THE ROBUSTNESS OF INCOMPLETE PENAL CODES IN REPEATED INTERACTIONS

OLIVIER GOSSNER

PRELIMINARY AND INCOMPLETE

Abstract. We study the robustness of equilibria with regards to small payoff perturbations of the dynamic game. We find that complete penal codes, that specify player’s strategies after every history, can have at best limited robustness and may even fail to exist for some games. We define incomplete penal codes as partial descriptions of equilibrium strategies. We construct incomplete penal codes that are robust to perturbations of the game dynamics. These codes generate a Folk Theorem in a class of stochastic games with incomplete information.

Keywords: Repeated Games, incomplete information, stochastic games, Folk Theorem, robust equilibrium.

1. Introduction

The theory of repeated games explains how long-term interactions may induce cooperation. The celebrated Folk Theorem (Aumann and Shapley, 1994; Rubinstein, 1979; Fudenberg and Maskin, 1986) fully characterizes the set of perfect equilibrium payoffs when players are sufficiently patient. Strategies that sustain these payoffs are also well understood, due to characterizations by Abreu (1988) and Abreu, Pearce, and Stacchetti (1990). The simple benchmark model of infinitely repeated games with discounting provides insights into long-term relationships between as diverse actors as countries, firms, contractors, and spouses.

But how much are the conclusions of the Folk Theorem driven by simplifying assumptions in the model? In particular, the benchmark model assumes stationary payoffs and common knowledge of player’s payoffs functions, and the equilibrium analysis of Abreu (1988), Abreu, Pearce, and Stacchetti (1990), and Fudenberg and Maskin (1986) relies on this assumption. In real life situations on the other hand, players’
utilities vary through time, are affected by past decisions, and players may possess asymmetric information about their preferences.

The aim of this paper is to study the robustness of equilibrium strategies to small, but as arbitrary as possible, deviations from the benchmark model of discounted repeated games. Such robustness is highly desirable when interpreting equilibria as codes of law (or conventions, or social norms). These codes, as for instance civil law, etiquette, driving codes, internet protocols, customs, help resolve the coordination problem arising from the huge multiplicity of equilibria in repeated games. They tend to be stable through time, and apply to agents with similar, but non-identical preferences, hence they must be robust to idiosyncratic shocks in agents’ preferences.

Our starting point is a discounted repeated game, which we call our benchmark game. We introduce a series of perturbations of the benchmark game, which we call the real games. In real games, we allow payoffs to be stochastic, i.e., stage payoffs can depend on past histories as well as present actions. Since our goal is to study the design of strategies that form equilibria of all real games that are sufficiently close to a given benchmark game, we first need to define a proximity notion between games. It turns out that two parameters are needed to describe such proximity: closeness in payoffs, and influence of the past.

We say that a real game is $\varepsilon$-close to the benchmark game if, for every past history, the payoffs (as a function of chosen actions) in the real game are $\varepsilon$-close to the payoffs in the benchmark game. Thus, independently of the past history, the benchmark game captures players’ current preferences in the real game up to $\varepsilon$.

Payoff proximity may not be enough to capture players’ strategic incentives in a repeated game, especially if these players are patient. Indeed, strategic choices in any stage can have a long impact on future payoffs, and, even if the impact on each future stage payoff function is bounded by $\varepsilon$, the cumulative impact can be unbounded. Since strategic incentives are driven by the cumulative impact of choices on future payoffs, a bound on the cumulative influence on present choices on future payoffs is needed. We measure this influence with a notion of \textit{influence of the past}. In words, a real game has $M$ influence of the past if, given any two past histories in the game, and assuming that players follow the same course of actions following each of these histories, the cumulative difference in payoffs following these two histories does not exceed $M$.

The appropriate measure of proximity of a real game to a benchmark game thus consists of the pair of parameters $\varepsilon$ and $M$, and we define a
($\varepsilon, M$) version of a benchmark game as any real game that has payoffs $\varepsilon$-close to the benchmark game and $M$ influence of the past.

It is useful to think of code robustness as a design problem faced by a law-maker whose aim is to device an equilibrium of the repeated game, henceforth called a penal code, with minimal knowledge of player’s patience levels and of player’s preferences. Given a benchmark game, the goal is to design a penal code that respects strategic incentives under the assumptions that the benchmark game is a good approximation of the real game and that players are sufficiently patient. We analyze both complete penal codes, which are complete description of players’ strategies, hence that prescribe a (possibly random) choice of actions after any history in the repeated game, and incomplete penal codes, which provide such prescriptions after some histories only.

Our main findings are the limited robustness of complete penal codes on the one hand, and the existence of robust incomplete penal codes that allow to generate a Folk Theorem in the other.

The limited robustness of complete codes is shown in two results. We define a ($\varepsilon, M$)-robust complete penal code as a complete code that is a perfect equilibrium of every ($\varepsilon, M$) version of this game, if players are sufficiently patient. Our Theorem 1 shows that, given any benchmark game, there exists a constant $M$ of same order of magnitude as the payoffs in the stage game such that, for any $\varepsilon > 0$, no ($\varepsilon, M$)-robust complete penal code exists. Hence, for a code to be robust, influence of the past must be limited. Proposition 1 shows that, furthermore, in a non-degenerate class of games, there exist no ($\varepsilon, M$)-robust complete penal code, no matter how small $\varepsilon$ and $M$. For these games, ($\varepsilon, M$) robustness is simply impossible to fulfill using complete penal codes.

For an incomplete penal code to implement a payoff in a real game, two characteristics are required. The first one is that the code is prescriptive enough so that it unambiguously induces the desired payoff as long as players respect the code in histories at which is it defined. The second characteristic is that the code should be incentive compatible whenever it prescribes choices to the players. The difficulty we must address is that this incentive compatibility depends on player’s behavior on histories where the code is silent. Our approach is to consider that this behavior is itself dictated by strategic incentives, hence must form an (subgame perfect) equilibrium of the auxiliary game in which players choose their actions when the code is silent, but respect the code whenever it prescribes choices. We say that a code is credible is a given real game when every such equilibrium of the auxiliary game, together with the incomplete code, forms a subgame perfect equilibrium. Hence
our approach does not only impose that strategic incentives on the domain of the code are respected for some equilibrium behavior outside of this domain, but for every such equilibrium behavior. This requirement is motivated by the fact that a law maker who designs a code cannot credibly force players to coordinate on one equilibrium rather than another in situations where rules are silent on what behavior to adopt.

Given a benchmark game, we say that an incomplete code is \((\varepsilon, M)\) robust, when it is credible in every \((\varepsilon, M)\)-version of the benchmark game if players are sufficiently patient. Our main result in incomplete codes is a Folk Theorem. We show that if the stage game satisfies Fudenberg and Maskin (1986)'s full dimensionality condition, for any payoff vector \(x\) that is feasible and individually rational, if \(\varepsilon\) is sufficiently large, for any value of \(M\), there exists an \((\varepsilon, M)\) robust incomplete code that implements \(x\). Note that the code depends on \(M\) (as well as \(x, \varepsilon\)) but there is no upper bound on \(M\). Codes can be designed to be \((\varepsilon, M)\) robust for arbitrarily large values of the influence of the past.

Several conclusions can be drawn from these results. First, limited robustness of complete codes, or even their failure to exist, show that equilibria of repeated games are quite dependent on the modeling details. It also shows that designing rules that respect strategic incentives independently of these details is difficult at best, or worse, impossible. Incomplete codes deliver a much more optimistic message: Such codes exist that are both sufficiently normative to implement a chosen payoff, while being permissive enough never to violate the players' strategic incentives. Furthermore, a law maker who designs a robust incomplete code has the power of the Folk Theorem in choosing the payoff implemented.

1.1. Benchmark game and real games. We first describe the benchmark game, then the real game.

Given a finite set \(X\), \(\Delta(X)\) represents the set of probability distributions over \(X\). For a vector \(x\) in \(\mathbb{R}^I\), we let \(\|x\|_\infty\) represent \(\max_i \{|x_i|\}\).

1.1.1. The benchmark game. Our benchmark is the standard model of repeated discounted games which we recall here.

The set of players is a finite set \(I\) and \(A_i\) is player \(i\)'s finite action set, so \(S_i = \Delta(A_i)\) is player \(i\)'s set of mixed strategies. The set of action profiles is \(A = \prod_i A_i\), and in the benchmark stage game, the vector payoff function is \(\hat{g} : A \to \mathbb{R}\), with \(\hat{g}(a) = (\hat{g}_i(a))_i\).

A history of length \(t\) is an element of \(H_t = A^t\), with the convention that \(H_0 = \{\emptyset\}\). A strategy for player \(i\) is a mapping \(\sigma_i : \cup_{t \geq 0} H_t \to S_i\).
A profile of strategies $\sigma = (\sigma_i)_i$ induces a probability distribution $P_\sigma$ over the set of plays $H_\infty = (A)^\mathbb{N}$, and we let $E_\sigma$ denote the expectation under this probability distribution. The normalized discounted payoff to player $i$ from the play $h_\infty = (a_t)_{t \geq 1}$ is

$$\gamma_i(h_\infty) = (1 - \delta) \sum_{t \geq 1} \delta^{t-1} g_i(a_t).$$

Characterizations of the set of subgame perfect equilibrium payoffs of $\hat{G}_\delta$ are well known, and were first provided by Abreu (1988), and Abreu, Pearce, and Stacchetti (1990). The Folk Theorem by Fudenberg and Maskin (1986, 1991) characterizes the limit set of subgame perfect equilibrium payoffs when players are patient. The benchmark repeated game $\hat{G}_\infty$ represents the family of repeated games $(\hat{G}_\delta)$, parametrized by $\delta$. It is formally given by the data $(I, (A_i), \hat{g})$.

1.1.2. The real game. In the real game, stage payoffs may depend not only on current actions, but also on the past history. We thus consider the payoff at any stage as a function of the whole history up to that stage.\footnote{Formally, the real game is thus a stochastic game, where the set of states of nature is identified to the set of histories.} The payoff function is thus $g: \cup_{t \geq 1} H_t \rightarrow \mathbb{R}^I$, with the interpretation that $g_i(h_t)$ is the payoff to player $i$ at stage $t$ following history $h_t$.

In the $\delta$-discounted real game, the normalized discounted payoff to player $i$ from the infinite history $h_\infty = (a_t)_{t \geq 1}$ is

$$\gamma_i(h_\infty) = (1 - \delta) \sum_{t \geq 1} \delta^{t-1} g_i(h_t).$$

In Section 6, we consider games with incomplete information, and indicate further directions into which the model can be extended.

The real game $G_\infty$ is the class of repeated games $(G_\delta)$. Such a game is given by $(I, (A_i), g)$. As the set of players $I$ and the actions sets $A_i$ are fixed throughout the paper, strategy sets are the same in the benchmark game $\hat{G}_\delta$ and in the real game $G_\delta$. It is thus legitimate to ask under what conditions a profile of strategies in the benchmark game forms a subgame perfect Nash equilibrium of a corresponding real game.

2. Payoff proximity and influence of the past

Given two histories $h_t \in H_t$, $h'_t \in H_t$, $h_t, h'_t$ denotes the history in $H_{t+1}$ in which $h_t$ is followed by $h'_t$, and given $a_{t+1} \in A$, $h_t.a_{t+1} \in H_{t+1}$ represents $h_t$ followed by $a_{t+1}$.
2.1. **Payoff proximity.** Payoff proximity between a real game $G_{\infty}$ and a benchmark repeated game $\hat{G}_{\infty}$ is defined by comparing their stage payoff functions. Following history $h_{t-1}$, and assuming the action profile at stage $t$ is $a_t$, the payoff to player $i$ at stage $t$ is $g_i(h_{t-1},a_t)$ in the real game, and $\hat{g}_i(a_t)$ in the benchmark game. According to the following definition, payoffs in $\hat{G}_{\infty}$ and in $G_{\infty}$ are $\varepsilon$-close if all stage payoffs in $G_{\infty}$ are $\varepsilon$-close to the payoff induced by the same actions in $\hat{G}_{\infty}$.

**Definition 1.** Payoffs in $\hat{G}_{\infty}$ and in $G_{\infty}$ are $\varepsilon$-close if for every $t$, history $h_t \in H_t$ and $a \in A$:

\[ \|g(h_t, a) - \hat{g}(a)\|_\infty \leq \varepsilon \]

It is straightforward to see that, whenever payoffs in $\hat{G}_{\infty}$ and in $G_{\infty}$ are $\varepsilon$-close, for every discount factor $\delta$, the expected normalized payoffs induced in both games by a strategy profile $\hat{\sigma}$ in $\hat{G}_\delta$ are also $\varepsilon$-close. Hence, seen as games in normal forms, the payoff functions of $G_\delta$ and $\hat{G}_\delta$ differ by at most $\varepsilon$. One may conjecture that a strict equilibrium of every $\hat{G}_\delta$ for $\delta$ large enough remains an equilibrium of every sufficiently close real game. This is not true, as shown by the following example.

**Example 1 (The moral prisoner).** Consider the stage benchmark game $\hat{G}$ of the prisoner’s dilemma, with payoff function:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3,3</td>
<td>-1,4</td>
</tr>
<tr>
<td>D</td>
<td>4,-1</td>
<td>0,0</td>
</tr>
</tbody>
</table>

The prisoner’s dilemma

The trigger strategies recommend to play C if no player has ever played D in the past, and D otherwise. They are a subgame perfect equilibrium of the benchmark discounted game $\hat{G}_\delta$ for every $\delta \geq \frac{1}{4}$. In fact, they are the strategies that implement (C, C) for the largest range of discount factors as shown by Friedman (1971). They also form an optimal penal code in the sense of Abreu (1988). Are they a perfect equilibrium of every repeated game that is sufficiently close to the prisoner’s dilemma, provided that players are patient enough?

Fix an arbitrary $\varepsilon > 0$. In the real game $G_\delta$, the payoff function is constructed in such a way that, if player 1 played C while player 2 played D, and (D, D) was played ever since, player 1 obtains payoff of $\varepsilon$ instead of 0 on (D, D). This payoff can be interpreted as a psychological “bonus” for having tried to induce cooperation with the other player.
More formally, \( g(h_{t-1} \cdot a_t) \) as a function of \( a_t \) takes one of the two following forms, depending on \( h_{t-1} \):

\[
\begin{array}{c|cc}
   & C & D \\
\hline
C & 3, 3 & -1, 4 \\
D & 4, -1 & 0, 0 \\
\end{array}
\]

Payoff function \( \hat{g} \)

\[
\begin{array}{c|cc}
   & C & D \\
\hline
C & 3, 3 & -1, 4 \\
D & 4, -1 & \varepsilon, 0 \\
\end{array}
\]

Payoff function \( g' \)

The left payoff matrix is simply the payoff matrix of the prisoner’s dilemma. The right payoff matrix gives a “bonus” of \( \varepsilon \) to player 1 on \((D, D)\), this payoff matrix applies after histories in which player 1 has been the last player to play \( C \) while the opponent played \( D \): Consider a history \( h_{t-1} = (a_1, \ldots, a_{t-1}) \), and let \( \tau = \inf \{ t' : \forall t'' > t', a_{t''} = (D, D) \} \) be the latest stage at which \((D, D)\) has not been played, thus taking the value \(-\infty\) is \((D, D)\) was always played. We let \( g(h_{t-1}, a_t) = \hat{g}(a_t) \) if \( a_\tau \neq (C, D) \) or \( \tau > -\infty \), and \( g(h_{t-1}, a_t) = g'(a_t) \) if \( a_\tau = (C, D) \).

Since payoff functions in both state are \( \varepsilon \)-close to the payoff function of the prisoner’s dilemma, payoffs \( \hat{G}_\delta \) and \( G_\delta \) are \( \varepsilon \)-close.

We claim that for every \( \varepsilon > 0 \), the trigger strategies do not form a subgame perfect equilibrium of \( G_\delta \) whenever \( \delta > \frac{1}{1+\varepsilon} \). Consider a history \( h_{t-1} \) such that \( a_\tau \neq (C, D) \) and such that \( a_{t-1} = (D, D) \), such as, for instance \((D, C), (D, D), \ldots, (D, D) \). After \( h_{t-1} \), the trigger strategies recommend both players to play \( D \) forever. By doing so, player 1 obtains a future discounted payoff of 0. By playing action \( C \) first, then \( D \) forever, player 1 obtains a total future discounted payoff of \( (1 - \delta)(-1) + \delta \varepsilon \), which is positive for \( \delta > \frac{1}{1+\varepsilon} \).

In this example playing \((D, D)\) after every history is a strict Nash equilibrium of the discounted benchmark game for every value of \( \delta \), but it is not a Nash equilibrium of the real game for \( \delta \) sufficiently close to one.

2.2. Influence of the past. The driving force of Example 1 is that, in some situations, player 1’s actions have a long-lasting effect on the payoffs. This implies that, although the effect of an action on the payoff at any later stage is at most \( \varepsilon \), the cumulative discounted effect can still be very large. We introduce the notion of influence of the past to bound this cumulative effect. Consider two histories \( h_{t-1}, h'_{t-1} \in H_{t-1} \), and a sequence of actions \((a_t, a_{t+1}, \ldots)\). The absolute difference in payoffs to player \( i \) at stage \( t + \tau \) following \((a_t, a_{t+1}, \ldots)\), depending on whether \( h_{t-1} \) or \( h'_{t-1} \) was the previous history, is:

\[
|g_i(h_{t-1}, (a_t, \ldots, a_{t+\tau})) - g_i(h'_{t-1}, (a_t, \ldots, a_{t+\tau}))|
\]
and the sum of these quantities over $\tau \geq 0$ is the cumulative effect of the past history being $h_t$ or $h'_t$ on future payoffs to player $i$, following $(a_t, a_{t+1}, \ldots)$. We look for a bound on this quantity that is uniform in $i, h_t, h'_t$, and $(a_t, a_{t+1}, \ldots)$.

**Definition 2.** A real game $G_\infty$ has $M$ influence of the past if, for every $t$, histories $h_{t-1}, h'_{t-1} \in H_{t-1}$, sequence of actions $(a_t, a_{t+1}, \ldots)$:

$$\sum_{\tau \geq 0} \| g(h_{t-1}, (a_t, \ldots, a_{t+\tau})) - g(h'_{t-1}, (a_t, \ldots, a_{t+\tau})) \|_\infty \leq M$$

Note that the definition is independent of $\delta$, and doesn’t make reference to any benchmark game. The proximity notion defined below accounts both for payoff proximity and influence of the past.

**Definition 3.** A $(\varepsilon, M)$ version of a benchmark game $\hat{G}_\infty$ is a real game $G_\infty$ such that

1. Payoffs in $\hat{G}_\infty$ and in $G_\infty$ are $\varepsilon$-close,
2. $G_\infty$ has $M$ influence of the past.

The following example exhibits values of $\varepsilon$ and $M$ such that trigger strategies are, for $\delta$ large enough, a subgame perfect equilibrium in any $(\varepsilon, M)$ version of the prisoner’s dilemma.

**Example 2 (Robustness of trigger strategies).** Consider the benchmark repeated game $\hat{G}_\infty$ of the prisoner’s dilemma of Example 1. Let $\varepsilon = \frac{1}{2}$ and $M = \frac{2}{3}$. We claim that the trigger strategies form a subgame perfect Nash equilibrium of every $(\varepsilon, M)$ version of the benchmark repeated prisoner’s dilemma provided that $\delta$ is large enough. We apply the one-shot deviation principle (Blackwell, 1965).

The one-shot gain from deviating from $C$ (i.e., after a history containing $C$’s only) is bounded above by $(1 - \delta)(1 + 2\varepsilon)$ whereas the future loss is bounded below by $\delta(3 - 2\varepsilon)$.

The one-shot loss from deviating from $D$ (i.e., after a history containing at least one $D$) is bounded below by $(1 - \delta)(3 - 2\varepsilon)$ whereas the future gain is bounded above from $\delta(1 - \delta)M$.

It follows that for every $\delta \geq \frac{3}{4}$, trigger strategies form a subgame perfect equilibrium of every $(\varepsilon, M)$ version of $\hat{G}_\infty$.

3. **Limited robustness of complete codes**

A strategy profile $\sigma$ prescribes a choice to the players after every history, hence it is a complete rule of behavior. Since one of our objectives is to contrast the robustness of such codes with incomplete codes that
prescribe choices after some histories only, we refer in what follows to a strategy profile as a complete code.

We define a robust complete code of a benchmark game as a complete code that is a subgame perfect equilibrium of every game that is in a given neighborhood of that benchmark game, provided players are patient enough.

Definition 4. Given $\epsilon, M > 0$, $(\epsilon, M)$-robust complete code of $\hat{G}_\infty$ is a complete code $\sigma$ for which there exists $\delta_0$ such that for every $\delta \geq \delta_0$, $\sigma$ is a subgame perfect equilibrium of every $(\epsilon, M)$-version of $\hat{G}_\delta$.

Our first result shows that the robustness of complete codes is necessarily limited, in the sense that $(\epsilon, M)$ robustness imposes an upper bound on $M$ that depends only on the payoffs in the stage game.

Theorem 1. Consider a benchmark repeated game $\hat{G}_\infty$ and a player $i$ such that $A_i$ is not a singleton. With $M > \max_a g_i(a) - \min_a g_i(a)$ and $\epsilon > 0$, there exists no $(\epsilon, M)$-robust equilibrium of $\hat{G}_\infty$.

The intuition of the proof is the following. Consider a strategy profile $\sigma$, and a history $h_t$ that minimizes player $i$’s continuation payoff under $\sigma$. We construct a $(\epsilon, M)$-version of $\hat{G}_\infty$ in which player $i$’s one-stage loss from deviating is bounded by $\max_a g_i(a) - \min_a g_i(a)$, and this player’s cumulative discounted increase in his payoff function at future stages after that deviation is arbitrarily close (as $\delta \to 1$) to $M$. The fact that $h_t$ yields the worst continuation payoff ensures that this deviation of player $i$ cannot be punished by the other players.

Proof. Starting with a strategy profile $\sigma$ and a discount factor $\delta_0$, we construct a $G_\delta$ with $\delta > \delta_0$ that is a $(\epsilon, M)$ version of $\hat{G}_\infty$ such that $\sigma$ is not a subgame perfect Nash equilibrium of $G_\delta$.

Fix $i$ with at least two actions, and $M > \max_a g_i(a) - \min_a g_i(a)$. Choose $\epsilon' \in (0, \epsilon)$ and an integer $T$ such that $\epsilon'T = M$. Finally, let $\delta > \delta_0$ such that $\sum_{t=1}^{T} \delta_0^{t-1} > \max_a g_i(a) - \min_a g_i(a)$.

Given a strategy $\sigma_i$ for player $i$ the continuation strategy after $h_t$ is given by $\sigma_{i|h_t}(h'_t) = \sigma_i(h_t \cdot h'_t)$ for every $h'_t$. The profile of continuation strategies of $\sigma$ after $h_t$ is denoted $\sigma_{|h_t} = (\sigma_{i|h_t})_i$. The continuation payoff to player $i$ in $\hat{G}_\delta$ after history $h_t$ is:

$$\pi_i(h_t) = \mathbb{E}_{\sigma_{|h_t}} (1 - \delta) \sum_{\tau \geq 1} \delta^{\tau-1} g_i(a_{\tau})$$

Fix $t_0$ and a history $h_{t_0}$ such that

$$\pi_i(h_{t_0}) - \inf_{h_t} \pi_i(h_t) < (1 - \delta) \left( \sum_{t=1}^{T} \delta_0^{t-1} \epsilon' + \min_a g_i(a) - \max_a g_i(a) \right).$$
Now we construct an \((\varepsilon, M)\)-version \(G_\infty^\ast\) of \(\hat{G}_\infty\) such that \(\hat{\sigma}\) is not an equilibrium of \(G_\delta\). Fix \(a_i'\) such that \(\sigma_i(h_{t_0})(a_i') \neq 1\). In \(G_\infty\), player \(i\) receives a bonus of \(\varepsilon'\) during \(T\) stages for playing \(a_i'\) after \(h_{t_0}\). More precisely, \(g_j((a_1, \ldots, a_t)) = \hat{g}_j(a_i)\) if \(j \neq I\) or \(t \leq t_0 + 1\), and for every \(h_{t_0}' \in H_{t_0}, (a_{-i}, a_i) \in A_i, (a_1, \ldots, a_t) \in H_t:\n\)

\[
g_i(h_{t_0}', (a_{-i}, a_i))(a_1, \ldots, a_t) = \begin{cases} \hat{g}_i(a_i) + \varepsilon' & \text{if } h_{t_0}' = h_{t_0}, a_i = a_i', t \leq T, \\ \hat{g}_i(a_i) & \text{otherwise.} \end{cases}
\]

It is immediate from the definition of \(g\) that payoffs in \(G_\infty\) and \(\hat{G}_\infty\) are \(\varepsilon'\)-close and that \(G_\infty^\ast\) has \(\varepsilon'T = M\) influence of the past. Hence \(G_\infty^\ast\) is a \((\varepsilon, M)\) version of \(\hat{G}_\infty\).

Let \(a_i'' \neq a_i'\) be in the support of \(\sigma_i(h_{t_0})\). We show that in \(G_\delta\), playing \(a_i'\) yields after \(h_{t_0}\) yields a greater discounted payoff for player \(i\) than \(a_i''\), hence that player \(i\) has a profitable one-shot deviation from \(\sigma_i\).

The expected discounted payoff following \(a_i''\) after \(h_{t_0}\) is \(\pi_i(h_{t_0})\). The expected discounted payoff from playing \(a_i'\) after \(h_t\), then following \(\sigma_i\), is:

\[
(1 - \delta)g_i(a_i', \sigma_i(h_{t_0})) + \delta E_{\sigma_i} \pi_i(h_{t_0}.(a_i', a_{-i})) + (1 - \delta) \sum_{t=1}^{T} \delta^t \varepsilon'
\]

\[
\geq (1 - \delta) \min_a g_i(a) + \delta \inf_{h_t} \pi_i(h_t) + (1 - \delta) \sum_{t=1}^{T} \delta^t \varepsilon'
\]

\[
= (1 - \delta) \left( \min_a g_i(a) - \inf_{h_t} \pi_i(h_t) + \sum_{t=1}^{T} \delta^t \varepsilon' \right) + \inf_{h_t} \pi_i(h_t)
\]

\[
\geq (1 - \delta) \left( \min_a g_i(a) - \max_a g_i(a) + \sum_{t=1}^{T} \delta^t \varepsilon' \right) + \inf_{h_t} \pi_i(h_t)
\]

\[
> \pi_i(h_{t_0})
\]

Hence, \(G_\infty^\ast\) is an \((\varepsilon, M)\) version of \(\hat{G}_\infty\) such that, for every \(\delta > \delta_0\), \(\hat{\sigma}\) is not a subgame perfect Nash equilibrium of \(G_\delta\). This implies that \(\sigma\) is not a \((\varepsilon, M)\)-robust complete code of \(\hat{G}_\infty\).

Theorem 1 provides an upper bound on \(M\) such that an \((\varepsilon, M)\) robust complete code exists. It thus provides a bound on the robustness of any complete code. The next example shows that robust complete codes can fail to exist. More precisely, there are games that admit no \((\varepsilon, M)\)-robust complete codes, no matter how small (but positive) \(\varepsilon\) and \(M\) are.
Example 3 (A game with no robust complete code). Consider the following stage benchmark two player game:

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>2,1</td>
<td>-1,2</td>
</tr>
<tr>
<td>B</td>
<td>-1,1</td>
<td>1,1</td>
</tr>
</tbody>
</table>

\( \hat{G} \)

We show that no repetition of \( \hat{G} \) has an equilibrium in pure strategies. The min-max level for each player in pure strategies is 1. Hence any equilibrium in pure strategies should yield an expected of at least 1 to each player. But no payoff that weakly Pareto dominates \((1,1)\) is feasible, as every feasible vector payoff \( x = (x_1, x_2) \) satisfies \( x_1 + x_2 \leq 1 \).

The non-existence of robust codes follows from Proposition 1 below and from the fact that any robust complete code of \( \hat{G}_\infty \) is necessarily an equilibrium of \( \hat{G}_\delta \) for \( \delta \) sufficiently large.

Proposition 1. Let \( M, \varepsilon > 0 \). If \( \sigma \) is an \((M, \varepsilon)\)-robust complete code of \( \hat{G}_\infty \), then \( \sigma \) is a pure strategy profile.

Proof. Fix a player \( i \), a history \( h_t \) and an action \( \tilde{a}_i \in A_i \). Let \( \varepsilon' = \inf(M, \varepsilon) > 0 \). We define a real game \( G_\infty \) by \( g_j(h_{t'}, a_{t'+1}) = \hat{g}_j(a_{t'+1}) \) if \( h_{t'} \neq h_t \) or \( j \neq i \), and

\[
g_i(h_t, (a_i, a_{-i})) = \begin{cases} 
\hat{g}_i(a_i, a_{-i}) + \varepsilon' & \text{if } a_i = \tilde{a}_i, \\
\hat{g}_i(a_i, a_{-i}) & \text{otherwise.}
\end{cases}
\]

It is immediate that \( G_\infty \) is an \((\varepsilon', M)\)-version of \( \hat{G}_\infty \).

For any \( \delta > 0 \), the expected discounted payoff to player \( i \) playing any \( a_i \neq \tilde{a}_i \) following \( h_t \) is the same in \( G_\delta \) and in \( \hat{G}_\delta \), while the expected discounted payoff playing \( \tilde{a}_i \) after \( h_t \) is \( (1 - \delta)\varepsilon \) larger in \( G_\delta \) compared to \( \hat{G}_\delta \). This shows that player \( i \) cannot be indifferent between actions \( a_i \) and \( \tilde{a}_i \) both in \( G_\delta \) and in \( \hat{G}_\delta \). Hence, if \( \sigma \) is \((\varepsilon, M)\)-robust, \( \sigma_i(h_t) \) puts either probability 0 or probability 1 on \( \tilde{a}_i \). Since the same reasoning applies for every \( i, \tilde{a}_i \) and \( h_t, \sigma \) must be a profile of pure strategies. □

4. A Folk Theorem in Robust Incomplete Codes

Given the negative results of Section 3, we investigate the robustness of incomplete codes and the payoffs that can be generated by such codes. Our main result is Theorem 2 that provides a Folk Theorem in incomplete codes.

Definition 5. An incomplete code of \( \hat{G}_\infty \) is given by a subset \( D \) of \( \cup_{t \geq 1} H_t \) with a family \( s = (s_i)_i \) of mappings \( s_i : D \to A_i \).
An incomplete code $s$ therefore specifies player’s behaviors on the domain $D$, and leaves the behavior free outside of this domain. According to the following definition, an incomplete code induces a given payoff vector if the specification of the behavior on $D$ is sufficient to determine player’s payoffs if they are patient enough.

**Definition 6.** An incomplete code $s$ induces the vector payoff $x$ in $\hat{G}_\infty$ if, for every strategy profile $\sigma$ such that $\sigma$ coincides with $s$ on $D$,

$$\lim_{\delta \to 1} E_\sigma(1 - \delta) \sum_{t \geq 1} \delta^{t-1} \hat{g}(a_t) = x.$$

Incomplete codes let players’ strategic behavior outside of $D$ depend on the instance of the real game $G_\infty$ that they are playing as well as on $\delta$. Given $G_\infty$ and $\delta$, we define a completion of $s$ in a $G_\delta$ as a subgame perfect Nash equilibrium of the game in which players choose strategies that coincide with $s$ on $D$, and make arbitrary choices outside of $D$.

**Definition 7.** Let $s = (s_i)_i$ be an incomplete code of $\hat{G}_\infty$. A completion of $s$ in $G_\delta$ is a strategy profile $\sigma = (\sigma_i)_i$ such that:

- $\sigma_i$ coincides with $s_i$ on $D$,
- for every $h_t \notin D$, $\sigma_i$ maximizes $i$’s continuation payoff in $G_\delta$ given $(\sigma_j)_{j \neq i}$.

A completion $\sigma$ of an incomplete code $s$ is not necessarily a subgame perfect Nash equilibrium of the game $G_\delta$, as incentive conditions may fail on $D$. The notion of credible incomplete code requires that every completion $\sigma$ of $s$ is a subgame perfect Nash equilibrium of $G_\delta$. A credible code is self-enforcing since no matter what completion is selected, players have incentives to respect the code whenever it is defined.

**Definition 8.** An incomplete code $s$ is credible in $G_\delta$ if every completion of $s$ in $G_\delta$ is a subgame perfect equilibrium of $G_\delta$.

Consider the situation of the designer of an incomplete code who has a benchmark game in mind. This designer would like to ensure that the designed code is credible, as long as player’s strategic incentives are not too far away from the ones given by the benchmark game, and players are sufficiently patient. This notion is captured by the definition of a robust incomplete code below.

**Definition 9.** An $(\varepsilon, M)$-robust incomplete code is an incomplete code $s$ for which there exists $\delta_0$ such that $s$ is credible in every $(\varepsilon, M)$ version of $G_\delta$ when $\delta \geq \delta_0$. 
4.1. **Robust incomplete codes for the prisoner’s dilemma.** Theorem 1 provides an upper bound on $M$ such that a complete code can be $(\varepsilon, M)$-robust. We now present $(\varepsilon, M)$-robust incomplete codes implementing cooperation in the repeated prisoner’s dilemma for $M$ arbitrarily large, thus showing that incompleteness increases robustness.

Take up the game $\hat{G}$ of the prisoner’s dilemma of Example 1. The incomplete code is defined by an algorithm. Strategies start on MP (the Main Path), and any deviation of player $i$ from MP or MP$j$ triggers a punishment phase $P(i)$. The length $P$ of punishment phases is a parameter of the code. The code is incomplete as it doesn’t specify player’s actions during a punishment phase.

**Proposition 2.** Let $\varepsilon < \frac{2}{3}$, and $M > 0$. If $P$ is a multiple of 3 such that

$$P > \frac{1 + 2\varepsilon + M}{\frac{2}{3} - 2\varepsilon},$$

then the code is a $(\varepsilon, M)$-robust code of the repeated prisoner’s dilemma.

**Proof.** We first show that if $\delta$ is large enough, in every sequential completion of the code, player $j \neq i$ plays $D$ $P$ times during $P(i)$. The payoff to player $j$ for playing $D$ $P$ times is at least $-\varepsilon$ for each of the $P$ stages, followed by a cycle consisting of $3 - \varepsilon$, $3 - \varepsilon$, $4 - \varepsilon$, (the case in which the payoff of $4 - \varepsilon$ comes last being is the least favorable of the three possibilities). In average discounted payoffs, this is at least:

$$\delta^P(3 + \frac{\delta^2}{1 + \delta + \delta^2}) - \varepsilon.$$

which converges to $\frac{10}{3} - \varepsilon$ as $\delta \to 1$.

Any other strategy of $j$ during $P(i)$ gives at most $4 + \varepsilon$ during $P$ stages, followed by a cycle of $3 + \varepsilon$, $3 + \varepsilon$, $-1 + \varepsilon$ (the payoff of $-1 + \varepsilon$ coming last is the most favorable), which corresponds to the maximal average discounted payoff of:

$$(1 - \delta^P)4 + \delta^P(3 - \frac{4\delta^2}{1 + \delta + \delta^2}) + \varepsilon.$$
which converges to $\frac{5}{3} - \varepsilon$ as $\delta \to 1$. Since $\varepsilon < \frac{5}{6}$, for $\delta$ large enough, player $i$’s payoff is larger when playing $D$ for $P$ periods than with other sequence of actions.

We now show that, for $\delta$ large enough, no player has incentives to deviate from MP. The payoff with no deviation is at least $3 - \varepsilon$. The payoff with and after a deviation is at most $(1 - \delta)4 + \delta^{P+1}(3 - \frac{4\delta^2}{1 + \delta + \delta^2}) + \varepsilon$, which converges to $\frac{5}{3} + \varepsilon$ when $\delta \to 1$. Hence deviations are not profitable for $\delta$ large enough since $\varepsilon < \frac{2}{3}$.

We finally show that there is no profitable deviation from MP$_i$ for $\delta$ large enough by comparing gains and losses from deviations at the deviation stage, during punishment, and after the punishment.

- The immediate gain from a deviation is at most $(1 - \delta)(1 + 2\varepsilon)$.
- During $P(i)$, the loss due to the punishment is at least $\delta(1 - \delta^P)(3 - \frac{4\delta^2}{1 + \delta + \delta^2}) - 2\varepsilon$.
- After $P(i)$, the play is the same whether player $i$ has deviated or not, but the streams of payoffs may not be identical since they depend on past play. Since the game has $M$ influence of the past, the cumulative (undiscounted) difference in payoffs is at most $M$, hence the cumulative gain from the deviation after $P(i)$ is at most $(1 - \delta)\delta^{P+1}M$.

The total net loss from the deviation is bounded below by

$$(1 - \delta)\left(-1 - 2\varepsilon + \delta\frac{1 - \delta^P}{1 - \delta}(-2\varepsilon + 3 - \frac{4\delta^2}{1 + \delta + \delta^2}) - \delta^{P+1}M\right)$$

When $\delta \to 1$, the term in the larger parenthesis converges to $\frac{5}{3}P - 1 - (P + 1)2\varepsilon - M$, which is positive by choice of $P$. Hence for $\delta$ large enough, deviations from MP$_i$ are not profitable.

4.2. A Folk Theorem in robust codes.

**Theorem 2.** Let $x$ be a feasible and strictly individually rational vector payoff. Assume $G$ satisfies the full dimensionality assumption. There exists $\varepsilon$ such that, for every $M > 0$, there exists a $(\varepsilon, M)$ robust code of $G_\infty$ that induces $x$.

Note that every payoff that is induced by a robust code is necessarily induced by an equilibrium of $G_\delta$. Hence, every such payoff is necessarily feasible and individually rational. This shows that Theorem 2 fully characterizes (the closure of) the set of payoffs that are induced by robust codes.

Note that it is unavoidable that $\varepsilon$ depends on $x$. 
Corollary 1. Let $M > 0$, and $(\epsilon_n)_n \to 0$, $(\delta_n)_n \to 1$. For every $n$, let $G^\epsilon_n$ be a $(\epsilon_n, M)$-version of $G_\infty$. Then the closure of the set of subgame perfect Nash equilibrium payoffs of $G^\epsilon_n$ goes to $F \cap IR$ as $n$ goes to $\infty$.

5. Construction of robust codes

5.1. Overview. Let $x \in F \cap IR$, and $M > 0$. We exhibit $\epsilon$ and, for every $M$, construct an incomplete code $s$ such that:

1. $s$ implements $x$,
2. there exists $\delta$ such that, given any $(\epsilon, M)$-version $G^\epsilon_\infty$ of $\hat{G}_\infty$, $s$ is credible in every $G^\delta_\infty$.

The structure of $s$ is as follows. A Main Path consists of a sequence of actions that implements $x$. For each subset $J$ of $I$, a Reward Path for $J$ consists of a sequence of actions that, compared to $x$, yields a bonus to players in $J$, and a penalty to other players. If a player deviates from either the Main Path or from some Reward Path, players enter a Punishment Phase lasting $P$ stages during which the code doesn’t prescribe strategies. After these $P$ stages, a joint statistical test is applied over the actions played during the punishment phase to determine a subset $J$ of effective punishers. The Reward Path for $J$ is then played for some $R$ stages, after which players revert to the Main Path.

The Reward Path ensures that after any deviation, every punisher has incentives to pass the statistical test. The test is constructed in such a way that (i) Each player (except for the punished player) can ensure to be an effective punisher with high probability (ii) Conditional on every punisher being effective, the payoff received by the punished player during the punishment phase is closed to the min max value.

5.2. Selection of reward payoffs. Throughout the proof, we assume wlog. that for every $a \in A$, $\|g(a)\|_\infty \leq 1$. Given $x$ in the interior of $F \cap IR$, we choose $r > 0$ such that, i) for every $i \in I$, $v_i + 3r < x_i$, and ii) for every subset $J$ of $I$, the vector payoff $x^J$ given by:

$$
 x^J_i = \begin{cases} 
 x_i + r & \text{if } i \in J \\
 x_i - r & \text{otherwise.}
\end{cases}
$$

is in $F$. Such $r > 0$ exists because $x$ is interior and because of the full dimensionality assumption. The payoff vector $x^J$ is a reward for players in $J$ and only for those players.

5.3. The joint statistical test. We construct the test $\Phi^i_\beta$, parametrized by $\beta > 0$, according to which the set $J \subseteq I - \{i\}$ of effective punishers of player $i$ is computed. $\Phi^i_\beta$ inputs a history of length $P$ during
the punishment phase, and outputs a subset $J$ of effective punishers. Formally, $\Phi^i_{\beta} : \cup_{T \geq P} H_T \rightarrow 2^{I-i}$.

Let $(m^j_i)_{j \neq i} \in \prod_{j \neq i} S_j$ be a profile of mixed strategies achieving the min max in mixed strategies against player $i$. Given a history $h_T = (a_1, \ldots, a_T) \in H_T$, $T \geq P$ and $a \in A$, we let $n_{h_T}(a) = \#\{t, a_T = a\}$ denote the number of occurrences of $a$ during the last $P$ stages in $h_T$. The number of occurrences of $a \neq j \in A_j$ in the last $P$ stages of $h_T$ is $n_{h_T}(a-j) = \sum_{a_j \in A_j} n_{h_T}(a-j, a_j)$. $\Phi^i_{\beta}(h_T)$ is defined as the set of players $j \neq i$ such that:

$$\frac{1}{P} \left| \sum_{a_j \mid a \neq j} n_{h_T}(a-j, a_j) - m^j_i(a_j)n_{h_T}(a-j) \right| < \beta.$$

In order to pass the test, the action frequency of player $j$ must be close to the frequency given by $m^j_i$, independently of the actions chosen by the other players.

The test $\Phi^i_{\beta}$ possesses two major properties: achievability and efficiency.

According to the achievability property, if $P$ is large enough, each player $j \neq i$ can, by playing the minmax strategy $m^j_i$, guarantee to pass the test with probability arbitrarily close to 1. This property, combined with large rewards, will ensure that in every completion of $G_\delta$, all punishers are effective with large probability provided $\delta$ is close enough to 1. More formally, if $\tilde{m}^j_i$ represents the strategy of player $j$ in the repeated game that plays $m^j_i$ at every stage, we have:

**Lemma 1 (Achievability).** Let $\alpha, \beta > 0$. There exists $P_0(\alpha, \beta)$ such that, for every $P \geq P_0$ and every strategy profile $\sigma_{-j}$,

$$P_{\tilde{m}^j_i, \sigma_{-j}}(i \in \Phi^i_{\beta,j}(h_P)) \geq 1 - \alpha.$$

The efficiency property states that, if all punishers are effective, then the payoff received by the punished player is close to the min max payoff.

**Lemma 2 (Efficiency).** Let $r > 0$, there exists $\beta > 0$ such that, with $h_P = (a_1, \ldots, a_P)$, $\Phi^i_{\beta}(h_P) = I - \{i\}$ implies

$$\frac{1}{P} \sum_{t=1}^P g_i(a_t) < v_i + r.$$
at least $\alpha$, is:

$$V_i(\alpha, \beta) = (1 - (I - 1)\alpha) \max_{\Phi^*_j(a_1, \ldots, a_t) = I \setminus \{i\}} \frac{1}{\mathcal{P}} \sum_{t=1}^{P} g_i(a_t) + (I - 1)\alpha$$

From Lemma 2, $\lim_{\alpha, \beta \to 0} V_i(\alpha, \beta) = v_i$ for every $i$. We now fix $\alpha$ and $\beta$ such that $V_i(\alpha, \beta) < v_i + r$. This ensures that $V_i(\alpha, \beta) < x_i - 2r$ for every $i$.

5.4. **Definition of the incomplete code.** The incomplete code is parametrized by $P$ and $R$. The parameter $P$ is the duration (equal for all $i$) of a punishment phase $P(i)$ against player $i$, and $R$ is the duration (equal for all $J$) of a reward phase $R(J)$ following a punishment phase.

We select, both for $x$ and for each $x^J$, a sequence of actions that implements this vector payoff. As shown by the following lemma, this sequence can be selected in such a way that the average payoff vector over any $T$ consecutive periods converges to the target payoff vector at a speed $\frac{1}{\sqrt{T}}$.

**Lemma 3.** Let $y \in \mathcal{F}$. There exists a sequence of action profiles $\tilde{a} = (a_t)_t$ and a constant $K > 0$ such that for every $T \geq 1$,

$$\left\| \frac{1}{T} \sum_{t=1}^{T} g(a_t) - y \right\|_{\infty} < \frac{K}{\sqrt{T}}$$

We say that such a sequence $\tilde{a} = (a_t)_t$ of actions implements $y$.

Let $\tilde{a} = (a_t)_t$ be a sequence of actions that implements $x$. We can (see ... ) choose $\tilde{a}$ such that, for some constant $B \geq 0$, the average payoff over $B$ consecutive stages along $\tilde{a}$ is $\frac{e}{2}$-close to $x$. For $J \subseteq I$, let $\tilde{a}^J = (a_t^J)_t$ implement $x^J$. We let $K$ be a constant as in Lemma 3 that applies both to $\tilde{a}$ and to $\tilde{a}^J$ for every $J$.

In the incomplete code below, strategies start on the Main Path, denoted MP.

**MP** Play $\tilde{a}$, using action profile $a_t$ at stage $t$;

**P(i)** If player $i$ deviates from MP or $R(J)$, the code is incomplete for $P$ periods. At the end of this phase, continue to $R(J)$, where $J$ is the set of effective computed using $\Phi^*_j$.

**R(J)** Play the sequence $\tilde{a}^J$, in reverse order: $(a_t^J, \ldots, a_t^1)$. Then return to MP.

5.5. **Condition for effective punishments.** We show that if rewards are large enough and $P$ is such that any punisher can ensure to be effective with probability at least $1 - \frac{\alpha}{2}$, then in any sequential completion, every punisher is effective with probability at least $1 - \alpha$. 


Given an history \(h_T\), strategy profile \(\sigma\) and \(h_t \in H_t\) we let \(P_{\sigma,h_T}(h_T,h_t)\) denote the probability of \(h_t\) induced by \(\sigma\) following \(h_T\).

**Definition 10.** Punishments are \((\varepsilon, M, \alpha)\)-effective if there exists \(\delta_0\) such that, for any \((\varepsilon, M)\)-version \(G_\delta\) of \(\hat{G}\), for every completion \(\sigma\) of \(s\) in \(G_\delta\), and every history \(h_T\) that ends with a deviation of player \(i\)

\[
P_{\sigma,h_T}(\Phi^i_{\beta,j}(h_T+p) \neq I-\{i\}) \leq 1 - (I-1)\alpha.
\]

**Lemma 4.** If

1. \(P \geq P_0(\alpha, \beta)\)
2. \[\frac{1}{R} \left( M + K\sqrt{R} + P(1+\varepsilon) \right) \leq \frac{\alpha r}{2} - \varepsilon,\]

then punishments are \((\varepsilon, M, \alpha)\)-effective.

The selection of \(\alpha\) and \(\beta\) ensures that, if punishments are \((\varepsilon, M, \alpha)\)-effective, then for \(\delta\) sufficiently large, for any \((\varepsilon, M)\)-version \(G_\delta\) of \(\hat{G}\), and for every completion \(\sigma\) of \(s\) in \(G_\delta\), the expected average payoff to player \(i\) during any punishment phase \(P(i)\) is less than \(x_i - 2r\).

5.6. **Robustness of the incomplete code.** Remember that for \(x\) in the interior of \(F \cap IR\), the parameters \(r, \alpha\) and \(\beta\) are fixed. We now let \(\varepsilon = \frac{\alpha r^2}{8}\). Given \(M\), we exhibit values of \(P\) and \(R\) such that the incomplete code described is \((\varepsilon, M)\)-robust.

**Lemma 5.** If punishments are \((\varepsilon, M, \alpha)\)-effective and

\[Pr \geq 2K\sqrt{R} + 2\varepsilon (P + R + 1) + 2 + B + M,\]

then the incomplete code \(s\) is \((\varepsilon, M)\)-robust.

The following Lemma concludes the proof of Theorem 2.

**Lemma 6.** Given any \(M\), there exist \(P, R\) that satisfy the conditions of Lemma 4 and Lemma 5.

### 6. Incomplete Information

We so far have assumed complete information, in that all players have common knowledge of the payoff function. This has the virtue of keeping the model relatively simple. This also allows to show complete codes are non-robust when a small class of perturbations of the benchmark game is considered. On the other hand, the complete information assumption seems quite strong, and it is desirable to design codes that are robust to incomplete information.

In this section, we introduce a very general class of repeated games with incomplete information, define closeness of such a game to a
benchmark game (with complete information), and introduce the notion of robustness of a code in incomplete information, called tail robustness. When particularized to games with complete information, tail robustness is a stronger notion than robustness.

We then show that the results of section 5 are tail-robust in incomplete information, which strengthens the proof of theorem 2 by both requiring a stronger notion of robustness, and extending the class of real games considered.

6.1. **Real games with incomplete information.** The play of a real game with incomplete information starts at \( t = 0 \) when a state of nature \( k \) is drawn from a finite set \( K \) according to a probability distribution \( q_K \in \Delta(K) \). At the end of each stage \( t = 0, 1, \ldots \), each player \( i \) receives a private signal \( z_{t,i} \) in a finite set \( Z_i \), and the profile of signals is \( z_t = (z_{t,i}) \). A complete history after stage \( t \) is \( h_t = (k, z_0, a_1, z_1, \ldots, a_t, z_t) \), and the distribution of signal profiles at stage \( t \) is \( q(h_t, a_{t+1}) \). The payoff profile at stage \( t \) is \( g(h_t) \). The real game with incomplete information \( G_\infty \) is given by \( (q_K, q, g) \). The set of private histories to player \( i \) of length \( t \) is \( H_{i,t} = Z_i \times (A \times Z_i)^t \), and a strategy for player \( i \) in the real game with incomplete information is a mapping \( \sigma_i : \cup_{t \geq 1} H_{i,t-1} \rightarrow \Delta(A_i) \).

The case in which each \( Z_i \) is a singleton corresponds to the model of real games with complete information.

The model of real games with incomplete information is very general in that signals can depend arbitrarily on actions and on the state of nature, in particular, it does not restrict to the classical case of incomplete information on one side (as in reputation games, or zero-sum games with incomplete information), or where information is public (as in games where players learn the payoff function). To the best of our knowledge, nothing is known in the literature on such general stochastic games with incomplete information.

6.2. **Payoff proximity and influence of the past.** We generalize the notions of payoff proximity and influence of the past to real games with incomplete information. A benchmark game \( (I, (A_i)_i, \hat{g}) \) is given as before.

We say that a real game with incomplete information \( G_\infty \) is \( \varepsilon \)-close to a benchmark game \( \hat{G}_\infty \) (with same action spaces) if for every history \( h_t \), action profile \( a \) and signal profile \( z \),

\[ \|g(h_t, (a, s)) - \hat{g}(a)\|_\infty \leq \varepsilon. \]

A history \( h_{t-1} \) after stage \( t - 1 \) and a sequence of action profiles \( \check{a} = a_1, \ldots \) induce a probability distribution \( P_{h_{t-1}, \check{a}} \) on sequences of
signal profiles \( z_t, \ldots \), hence on histories \( h_\tau \) after any stage \( \tau \geq t \). We say that a real game with incomplete information has \( M \) influence of the past if, for every histories \( h_{t-1}, h'_{t-1} \) and sequence of action profiles \( \hat{a} \),

\[
\sum_{\tau \geq t} \| E_{P_{ht-1}, \hat{a}} g(h_\tau) - E_{P_{ht'-1}, \hat{a}} g(h_\tau) \|_\infty \leq M.
\]

Remark that when every \( Z_i \) is a singleton, a game with incomplete information is \( \varepsilon \)-close to a benchmark game in the sense above if and only if the corresponding game with complete information is \( \varepsilon \)-close to that game in the sense of definition 1, and that it has \( M \)-influence of the past in the sense above if and only if it has \( M \) influence of the past in the sense of definition 2. Thus, the following definition of a \((\varepsilon, M)\) version is an extension of definition 3 to games with incomplete information.

**Definition 11.** A real game with incomplete information \( G_\infty \) is an \((\varepsilon, M)\) version of a benchmark game \( \hat{G}_\infty \) if \( G_\infty \) is \( \varepsilon \)-close to \( \hat{G}_\infty \) and if \( G_\infty \) has \( M \) influence of the past.

In order to define sequential equilibria, we endow the strategy space of each player with the product topology. A belief system \( \mu \) specifies, for every private history in \( Z_i \times (A \times Z_i)^t \) a probability distribution over the set of histories after stage \( t \). We also endow the space of belief systems with the product topology.

6.3. **Tail credibility.** The straightforward way to extend the notion of completion of an incomplete code to games with incomplete information \( G_\delta \) would consist in considering a sequential equilibrium of the game in which players choose their strategies when the code doesn’t prescribe strategic choices, and payoffs are in \( G_\delta \). This raises the issue of defining player’s beliefs \( \mu \) on histories following a player’s deviation, say player \( i \) at stage \( t \). In order for \( \mu \) to correspond to the beliefs arising from \( i \)’s strategic incentives to deviate at stage \( t \), one needs to leave the choice of player \( i \)’s action at \( t \) endogenous, thus part of the definition of a completion. This requirement is in contradiction with the assumption that players prescribe to the code whenever it is defined.

We solve this difficulty by defining completions assuming that players follow the code after some stage \( T \) only, and later letting \( T \) go to infinity. Since player’s strategies define all their choices up to stage \( T \), all beliefs up to that stage, including following deviations, are well defined in sequential equilibria.

Note that the approach imposes a weaker requirement than assuming that players follow the code whenever defined, since it assumes that
players bind to the code after some stage $T$ only. The corresponding notion of robustness is thus stronger than the initial notion.

We now make definitions formal. Every history $h_t = (a_1, \ldots, a_t)$ in $G_\infty$ corresponds to the set $\{(s_{i,0}, a_1, z_{i,1}, \ldots, a_t, z_{i,t}), (z_{i,0}, \ldots, z_{i,t}) \in Z_i^{t+1}\}$ of private histories to player $i$ in $G_{\delta}$. Given an incomplete code $s$, its domain $D_s$ thus defines a subset of the set $\bigcup_{t \geq 0} H_{i,t}$ of private histories to player $i$ in $G_{\delta}$. The code $s = (s_i)$, naturally defines a mapping, still denoted $s_i$, from $D_{si}$ to $A_i$. The set $D_{siT} = \cap_{t \leq T} H_{i,t} \cup D_{si}'$ is the set of private histories to player $i$ that have either length less or equal than $t$, that do not belong to the domain of the code. We let $\Sigma_{siT} = \{\sigma: D_{siT} \rightarrow A_i\}$ be the set of strategies to player $i$ defined over $D_{siT}$. A strategy $\sigma_{siT}$ together with the code $s$ define a strategy $\sigma_i$ in $\Sigma_i$ given by $\sigma_i(h_{i,t}) = \sigma_{siT}(h_{i,t})$ if $h_{i,t} \in D_{siT}$ and $\sigma_i(h_{i,t}) = s_i(h_{i,t})$ otherwise. Thus, a profile of strategies $\sigma_{siT}$ defines a payoff for each player $i$ in $G_{\delta}$.

To define a completion of $s$ at $T$, we assume that players follow $s$ after stage $T$, and consider a sequential equilibrium of the remaining game.

**Definition 12.** A completion at $T$ of an incomplete code $s$ in $G_{\delta}$ is a sequential equilibrium in the game in which each player $i$’s strategy set is $\Sigma_{siT}$ and payoffs are as in $G_{\delta}$.

Now we define the corresponding notions of credibility at $T$ and tail credibility of an incomplete code. A code $s$ is credible at $T$ if every completion of $s$ at $T$ coincides with $s$ on its domain.

**Definition 13.** An incomplete code $s$ is credible at $T$ in $G_{\delta}$ if every completion $\sigma_{siT} = (\sigma_{siT})_i$ at $T$ of $s$ in $G_{\delta}$ is such that $\sigma_{siT}(h_{i,t}) = s_i(h_{i,t})$ for every $h_{i,t} \in D_{si}$. An incomplete code $s$ is tail credible if it is credible at $T$ for every $T$.

The notion of tail-credibility is stronger than the original credibility notion, since it relies only on the assumption that players eventually follow the code (after some stage $T$), instead of assuming that they follow the code whenever it is defined. Finally, the corresponding notion of tail-robustness is defined as follows:

**Definition 14.** An incomplete code $s$ is $(\varepsilon, M)$-tail robust if there exists $\delta_0$ such that $s$ is credible in every $G_{\delta}$ such that $G_\infty$ is an $(\varepsilon, M)$ version of $\hat{G}_\infty$ for $\delta \geq \delta_0$.

If every $Z_i$ is a singleton, a code which is $(\varepsilon, M)$-tail robust is also robust in the sense of definition 9.
6.4. Robustness results with incomplete information.

**Theorem 3.** The incomplete codes of section 5 are $(\varepsilon, M)$-tail robust in incomplete information.

We obtain as an immediate corollary the following result.

**Corollary 2.** Let $x$ be a feasible and strictly individually rational vector payoff. Assume $\hat{G}$ satisfies the full dimensionality assumption. There exists $\varepsilon$ such that, for every $M > 0$, there exists a code that induces $x$ and that is $(\varepsilon, M)$-tail robust in incomplete information.

We derive the following Folk Theorem for stochastic games with incomplete information.

**Theorem 4.** Assume $G$ satisfies the full dimensionality assumption and let $x$ be feasible and strictly individually rational. Given $d$ and $M > 0$, there exists $\delta_0 < 1$ and $\varepsilon > 0$ such that, for every $(\varepsilon, M)$-version $G_\infty$ of $\hat{G}_\infty$ with incomplete information and every $\delta > \delta_0$, there exists a sequential equilibrium payoff $y$ of $G_\delta$ such that $|y_i - x_i| < d$ for every player $i$.

**References**


Appendix A. Proof of Lemma 3

Lemma 3 is a consequence of the next lemma:

Lemma 7. Let $X$ be a compact subset of $\mathbb{R}^I$, and $y \in \text{co}X$. There exists a sequence $(x_i)_t$ of elements of $X$ and a constant $K > 0$ such that for every $T \geq 1$,

$$\left\| \frac{1}{T} \sum_{t=1}^{T} x_t - x \right\|_\infty < \frac{K}{\sqrt{T}}.$$

Proof. We prove the lemma using the norm $\|x\|_2$ given by $\sqrt{\sum_i x_i^2}$ instead of $\|x\|_\infty$. This is sufficient since $\|x\|_\infty \leq \|x\|_2$.

The case in which $y \in X$ is immediate. Otherwise, by Caratheodory’s Theorem, let $X' \subseteq X$ of cardinality at least 2 and at most $I + 1$ such that $y$ is a convex combination of elements of $X'$ with positive weights. Let $U$ be the intersection of a ball of radius 1 around $y$ with the space spanned with $X'$.

$$U = \{ z \text{ s.t. } \|z\|_2 = 1 \} \cap \{ \sum_{i \in I} \lambda_i x_i, (\lambda_i)_i \in \mathbb{R}^I, (x_i)_i \in X' \}$$

A separation argument shows that, for every element $z$ of $U$, there exists $x \in X'$ such that:

$$\langle x - y, z \rangle < 0$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product. By continuity of the scalar product, there exists $\rho > 0$ such that for every $z \in U$, there exists $x \in X'$ such that

$$\langle x - y, z \rangle \leq -\rho.$$

We construct $(x_i)_t$ inductively. Let $\gamma_t = \frac{1}{t} x_t$. If $t = 1$ or $\gamma_t = y$, let $x_{t+1} \in X'$ be arbitrary. Otherwise, by the above applied to $z = y + \frac{\gamma_t - y}{\|\gamma_t - y\|_2}$, we can choose $x_{t+1} \in X'$ such that:

$$\langle x_{t+1} - y, \gamma_t - y \rangle \leq -\rho \|y - \gamma_t\|_2$$
Let $b_t = t\|y - \gamma_t\|_2^2$. For every $T$, either $b_T = 0$, or

$$b_{T+1}^2 = \|\sum_{t=1}^{T+1} x_t - (T+1)Y\|_2^2$$

$$= b_T^2 + 2\langle\sum_{t=1}^{T} a_t - Ty, x_{T+1} - y\rangle + \|a_{T+1} - y\|_2^2$$

$$\leq b_T^2 - 2\rho b_T + M^2$$

where $M = \max_{x \in X'} \|y - x\|_2$. Note that $\langle x - y, z \rangle \geq -\|x - y\|_2$ for every $x \in X'$, $z \in U$ implies $M \geq \rho$. With $K = \frac{M^2}{\rho}$, we prove by induction that $b_T \leq K$ for every $T$. This is true for $T = 1$, since $b_1 = \|y - x_1\|_2 \leq M \leq \frac{M^2}{\rho}$. Now assume that $b_T \leq K$. Either $b_T = 0$, in which case $b_{T+1} = \|y - x_{T+1}\|_2 \leq M \leq \frac{M^2}{\rho}$, or $b_T \neq 0$, and in this case:

$$b_{T+1}^2 \leq \max_{0 \leq x \leq K} (x^2 - 2\rho x) + M^2$$

$$\leq \max \{K^2 - 2\rho K, 0\} + M^2$$

$$\leq \max \{K^2 - M^2, M^2\}$$

$$\leq K^2.$$

\[\square\]

**Appendix B. A Lemma**

In the proofs of Lemmata 4 and 5, we provide bounds on the streams of expected payoffs to players. The following lemma allows to derive bounds on the discounted sum of payoffs from bounds on the streams of payoffs.

**Lemma 8.** Let $T_0 \in \mathbb{N}$ and $a > 0$. There exists $\delta_0$ such that, for every $\delta \geq \delta_0$, and every sequence $(x_t)_t$ with values in $[-1, 1]$,

$$\sum_{t=1}^{T_0} x_t - \sum_{t=T_0+1}^{\infty} |x_t| > a$$

implies

$$\forall \delta \geq \delta_0, \sum_{t=1}^{\infty} \delta^t x_t > 0.$$

**Appendix C. Proof of Lemma 4**

Let $\delta_0$ be obtained from Lemma 8 using $T_0 = P + R$ and $a = \frac{M}{1 + \varepsilon}$. Let $\delta > \delta_0$, and $\sigma$ be a completion of $s$ in $G_\delta$. Consider a history $h_T$
ending with a deviation of player \( i \). We define
\[
x_t = \frac{E_{\sigma, m_j} g_j(h_{T+t}) - E\sigma g_j(h_{T+t})}{1+\varepsilon}
\]
with values in \([-1, 1]\), and let \( p = P_{\sigma, h_T} (i \in \Phi^i_{\beta, j}(h_{T+P})) \). We have the following bounds on payoffs from stage \( T + 1 \) to \( T + P \), and from \( T + P + 1 \) to \( T + P + R \):
\[
\sum_{t=1}^{P} E_{\sigma, m_j} g_j(h_{T+t}) \geq -P(1 + \varepsilon)
\]
\[
\sum_{t=1}^{P} E\sigma g_j(h_{T+t}) \leq P(1 + \varepsilon)
\]
\[
\sum_{t=P+1}^{P+R} E_{\sigma, m_j} g_j(h_{T+t}) \geq R((1 - \frac{\alpha}{2})(x_i + r) + \frac{\alpha}{2}(x_i - r)) - K\sqrt{R} - R\varepsilon
\]
\[
\sum_{t=P+1}^{P+R} E\sigma g_j(h_{T+t}) \leq R((1 - p)(x_i + r) + p(x_i - r)) + K\sqrt{R} + R\varepsilon
\]

Now we derive, using condition (2) of Lemma 4:
\[
(1 + \varepsilon) \sum_{t=1}^{P+R} x_t \geq -2 \left( P(1 + \varepsilon) + K\sqrt{R} + R\varepsilon \right) + 2Rr(p - \frac{\alpha}{2})
\]
\[
\geq -Rr\alpha + 2M + 2Rr(p - \frac{\alpha}{2})
\]
\[
\geq 2M + Rr(2p - \alpha)
\]

Since \( G \) has \( M \)-influence of the past, \( \sum_{t=R+P+1}^{\infty} |x_t| \leq M \), and
\[
\sum_{t=1}^{P+R} x_t - \sum_{t=P+R+1}^{\infty} |x_t| > \frac{M + Rr(2p - \alpha)}{1+\varepsilon}.
\]

From Lemma 8, \( p \geq \frac{\alpha}{2} \) implies \( \sum_{t=1}^{P+R} \delta^ix_t > 0 \), hence, following \( h_T \), playing \( m_i^j \) for \( P \) stages then following \( s \) would constitute a profitable deviation. This is a contradiction, therefore \( p \geq \frac{\alpha}{2} \).

**APPENDIX D. PROOF OF LEMMA 5**

Let \( \delta_0^i \) be obtained from Lemma 8 using \( T_0 = P + R + 1 \) and \( a = \frac{Pr}{2(1+\varepsilon)} \).

Let \( \delta_0^i \) be such that, for any \( (\varepsilon, M) \)-version \( G_\delta \) of \( \hat{G} \), for every completion \( \sigma \) of \( s \) in \( G_\delta \), and every history \( h_T \) that ends with a deviation of player \( i \)
\[
P_{\sigma, h_T} (\Phi^i_{\beta, j}(a_{T+1}, \ldots, a_{T+P}) \neq I - \{i\}) \leq 1 - (I - 1)\alpha.
\]
And let $\delta_0 = \max\{\delta'_0, \delta''_0\}$. Consider $\delta \geq \delta_0$, a $(\varepsilon, M)$-version $G$ of $\hat{G}$, and a completion $\sigma$ of $s$ in $G_\delta$.

We need to show that deviations after a history $h_T$ on the Main path or on some Reward Path cannot be profitable. Let $\sigma'_i$ be a strategy of player $i$ that deviates after $h_T$ and reverts to $\sigma_i$ once the punishment phase following the deviation is over, and denote $\sigma' = (\sigma_{-i}, \sigma'_i)$. Assume that, in $h_T$, $R'$ stages of a reward phase remain to be played, with $R' = 0$ if $h_T$ is on the Main Path. Let

$$x_t = \frac{E_{\sigma, h_T} g_j(h_{T+t}) - E_{\sigma', h_T} g_j(h_{T+t})}{1 + \varepsilon}.$$ 

We have:

$$\sum_{t=1}^{R'} E_{\sigma, h_T} g_j(h_{T+t}) \geq R' (x_i - r - \varepsilon) - K\sqrt{R'}$$

$$\sum_{t=R'+1}^{P+R+1} E_{\sigma, h_T} g_j(h_{T+t}) \geq (P + R - R' + 1)(x_i - \frac{r}{2} - \varepsilon) - B$$

$$\sum_{t=1}^{P+1} E_{\sigma', h_T} g_j(h_{T+t}) \leq 1 + P(V_i(\alpha, \beta) + \varepsilon)$$

$$\sum_{t=P+2}^{P+R+1} E_{\sigma', h_T} g_j(h_{T+t}) \leq R(x_i - r + \varepsilon) + K\sqrt{R}$$

Using the above inequalities, then the condition of Lemma 5, we deduce that:

$$(1 + \varepsilon) \sum_{t=1}^{P+R+1} x_t \geq - \left( K(\sqrt{R} + \sqrt{R'}) + 2\varepsilon(P + R + 1) + 1 - x_i + B \right) + \frac{3}{2} Pr$$

$$\geq - \left( 2K\sqrt{R} + 2\varepsilon(P + R + 1) + 2 + B \right) + \frac{3}{2} Pr$$

$$\geq \frac{1}{2} Pr + M$$

Note that $\sigma$ and $\sigma'$ induce the same path of actions from stage $P + R + 2$ on, and therefore, since $G$ has $M$ influence of the past:

$$(1 + \varepsilon) \sum_{t=P+R+2}^{\infty} |x_t| \leq M.$$
We now obtain
\[ \sum_{t=1}^{P+R+1} x_t - \sum_{t=P+R+2}^{\infty} |x_t| \geq \frac{Pr}{2(1 + \varepsilon)}. \]

Lemma 8 implies that:
\[ \sum_{t=1}^{\infty} \delta^t E_{\sigma, h_T} g(h_{T+t}) > \sum_{t=1}^{\infty} \delta^t E_{\sigma^*, h_T} g(h_{T+t}). \]

Hence that a deviation from \( \sigma \) on the Main Path or any Reward Path is not profitable.

**APPENDIX E. PROOF OF LEMMA 6**

Given \( \varepsilon = \frac{\alpha r^2}{8} \), note that
\[ \varepsilon (1 + \frac{r}{2} + \frac{\alpha r}{2}) < \frac{\alpha r^2}{4}, \]
hence
\[ \frac{2 - 2\varepsilon}{\varepsilon} > \frac{1 + \varepsilon}{\frac{\alpha r}{2} - \varepsilon}. \]

Let \( \lambda \) be such that
\[ \frac{2 - 2\varepsilon}{\varepsilon} > \lambda > \frac{1 + \varepsilon}{\frac{\alpha r}{2} - \varepsilon}. \]
and let \( R \) be the inferior integer part of \( \lambda P \). For \( P \) large enough, condition (2) of Lemma 4 is satisfied, since
\[ 1 + \varepsilon \leq (\frac{\alpha r}{2} - \varepsilon)\lambda, \]
and the condition of Lemma 5 is also satisfied, since
\[ r \geq 2\varepsilon (\lambda + 1). \]
Thus, if \( P \) is large enough and \( P \geq P_0(\frac{\alpha}{2}, \beta) \), punishments are \((\varepsilon, M, \alpha)\)-effective by Lemma 4, and from Lemma 5 the code is \((\varepsilon, M)\)-robust.