficiently high probability. But we have not checked this out, and in particular the required probability might depend on the effective length α of Γ.

12. Related Recent Literature

That the "gang of four" result (see Section 4) cannot be viewed as a truly endogenous derivation of cooperation is underscored by the work of Fudenberg and Maskin (1986), who showed that by perturbing with another strategy rather than tit-for-tat, one can support any feasible individually rational outcome in a finitely repeated game.[25]

As far as we know, the first to find such a derivation—i.e., conditions that *necessarily* lead to Pareto optimality in a noncooperative frame-

[25] Though not necessarily as the unique equilibrium outcome.

work—were Kalai and Samet (1985). Their result deals with *unanimity games*,[26] defined as games in which the action space is the same for all players, and the payoffs vanish unless all choose the same action (i.e., choose a *diagonal* action n-tuple). The game is played (the authors say "attempted") k times, with the payoff to the k-attempt game defined as that of the first attempt at which agreement is reached, i.e., a diagonal element is chosen; if this never happens, the payoff is zero to all. The authors show that if k is large enough,[27] then in the k-stage game, all persistent equilibria (Kalai and Samet, 1984) satisfying a certain natural symmetry condition are Pareto optimal.

More recently, Ben Porath and Dekel (1987p) have considered games of *mutual* interest, defined as games with a unique Pareto optimal action pair.[28] They showed that under certain conditions, if a player may "burn" utility (i.e., unilaterally lower his payoff), then only the Pareto optimal outcome survives iterated deletion of weakly dominated strategies.[29] Note that this result involves only one stage, so there is no learning, even implicitly. The burning rights must be asymmetric; either only one player may burn (in two-person games) or they must do their burning in a specified order (nonsimultaneous).

Both the foregoing results, unlike ours, apply (or can be extended) to n players, not just two.

Yet another approach, due to Matsui (1989), uses the idea of "information leakage." In Matsui's model, before play starts, each player chooses a strategy for the infinitely repeated game; then with a small probability, i's strategy is revealed to j (without i finding this out); j may then change his strategy at a small cost. If the leakage is one-sided—only $P2$, say, may discover $P1$'s strategy, but $P1$ cannot discover $P2$'s—and if the one-shot game has "strictly individually rational"[30] payoffs, then subgame perfect pure equilibria exist in the repeated game, and all such equilibria have Pareto optimal payoffs. If the leakage is two-sided, an optimality theorem is proved, but not an existence theorem.

Among recent results, perhaps the most closely related to ours is that of Fudenberg and Levine (1989). They consider a repeated two-person game in which one of the two players is replaced by a succession of "tempo-

[26] Often called *games of coordination* in the literature.

[27] For example, larger than the number of actions.

[28] As distinguished from games with common interests, in which there is a unique Pareto optimal *payoff*, which may result from different action n-tuples. In games of mutual interest, the unique Pareto optimal action n-tuple provides a sort of focal point in the sense of Schelling (1960).

[29] In particular, all stable equilibria (Kohlberg and Mertens, 1986) are Pareto optimal.

[30] Yielding each player more than his minmax over pure actions. In a game with common interests, the unique Pareto optimal payoff is always strictly individually rational (unless it is the only payoff).

rary" players with identical characteristics, each of whom observes the outcome of all previous plays, but herself plays—and is paid—in one stage only. The "permanent" player is perturbed by a class of automata that is arbitrary except that it must assign positive probability to each constant strategy[31]; for example, that is so if all finite automata, or all Turing machines, are considered possible. The conclusion is that in equilibrium, the permanent player must get the outcome that is best possible for him, subject to each temporary player getting at least her maxmin. Bounded recall is irrelevant in this context because the temporary player does not have to worry about being "punished" for a deviation in the future, since she has no future.

A feature tying the perturbation literature together is that the players tend to mimic the perturbation. As a result, the perturbation "takes over," the main strategies become indistinguishable from it; the rational take on the protective coloring of the irrational. The distinguishing feature of the later literature of this genre[32] is an endogenous mechanism that selects "cooperative" strategies from among a relatively large and amorphous class of "unbiased" strategies comprising the perturbation.

Finally, we mention Gilboa and Samet (1987p). In their work, which does not use perturbations, $P1$ (the "weak" player) is restricted to using a certain kind of automaton (e.g., bounded recall automata), while $P2$ is unrestricted. When the game is repeated infinitely often, $P2$ has strategies that weakly dominate all others; if she uses such a strategy, and $P1$ maximizes against it, then the outcome is best possible for the *weak* player. (That is, he gets the outcome that is best possible for him, subject to $P2$ getting at least her maxmin; note the similarity of this conclusion to Fudenberg and Levine's.) As in the work of Ben Porath and Dekel, Fudenberg and Levine, and Matsui, the asymmetry between the players plays an important role here.

13. Conclusion

The work on equilibrium refinements since Selten's "trembling hand" (1975) indicates that rationality in games depends critically on *irrationality*. In one way or another, all refinements work by assuming that irrationality cannot be ruled out, that the players ascribe irrationality to each

[31] History-independent strategy that always prescribes the same action.

[32] Fudenberg and Levine (1989), Matsui (1989), and this paper, as distinguished from Kreps, Milgrom, Roberts, and Wilson (1982), and Fudenberg and Maskin (1986). Although Matsui does not use perturbations explicitly, the opportunity that one of the players may have to change his strategy may perhaps be viewed as a kind of endogenous perturbation.

other with a small probability. True rationality needs a "noisy," irrational environment; it cannot grow in sterile soil, cannot feed on itself only.

In most of the previous literature on refinements, the issue is the manner in which rational agents process the irrationality in the environment. In contrast, here we examine the effect of the *composition of the environment;* we look at the soil, not the plant.

Our conclusion is that when the environment is forgetful—when people, in general, do not bear grudges—then there is ground to hope that rational agents will cooperate.

And since the equilibrium strategies mimic the perturbation, the effect reinforces itself: The more people play like bounded recall automata, the more probability a rational agent will attach to the perturbation having bounded recall, so the more likely *he* will be to play like this.[33] The rational agent comes to resemble his irrational environment. We are what we eat; the plant becomes what it drinks from the soil, and then enriches the soil with more of the same.

ACKNOWLEDGMENTS

We gratefully acknowledge a seminal conversation with Mordecai Kurz that took place at Stanford in the summer of 1975, at which he suggested that there might be a connection between cooperation and bounded recall. But the nature of this connection remained elusive, despite repeated attempts during the ensuing years to pin it down. The idea that eventually led to the research reported here was conceived at the conference on repeated games that was held at Jerusalem in June of 1985. Conversations with S. Hart, A. Neyman, and J.-F. Mertens are gratefully acknowledged.

REFERENCES[34]

AUMANN, R. (1981). "Survey of Repeated Games," in *Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern,* pp. 11–42. Mannheim, Wien, Zurich: Wissenschaftsverlag, Bibliographisches Institut.

AXELROD, R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

BEN PORATH, E., AND DEKEL, E. (1987p). "Coordination and the Potential for Self-Sacrifice," RP 984. Stanford University Graduate School of Business (revised April 1988).

FUDENBERG, D., AND LEVINE, D. (1989). "Reputation and Equilibrium Selection in Games with a Single Patient Player," *Econometrica* **57,** in press.

FUDENBERG, D., AND MASKIN, E. (1986). "The Folk Theorem in Repeated Games with Discounting and with Incomplete Information," *Econometrica* **54,** 533–554.

---

[33] See (vi) of Section 11.

[34] p stands for "preprint."

GILBOA, Y., AND SAMET, D. (1987p). "Bounded vs. Unbounded Rationality: The Strength of Weakness," WP 16. Foerder Institute for Economic Research, Tel Aviv University, July.

KALAI, E., AND SAMET, D. (1984). "Persistent Equilibria in Strategic Games," *Int. J. Game Theory* **13,** 129–144.

KALAI, E., AND SAMET, D. (1985). "Unanimity Games and Pareto Optimality," *Int. J. Game Theory* **14,** 41–50.

KOHLBERG, E., AND MERTENS, J. F. (1986). "On the Strategic Stability of Equilibria," *Econometrica* **54,** 1003–1037.

KREPS, D., MILGROM, P., ROBERTS, J., AND WILSON, R. (1982). "Rational Cooperation in the Finitely Repeated Prisoner's Dilemma," *J. Econ. Theory* **27,** 245–252.

KREPS, D., AND WILSON, R. (1982). "Sequential Equilibria," *Econometrica* **50,** 863–894.

MATSUI, A. (1989). "Information Leakage Forces Cooperation," *Games Econ. Behav.* **1,** 94–115.

MYERSON, R. B. (1978). "Refinement of the Nash Equilibrium Concept," *Int. J. Game Theory* **7,** 73–80.

NEYMAN, A. (1985). "Bounded Complexity Justifies Cooperation in the Finitely Repeated Prisoner's Dilemma," *Econ. Lett.* **19,** 227–230.

RUBINSTEIN, A. (1979). "Equilibrium in Supergames with the Overtaking Criterion," *J. Econ. Theory* **21,** 1–9.

RUBINSTEIN, A. (1986). "Finite Automata Play the Repeated Prisoner's Dilemma," *J. Econ. Theory* **39,** 83–96.

SCHELLING, T. C. (1960). *The Strategy of Conflict.* Cambridge: Harvard Univ. Press.

SELTEN, R. (1975). "Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games," *Int. J. Game Theory* **4,** 25–55.