THE ROBUSTNESS OF INCOMPLETE PENAL CODES IN REPEATED INTERACTIONS

OLIVIER GOSSNER

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 players' strategies after every history, have only limited robustness. We define incomplete penal codes as partial descriptions of equilibrium strategies and introduce a notion of robustness for incomplete penal codes. We prove a Folk Theorem in robust incomplete codes that generates a Folk Theorem in a class of stochastic games.

Keywords: Repeated games, stochastic games, Folk Theorem, robust equilibrium.

1. Introduction

The theory of repeated games explains how long-term interactions may induce cooperation. The cornerstone of this theory is the celebrated Folk Theorem (Aumann and Shapley, 1994; Rubinstein, 1979; Fudenberg and Maskin, 1986), which 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 et al. (1990). The 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

Date: December 2020.

payoffs and common knowledge of player's payoffs functions, and the equilibrium analyses of Abreu (1988), Abreu et al. (1990), and Fudenberg and Maskin (1986) rely on this assumption. However, in many real life situations, players' utilities vary through time, and may be affected by past decisions on top of current ones.

The aim of this paper is to study the robustness of equilibrium strategies to small deviations from the benchmark model of discounted repeated games. Such robustness is highly desirable when equilibria are being interpreted as codes of law, conventions, or social norms. These codes, as for instance civil law, etiquette, driving codes, internet protocols, or customs, are stable through time, and the same code applies to agents with similar, but non-identical preferences. Hence, they must be robust to idiosyncratic shocks in agents' preferences and to different patience levels.

Our starting point is a discounted repeated game, which we call our benchmark game. We introduce a series of perturbed versions 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 a benchmark game and one of its perturbations. We describe such proximity using two parameters: closeness in payoffs and influence from the past.

We say that a real game is ε-close to the benchmark game if, for every past history, the stage payoffs in the real game are ε-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 ε. We consider this stage payoffs proximity as a basic requirement when defining closeness between repeated games.

It turns out that payoff proximity may not be enough to capture players' strategic incentives in a repeated game, especially if these players are very patient. Indeed, since real games are stochastic games, 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 some ε, the cumulative impact can be unbounded as the discount factor gets arbitrarily close to 1. Since strategic incentives are driven by the cumulative discounted impact of choices on future payoffs, a bound on the cumulative influence of present choices on future payoffs is needed. We measure this influence with a notion of *influence from the past*. In words, a real game has M influence from the past if, given any two past histories in the game, and assuming that players follow the same course of actions after 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 ε and M, and we define a (ε, M)-version of a benchmark game as any real game that has payoffs ε-close to the benchmark game and M influence from the past.

It is useful to think of code robustness as a design problem faced by a law-maker whose aim is to devise an equilibrium of the repeated game, henceforth called a penal code, with minimal knowledge of players' stage preferences and patience levels. Given a benchmark game, the law-maker's objective is to design penal codes that are incentive compatible 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 descriptions of players' strategies, hence 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 hand.

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

For an incomplete penal code to implement a payoff in a real game, two features are required. The first is that the code must be prescriptive enough so that it unambiguously induces the desired payoff as long as players follow the code in histories at which is it defined. The second feature 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 players' behaviors for histories where the code does not prescribe strategies. Our approach is to consider that this behavior is itself dictated by strategic incentives, and hence must form an (subgame perfect) equilibrium of the auxiliary game in which players choose their actions when the code is silent, but follow the code whenever it prescribes choices. We say that a code is *credible* in 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 followed for *some* equilibrium behavior outside of this domain, but for *every* such equilibrium behavior. This strong 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 (ε, M)-robust when it is credible in every (ε, M)-version of the benchmark game if players are sufficiently patient. Our main result is a robust Folk Theorem in incomplete penal codes. We show that if the stage game satisfies Fudenberg and Maskin's (1986) full dimensionality condition, for any payoff vector x that is feasible and individually rational, if ε is sufficiently small, for any value of M, there exists an (ε, M)-robust incomplete code that implements x. Note that the code depends on M (as well as x, ε) but there is no upper bound on M. Codes can be designed to be (ε, M)-robust for arbitrarily large values of the influence from the past.

Several conclusions can be drawn from these results. First, the 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 implemented payoff. The proof of our robust Folk Theorem does not rely on recursive techniques, but rather on statistical tests previously developed in the context of finitely repeated games by Gossner (1995). It provides an alternative proof of the Folk Theorem to the standard proof by Fudenberg and Maskin (1986). We note in passing that our folk theorem strategies are implementable by finite automata; see for instance Abreu and Rubinstein (1988), Aumann and Sorin (1989); Ben Porath (1993), and Neyman (1998) on games played by finite automata.

A strand of literature in repeated games focuses on robustness of equilibria to different patience levels for a fixed stage payoff function. The well studied notion of *uniform equilibrium* requires the same strategies to form approximate equilibria for all large enough patience levels. More recently, Cavounidis et al. (2018) studied strategies that form exact equilibria of all repeated games with large enough discount factors, and characterized the corresponding sets of equilibrium payoffs. This approach, however, does not perturb the stage payoff function. Other papers that introduced payoff perturbations in repeated games include Chassang and Takahashi (2011) in the context of equilibrium refinement *Ã la* global games, and Bhaskar et al. (2008) who study the purification of belief free strategies in repeated games.

Our results are connected to the literature of stochastic games. Under an irreducibility condition, Dutta (1995), Fudenberg and Yamamoto (2010) and Hörner et al. (2011) prove a Folk Theorem for stochastic games with finite state spaces. We prove (Theorem 1) a Folk Theorem for stochastic games with infinite state spaces with the further feature that equilibrium strategies between closed games coincide on the equilibrium path as well as on all histories that are not part of finite duration punishment phases. Our main result is not comparable with those previously obtained. On the one hand, our result relies on a proximity assumption between our game and a benchmark game, in particular the payoffs at every stage cannot be too far from the payoff in the benchmark game. On the other hand, we allow for a large class of transitions between states, and our stochastic games may not have a recursive structure. We refer the reader to the monographs by Mailath and Samuelson (2006) and Mertens et al. (2015) on repeated games.

The paper is organized as follows. Section 2 introduces benchmark and real games, and Section 3 defines the notions of proximity between repeated games. We study complete codes in Section 4 and incomplete codes in Section 5, where we present our Folk Theorem. The construction of robust incomplete codes is presented in Section 6.

2. Repeated games

2.1. **Benchmark game and real games.** 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 $\max_i\{|x_i|\}$.

2.1.1. *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; $S_i = \Delta(A_i)$ 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* $\hat{G}$, the vector payoff function is $\hat{g}\colon A \to \mathbb{R}^I$, 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 (behavioral) strategy for player i is a mapping $\sigma_i\colon \cup_{t\geq 0} H_t \to S_i$ and we let Σ_i denote the set of player i's strategies. A profile of strategies $\sigma = (\sigma_i)_i$ induces a probability distribution P_σ over the set of plays $H_\infty = (A)^{\mathbb{N}}$, and we let E_σ denote the expectation under this probability distribution. The normalized discounted payoff with discount factor $\delta \in (0,1)$ to player i from the play $h_\infty = (a_t)_{t\geq 1}$ is

$$\hat{\gamma}_i(h_\infty) = (1-\delta)\sum_{t\geq 1}\delta^{t-1}\hat{g}_i(a_t).$$

The *discounted benchmark game* $\hat{G}_\delta$ is the game with strategy sets $(\Sigma_i)_i$ and payoff function $\mathrm{E}_\sigma\,\hat{\gamma}_i(h_\infty)$. Characterizations of the set of subgame perfect equilibrium payoffs of $\hat{G}_\delta$ were provided by Abreu (1988), and Abreu et al. (1990). The Folk Theorem by Fudenberg and Maskin (1986, 1991) characterizes the limit set of subgame perfect equilibrium payoffs when players are patient.

Since we are interested in equilibria that are independent of δ for δ large enough, it is useful to call *benchmark game* and denote by $\hat{G}_\infty$ the family of repeated games $(\hat{G}_\delta)$, parametrized by δ.

2.1.2. *The real game.* In the *real game*, stage payoffs may depend not only on current actions, but also on the past history. The payoff at any stage is thus a function of the whole history up to that stage.[1] The stage payoff function is $g\colon \cup_{t\geq 0} H_t \to \mathbb{R}^I$, with the interpretation that $g_i(h_t)$ is the payoff to player i at stage t following history h_t.

The normalized δ-discounted payoff to player i from the infinite history h_∞ is

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

The δ-*discounted real game* G_δ is the game with strategy sets $(\Sigma_i)_i$ and payoff function $\mathrm{E}_\sigma\,\gamma_i(h_\infty)$.

We call *real game*, and denote by G_∞, the class of repeated games (G_δ). As the set of players I and the action sets A_i are fixed throughout the paper, strategy sets

[1]Formally, the real game can be viewed as a stochastic game, where the set of states of nature is identified to the set of histories.

Σ_i are the same in the benchmark game and in the real game. It is thus legitimate to compare equilibrium strategies in both frameworks.

3. Payoff proximity and influence from the past

Given two histories $h_t \in H_t$ and $h'_\tau \in H_\tau$, $h_t.h'_\tau$ denotes the history in $H_{t+\tau}$ in which h_t is followed by h'_τ. For $a \in A$, $h_t.a \in H_{t+1}$ represents h_t followed by a.

3.1. **Payoff proximity.** We define payoff proximity between a real game G_∞ and a benchmark game $\hat{G}_\infty$ 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_∞ are ε-close if all stage payoffs in both games are ε-close.

Definition 1. *Payoffs in $\hat{G}_\infty$ and in G_∞ are ε-close if for every 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_∞ are ε-close, for every discount factor δ, the expected normalized payoffs induced in both games by the same strategy profile are also ε-close. Hence, in their normal form representations, the payoff functions of G_δ and $\hat{G}_\delta$ differ by at most ε. One may conjecture that a strict equilibrium of $\hat{G}_\delta$ 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 benchmark stage game $\hat{G}$ of the prisoner's dilemma, with payoff function:*

	C	D
C	$3,3$	$-1,4$
D	$4,-1$	$0,0$

The prisoner's dilemma

The trigger strategies specify 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, Friedman (1971) showed that these strategies implement an infinite repetition of (C, C) for the largest possible range of discount factors. 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 discounted real game G_δ, the payoff function is such that, if player 1 played C while player 2 played D at some point in the past, and (D, D) was played ever since, player 1 obtains payoff of ε 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}:

	C	D
C	$3,3$	$-1,4$
D	$4,-1$	$0,0$

Payoff function $\hat{g}$

	C	D
C	$3,3$	$-1,4$
D	$4,-1$	$\varepsilon,0$

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 ε 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. More formally, 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, taking the value* $-\infty$ *if* (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 states are ε*-close to the payoff function of the prisoner's dilemma, payoffs* $\hat{G}_\infty$ *and* G_∞ *are* ε*-close.*

We claim that for every $\varepsilon > 0$*, the trigger strategies do not form a subgame perfect equilibrium of* G_δ *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* δ*, but it is not a Nash equilibrium of the real game for* δ *sufficiently close to one. It follows that the trigger strategies are not a subgame perfect equilibrium of the real game when* δ *is large enough, no matter how small* ε *is.*

3.2. **Influence from 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 ε, the cumulative discounted effect can still be very large. We introduce the notion of influence from 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_∞ has M influence from the past if, for every pair of histories $h_{t-1}, h'_{t-1} \in H_{t-1}$ and for every 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 δ, and does not make reference to any benchmark game. The proximity notion defined below accounts for both payoff proximity and influence from the past.

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

1. *payoffs in $\hat{G}_\infty$ and in G_∞ are ε-close,*
2. *G_∞ has M influence from the past.*

We take up the prisoner's dilemma from Example 1 and exhibit values of ε and M such that trigger strategies are a subgame perfect equilibrium in any (ε, M)-version of the prisoner's dilemma when δ is large enough,

Example 2 (Robustness of trigger strategies)**.** *Consider the benchmark game $\hat{G}_\infty$ of the prisoner's dilemma in Example 1. Let $\varepsilon = \frac{1}{4}$ and $M = \frac{1}{2}$. We claim that the trigger strategies form a subgame perfect Nash equilibrium of every (ε, M) version of* *the benchmark prisoner's dilemma, provided that δ 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 Cs only) is bounded above by $(1-\delta)(1+2\varepsilon)$, whereas the future loss is bounded below by $\delta(3-2\varepsilon)$. It is verified that the loss exceeds the gain for $\varepsilon = \frac{1}{4}$ and $\delta \geq \frac{3}{8}$.

The one-shot loss from deviating from D (i.e., after a history containing at least one D) is bounded below by $(1-\delta)(1-2\varepsilon)$, whereas the future gain is bounded above by $\delta(1-\delta)M$. Hence, the loss exceeds the gain for $\varepsilon = \frac{1}{4}$ and $M = \frac{1}{2}$.

It follows that for every $\delta \geq \frac{3}{8}$, the trigger strategies form a subgame perfect equilibrium of every $(\frac{1}{4}, \frac{1}{2})$-version of $\hat{G}_\infty$.

4. Limited robustness of complete codes

A strategy profile σ prescribes a choice to the players after every history; 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 in a given neighborhood of that benchmark game, provided players are patient enough.

Definition 4. *Given $\varepsilon, M > 0$, an (ε, M)-robust complete code of $\hat{G}_\infty$ is a complete code σ for which there exists δ_0 such that for every $\delta \geq \delta_0$, σ is a subgame perfect equilibrium of every (ε, M)-version of $\hat{G}_\delta$.*

Example 2 shows that trigger strategies are a $(\frac{1}{4}, \frac{1}{2})$-robust code of the repeated prisoner's dilemma. The next result shows that the robustness of complete codes is necessarily limited, in the sense that it imposes an upper bound on the influence from the past that depends only on the payoffs in the stage game.

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

The intuition of the proof is the following. Consider a strategy profile σ and a history h_t that minimizes player i's continuation payoff under σ. We construct a (ε, M)-version of $\hat{G}_\infty$ in which, after a deviation of player i following h_t, this player's cumulative increase on payoffs at future stages is 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. This deviation is then profitable, provided i is patient enough, as the one-shot loss from deviating is bounded above by $\max_a g_i(a) - \min_a g_i(a)$ which is less than the cumulative future gains given by M.

Proof. We start with a strategy profile σ. We construct a real game G_∞ that is a (ε, M)-version of $\hat{G}_\infty$ such that σ is not a subgame perfect Nash equilibrium of G_δ for any δ large enough.

Fix i such that $|A_i| \geq 2$, and $M > \max_a g_i(a) - \min_a g_i(a)$. Choose $\varepsilon' \in (0, \varepsilon)$ and an integer T such that $\varepsilon' T = M$. Finally, let $\delta > \delta_0$ such that $\sum_{t=1}^{T} \delta_0^t \varepsilon' > \max_a g_i(a) - \min_a g_i(a)$.

Given the strategy σ_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 after h_t is denoted $\sigma_{|h_t} = (\hat{\sigma}_{i|h_t})_i$. The continuation payoff to player i in $\hat{G}_\delta$ after history h_t is given by:

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

Fix t_0 and a history h_{t_0} such that

$$\pi_i(h_{t_0}) - \inf_{h_t} \pi_i(h_t) < (1-\delta)\left(\sum_{t=1}^{T} \delta_0^t \varepsilon' + \min_a g_i(a) - \max_a g_i(a)\right).$$

Now we construct an (ε, M)-version G_∞ of $\hat{G}_\infty$ such that σ is not an equilibrium of G_δ. Fix a_i' such that $\sigma_i(h_{t_0})(a_i') \neq 1$. In G_∞, player i receives a bonus of ε' during T stages for playing a_i' after H_{t_0}. More precisely, $g_j((a_1, \ldots, a_t)) = \hat{g}_j(a_t)$ if $j \neq i$ or $t \leq t_0 + 1$, and for every $h_{t_0}' \in H_{t_0}$, $(a_{-i}, a_i) \in A$, $(a_1, \ldots, a_t) \in H_t$:

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

It is immediate from the definition of g that payoffs in G_∞ and $\hat{G}_\infty$ are ε'-close and that G_∞ has $\varepsilon' T = M$ influence from the past. Hence G_∞ is a (ε, 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_δ, 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 σ_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 σ_i, is:

$$
\begin{aligned}
& (1-\delta)g_i(a_i', \sigma_{-i}(h_{t_0})) + \delta \operatorname{E}_{\sigma_{-i}} \pi_i(h_{t_0}\cdot(a_i', a_{-i})) + (1-\delta)\sum_{t=1}^{T}\delta^t\varepsilon' \\
\geq\ & (1-\delta)\min_a g_i(a) + \delta \inf_{h_t} \pi_i(h_t) + (1-\delta)\sum_{t=1}^{T}\delta^t\varepsilon' \\
=\ & (1-\delta)\left(\min_a g_i(a) - \inf_{h_t}\pi_i(h_t) + \sum_{t=1}^{T}\delta^t\varepsilon'\right) + \inf_{h_t}\pi_i(h_t) \\
\geq\ & (1-\delta)\left(\min_a g_i(a) - \max_a g_i(a) + \sum_{t=1}^{T}\delta^t\varepsilon'\right) + \inf_{h_t}\pi_i(h_t) \\
>\ & \pi_i(h_{t_0}).
\end{aligned}
$$

Hence, G_∞ is an (ε, M)-version of $\hat{G}_\infty$ such that, for every $\delta > \delta_0$, σ is not a subgame perfect Nash equilibrium of G_δ. This implies that σ is not a (ε, M)-robust complete code of $\hat{G}_\infty$. □

Theorem 1 provides an upper bound on M such that an (ε, 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 fail to exist for some games. In fact, some games admit no (ε, M)-robust complete codes, no matter how small ε and M may be.

Example 3 (A game with no robust complete code)**.** *Consider the following benchmark two player stage game:*

	L	R
T	$2,-1$	$-1,2$
B	$-1,1$	$1,-1$

$\hat{G}$

We show that no discounted repetition $\hat{G}_\delta$ has an equilibrium in pure strategies. The min max *level for each player in pure strategies is 1. Hence any equilibrium in pure strategies has to give an expected payoff of at least 1 to each player. However, no payoff that weakly Pareto dominates $(1,1)$ is feasible, as every feasible vector payoff $x=(x_1,x_2)$ satisfies $x_1+x_2 \le 1$.*

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 δ sufficiently large.

Proposition 1. *Let $M,\varepsilon>0$. If σ is an (M,ε)-robust complete code of $\hat{G}_\infty$, then σ 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_∞ by giving a bonus ε' to action $\tilde{a}_i$ of player i following history h_t: $g_j(h'_{t'}.a_{t'+1}) = \hat{g}_j(a_{t'+1})$ 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_∞ is an (ε', 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_δ and in $\hat{G}_\delta$, while the expected discounted payoff playing $\tilde{a}_i$ after h_t is $(1-\delta)\varepsilon$ larger in G_δ compared to $\hat{G}_\delta$. This implies that player i cannot be indifferent between actions $\tilde{a}_i$ and some other action, both in G_δ and in $\hat{G}_\delta$. Hence, if σ is (ε, 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, if σ is (ε, M)-robust, it must be a profile of pure strategies. □

5. A Folk Theorem in robust incomplete codes

Given the negative results of Section 4, we investigate the robustness of incomplete codes and the payoffs that can be generated by such codes. Our main result is Theorem 2, which provides a Folk Theorem in robust 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 $s_i \colon D \to A_i$.*

An incomplete code s therefore specifies players' behaviors on the domain D, and leaves the behavior unrestricted 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 σ such that σ coincides with s on D,*

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

Incomplete codes let players' strategic behaviors outside of D depend on the instance of the real game G_∞ that they face as well as on δ. We define a completion of

s in a discounted real game G_δ as a subgame perfect Nash equilibrium of the game in which players follow s on D and choose arbitrary strategies outside of D.

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

- *σ_i coincides with s_i on D,*
- *for every $h_t \notin D$, σ_i maximizes i's continuation payoff in G_δ given $(\sigma_j)_{j \neq i}$.*

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

Definition 8. *An incomplete code s is* credible *in G_δ if every completion of s in G_δ is a subgame perfect equilibrium of G_δ.*

Consider the problem 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 the players' strategic incentives are not too far away from those 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 (ε, M)-robust incomplete code is an incomplete code s for which there exists δ_0 such that, for every (ε, M)-version G_∞ of $\hat{G}_\infty$ and every $\delta \geq \delta_0$, s is credible in G_δ.*

5.1. **Robust incomplete codes for the prisoner's dilemma.** Theorem 1 provides an upper bound on M such that a complete code can be (ε, M)-robust. We now present (ε, 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 $\mathrm{P}(i)$. Two variations MP^1 and MP^2 of the Main Path serve as potential rewards to the punisher after a punishment phase. The length P of a punishment phase is a parameter of the code. The code is incomplete, as it doesn't specify players' actions during a punishment phase.

MP Play (C, C)

MP^i Play (C_i, D_j) $(j \neq i)$ at stages $3k$, $k \in \mathbb{N}$, (C, C) at other stages

$\mathrm{P}(i)$ Play for P stages, after which:

- continue to MP^j if player $j \neq i$ played D at each of the P stages,
- continue to MP^i otherwise.

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

$$P > \frac{1 + 2\varepsilon + M}{\frac{5}{3} - 2\varepsilon},$$

then the code is a (ε, M)-robust code of the repeated prisoner's dilemma.

Proof. We first show that if δ is large enough, in every sequential completion of the code, player $j \neq i$ plays P times the action D during $\mathrm{P}(i)$. The payoff to player j for playing P times D 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 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 player 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 a 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 δ large enough, player i's payoff is larger when playing D for P periods than with other sequences of actions.

We now show that, for δ large enough, no player has an incentive to deviate from MP. The payoff with no deviation is at least $3 - \varepsilon$. The payoff with and following 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 $\delta \to 1$. Thus deviations are not profitable for δ large enough, since $\varepsilon < \frac{2}{3}$.

We finally show that there is no profitable deviation from MP^i for δ 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 prior actions. Since the game has M influence from the past, the cumulative (undiscounted) difference in payoffs is at most M, thus 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

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

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 the choice of P. Hence for δ large enough, deviations from MP^i are not profitable. □

5.2. **A Folk Theorem in robust codes.** Before we state our main result, we recall some standard definitions. The set F of *feasible payoffs* of the benchmark stage game $\hat{G}$ is the convex hull of the stage payoffs: $F = \mathrm{co}(g(A))$. The min max of player i is

$$v_i = \min_{s_{-i}\in\prod_{j\neq i}S_j}\max_{a_i\in A_i}\mathrm{E}_{s_{-i}}\, g_i(s_{-i},a_i),$$

and the set of *individually rational payoffs* is

$$IR = \{x\in\mathbb{R}^I, \forall i\ x_i \geq v_i\}.$$

Following Fudenberg and Maskin (1986), we say that $\hat{G}$ satisfies the *full dimensionality assumption* when the dimension of $F\cap IR$ equals the cardinality of I, namely when it contains an open ball of the form $B(x,\varepsilon)$ for $x\in F\cap IR$ and $\varepsilon > 0$.

Theorem 2. *Assume $\hat{G}$ satisfies the full dimensionality assumption and let x be a feasible and strictly individually rational vector payoff. There exists ε such that, for every $M > 0$, a (ε, M)-robust code of G_∞ that induces x exists.*

Note that every payoff that is induced by a robust code is necessarily a limit of equilibrium payoffs of $\hat{G}_\delta$; hence such a 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.

Theorem 2 is a robustness result in strategies; it shows that all games close to the same benchmark game have credible codes in common, and that these codes generate a Folk Theorem. A consequence of this result is a Folk Theorem for these games, expressed in terms of sets of equilibrium payoffs, as follows.

Corollary 1. *Let $M > 0$, $(\varepsilon_n)_n \to 0$, and $(\delta_n)_n \to 1$. For every n, let G^n_∞ be a (ε_n, M)-version of $\hat{G}_\infty$. Then the closure of the set of subgame perfect Nash equilibrium payoffs of $G^n_{\delta_n}$ goes to $F \cap IR$ as n goes to ∞.*

The inclusion of $F \cap IR$ in the limit set follows from Theorem 2. For the other inclusion, notice that in every equilibrium of a $G^n_{\delta_n}$, each player i receives at least $v_i - \varepsilon_n$, and that every average discounted payoff in $G^n_{\delta_n}$ is ε-close to F.

6. Construction of robust codes

6.1. **Overview.** Given x in the interior of $F \cap IR$ and $M > 0$, we find ε and, for every M, construct an incomplete code s such that:

(1) s implements x,
(2) there exists $\overline{\delta}$ such that, given any (ε, M)-version G_∞ of $\hat{G}_\infty$, s is credible in every G_δ when $\delta > \overline{\delta}$.

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 any player deviates from either the Main Path or from some Reward Path, the players enter a *Punishment Phase* for 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 R stages, after which the 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 punisher has a strategy that ensures 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 this player's min max value.

6.2. **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:

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

is in F. Such $r > 0$ exists because x is interior and because of the full dimensionality assumption. Compared to x, the payoff vector x^J is a reward for the players in J and a punishment for the others.

6.3. **The joint statistical test.** We construct the test Φ^i_β, parametrized by $\beta > 0$, according to which the set $J \subseteq I - \{i\}$ of effective punishers of player i is computed. Φ^i_β inputs a history of length $T \geq P$, and outputs a subset J of effective punishers. Formally, it is a mapping $\Phi^i_\beta \colon \cup_{T \geq P} H_t \to 2^{I - \{i\}}$.

Let $(m^i_j)_{j \neq i} \in \prod_{j \neq i} S_j$ be a profile of mixed strategies such that

$$v_i = \max_{a_i \in A_i} \mathrm{E}_{(m^i_j)_j}\, g_i(a_{-i}, a_i).$$

Given a history $h_T = (a_1, \ldots, a_T) \in H_T$, $T \geq P$ and $a \in A$, we let $n_{h_T}(a)$ denote the number of occurrences of a during the last P stages of h_T:

$$n_{h_T}(a) = \#\{t > T - P,\, a_t{=}a\}.$$

The number of occurrences of $a_{-j} \in A_{-j}$ 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:

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

In order to pass the test, i.e. to belong to $\Phi^i_\beta(h_T)$, the frequency of actions of player j must be close to m^i_j, independently of the actions chosen by the other players.

The test Φ^i_β 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^i_j at every stage, be guaranteed to pass the test with probability arbitrarily close to 1. This property, combined with large rewards, will ensure that in every completion of G_δ, all punishers are effective with large probability, provided δ is close enough to 1. More formally, if $\tilde{m}^i_j$ represents the strategy of player j in the repeated game that plays m^i_j at every stage, we have:

Lemma 1 (Achievability)**.** *Let $\alpha, \beta > 0$. There exists $P_0(\alpha, \beta)$ such that, for every $P \geq P_0(\alpha, \beta)$, every pair of players $i, j \in I$ and every strategy profile σ_{-j},*

$$P_{\tilde{m}^i_j, \sigma_{-j}}(j \notin \Phi^i_\beta(h_P)) < \alpha.$$

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)**.** *There exists $\beta > 0$ such that, for every player $i \in I$ and $T \geq P$, if $\Phi^i_\beta(h_T) = I - \{i\}$ then*

$$\frac{1}{P}\sum_{t=T-P+1}^{T} g_i(a_t) < v_i + \frac{r}{2}.$$

The proofs of Lemmata 1 and 2 can be found in Gossner (1995) (See Lemma 3.1 and 3.2 respectively).

A bound on the average expected stage payoff received during a punishment phase by player i in the benchmark game, if each punisher is effective with probability at least $1 - \alpha$, is:

$$V_i(\alpha,\beta) = (1-(I-1)\alpha)\max_{\Phi^i_\beta(a_1,\ldots,a_t)=I-\{i\}}\frac{1}{P}\sum_{t=T-P+1}^{T} g_i(a_t) + (I-1)\alpha.$$

By taking β according to 2, then α small enough, we now fix α and β such that $V_i(\alpha,\beta) < v_i + r$. This ensures that

$$V_i(\alpha,\beta) < x_i - 2r \tag{6.1}$$

for every i.

6.4. **Definition of the incomplete code.** The incomplete code is parametrized by P and R, where P is the duration (same for all players i) of a punishment phase $\mathrm{P}(i)$ against player i, and R is the duration (equal for all sets J of effective punishers) of a reward phase $\mathrm{R}(J)$ that follows 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 rate of $\frac{1}{\sqrt{T}}$.

Lemma 3. *Let $y \in F$. 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 $\tilde{a}$ of actions implements y.

Let $\tilde{a} = (a_t)_t$ be a sequence of actions that implements x. We can, as for example in the proof of Lemma 2 in Fudenberg and Maskin (1991), choose $\tilde{a}$ such that, for some constant $B \geq 0$, the average payoff over any B consecutive stages along $\tilde{a}$ is $\frac{r}{2}$-close to x. For $J \subseteq I$, let $\tilde{a}^J = (a_t^J)_t$ be a sequence of actions that implements 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 punishers given by Φ^i_β;
- R(J) Play the sequence $\tilde{a}^J$, in reverse order[2]: $(a^J_R, \ldots, a^J_1)$. Then return to MP.

6.5. **Condition for effective punishments.** We show that if R is large enough and P is such that any punisher has a strategy that ensures to be effective with probability at least $1 - \frac{\alpha}{2}$, then in any sequential completion, every punisher is effective with probability at least $1 - \alpha$.

[2]The sequences $\tilde{a}^J$ are defined such that the average payoffs converge to the target payoffs. Playing actions in reverse order ensure that, from any period on before the last ones, the average payoff of the remaining periods is close to the target.

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

Definition 10. *Punishments are* (ε, M)-effective *if there exists* δ_0 *such that, for any* (ε, M)*-version* G_∞ *of* $\hat{G}_\infty$*, for every completion* σ *of* s *in* G_δ*, and every history* h_T *that ends with a deviation of player* i*,*

$$P_{\sigma,h_T}(\Phi^i_\beta(h_{T+P}) \neq I-\{i\}) \leq (I-1)\alpha.$$

The next lemma shows that punishments are (ε, M)-effective under two conditions. The first states that the length of a punishment period must be long-enough so that playing the min max repeatedly ensures each punisher a large enough probability (at least $1-\alpha/2$) of being an effective punisher. The second states that the length R of a reward phase is large enough to provide sufficient incentives for each punisher to be effective with a large probability.

Lemma 4. *If*

(1) $P \geq P_0(\frac{\alpha}{2}, \beta)$

(2)
$$\frac{1}{R}\left(M + K\sqrt{R} + P(1+\varepsilon)\right) \leq \frac{\alpha r}{2} - \varepsilon,$$

then punishments are (ε, M)*-effective.*

A consequence of equation 6.1 is that, if punishments are (ε, M)-effective, then for δ sufficiently large, for any (ε, M)-version G_∞ of $\hat{G}_\infty$, and for every completion σ of s in G_δ, the expected average payoff to player i during any punishment phase $P(i)$ is less than $x_i - 2r$.

6.6. **Robustness of the incomplete code.** Remember that for x in the interior of $F \cap IR$, the parameters r, α and β 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 (ε, M)-robust.

Lemma 5. *If punishments are* (ε, M)*-effective and*

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

then the incomplete code s *is* (ε, M)*-robust.*

Lemma 4 imposes a lower bound on the length of a reward phase R. Long reward phases are needed to provide each punisher with enough incentives to being an effective punisher during any punishment phase.

On the other hand, Lemma 5 imposes an upper bound on the length of a reward phase R. This comes from the fact that payoffs in reward phases are only equal to their targets up to an approximation. If the reward phases were to last too long, a player who is not currently rewarded could have incentives to deviate, and trigger a new punishment phase, in the hope of receiving a better payoff in the reward phase that will follow.

The following Lemma states that both conditions on the lengths of phases are compatible, thus concluding the proof of Theorem 2.

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

References

Abreu, D. (1988): "On the Theory of Infinitely Repeated Games with Discounting," *Econometrica*, 56, 383–396.

Abreu, D., D. Pearce, and E. Stacchetti (1990): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–1063.

Abreu, D. and A. Rubinstein (1988): "The structure of Nash equilibrium in repeated games with finite automata," *Econometrica*, 56, 1259–1281.

Aumann, R. J. and L. S. Shapley (1994): "Long-term competition—A game theoretic analysis," in *Essays on game theory*, ed. by N. Megiddo, New York: Springer-Verlag, 1–15.

Aumann, R. J. and S. Sorin (1989): "Cooperation and bounded recall," *Games and Economic Behavior*, 1, 5–39.

Ben Porath, E. (1993): "Repeated games with finite automata," *Journal of Economic Theory*, 59, 17–32.

Bhaskar, V., G. Mailath, and S. Morris (2008): "Purification in the Infinitely-Repeated Prisoners' Dilemma," *Review of Economic Dynamics*, 11, 515–528.

Blackwell, D. (1965): "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 36, 226–235.

Cavounidis, C., S. Ghosh, J. Hörner, E. Solan, and S. Takahashi (2018): "Blackwell Equilibria," draft.

Chassang, S. and S. Takahashi (2011): "Robustness to incomplete information in repeated games," *Theoretical Economics*, 6, 49–93.

Dutta, P. (1995): "A Folk Theorem for stochastic games," *Journal of Economic Theory*, 66, 1–32.

Friedman, J. (1971): "A noncooperative equilibrium for supergames," *Review of Economic Studies*, 38, 1–12.

Fudenberg, D. and E. Maskin (1986): "The folk theorem in repeated games with discounting or with incomplete information," *Econometrica*, 54, 533–554.

——— (1991): "On the dispensability of public randomization in discounted repeated games," *Journal of Economic Theory*, 53, 428–438.

Fudenberg, D. and Y. Yamamoto (2010): "The Folk Theorem for irreducible stochastic games with imperfect public monitoring," *Journal of Economic Theory*, 146, 1664–1683.

Gossner, O. (1995): "The folk theorem for finitely repeated games with mixed strategies," *International Journal of Game Theory*, 24, 95–107.

Hörner, J., T. Sugaya, S. Takahashi, and N. Vieille (2011): "Recursive methods in discounted stochastic games: an algorithm for $\delta \to 1$ and a Folk Theorem," *Econometrica*, 79, 1277–1318.

Mailath, G. and L. Samuelson (2006): *Repeated Games and Reputations Long-Run Relationships*, Oxford University Press.

Mertens, J.-F., S. Sorin, and S. Zamir (2015): *Repeated games*, Cambridge University Press.

Neyman, A. (1998): "Finitely repeated games with finite automata," *Mathematics of Operations Research*, 23, 513–552.

Rubinstein, A. (1979): "Equilibrium in Supergames with the Overtaking Criterion," *Journal of Economic Theory*, 21, 1–9.

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 $y \in \mathrm{co}X$. There exists a sequence $(x_t)_t$ 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'.

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

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

$$<x - y, z> < 0$$

where $<.,.>$ 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

$$<x - y, z> \leq -\rho.$$

We construct $(x_t)_t$ 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:

$$<x_{t+1} - y, \gamma_t - y> \leq -\rho \|y - \gamma_t\|_2$$

Let $b_t = t\|y - \gamma_t\|_2^2$. For every T, either $b_T = 0$, or

$$\begin{aligned} b_{T+1}^2 &= \|\sum_{t=1}^{T+1} x_t - (T+1)Y\|_2^2 \\ &= b_T^2 + 2<\sum_{t=1}^{T} a_t - Ty, x_{T+1} - y> + \|a_{T+1} - y\|_2^2 \\ &\leq b_T^2 - 2\rho b_T + M^2 \end{aligned}$$

where $M = \max_{x \in X'} \|y - x\|_2$. Note that $<x - y, z> \geq -\|x - y\|_2$ 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:

$$\begin{aligned} b_{T+1}^2 &\leq \max_{0 \leq x \leq K} (x^2 - 2\rho x) + M^2 \\ &\leq \max\left\{K^2 - 2\rho K, 0\right\} + M^2 \\ &\leq \max\left\{K^2 - M^2, M^2\right\} \\ &\leq K^2. \end{aligned}$$

□

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 us 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 δ_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

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

Appendix C. Proof of Lemma 4

Let δ_0 be obtained from Lemma 8 using $T_0 = P + R$ and $a = \frac{M}{1+\varepsilon}$. Let $\delta > \delta_0$, and σ be a completion of s in G_δ. Consider a history h_T ending with a deviation of player i. We define

$$x_t = \frac{\mathrm{E}_{\sigma_{-i}, m_i^j}\, g_j(h_{T+t}) - \mathrm{E}_\sigma\, g_j(h_{T+t})}{1 + \varepsilon}$$

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 $T+P+R$:

$$\begin{aligned}
\sum_{t=1}^{P} \mathrm{E}_{\sigma_{-i}, m_i^j}\, g_j(h_{T+t}) &\geq -P(1+\varepsilon) \\
\sum_{t=1}^{P} \mathrm{E}_\sigma\, g_j(h_{T+t}) &\leq P(1+\varepsilon) \\
\sum_{t=P+1}^{P+R} \mathrm{E}_{\sigma_{-i}, m_i^j}\, g_j(h_{T+t}) &\geq R((1-\frac{\alpha}{2})(x_i + r) + \frac{\alpha}{2}(x_i - r)) - K\sqrt{R} - R\varepsilon \\
\sum_{t=P+1}^{P+R} \mathrm{E}_\sigma\, g_j(h_{T+t}) &\leq R((1-p)(x_i + r) + p(x_i - r)) + K\sqrt{R} + R\varepsilon.
\end{aligned}$$

Now we derive, using condition (2) of Lemma 4:

$$\begin{aligned}(1+\varepsilon)\sum_{t=1}^{P+R} x_t &\geq -2\left(P(1+\varepsilon)+K\sqrt{R}+R\varepsilon\right)+2Rr(p-\frac{\alpha}{2})\\ &\geq -Rr\alpha+2M+2Rr(p-\frac{\alpha}{2})\\ &\geq 2M+Rr(2p-\alpha).\end{aligned}$$

Since G has M-influence from 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 $\sum_{t=1}^{P+R}\delta^t x_t > 0$, hence, following h_T, playing m_j^i 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 δ_0' be obtained from Lemma 8 using $T_0 = P+R+1$ and $a = \frac{Pr}{2(1+\varepsilon)}$. Let δ_0'' be such that, for any (ε, M)-version G_δ of $\hat{G}$, for every completion σ of s in G_δ, and every history h_T that ends with a deviation of player i,

$$P_{\sigma,h_T}(\Phi^i_{\beta,j}(a_{T+1},\ldots,a_{T+P}) \neq I{-}\{i\}) \leq 1-(I-1)\alpha.$$

Also, let $\delta_0 = \max\{\delta_0', \delta_0''\}$. Consider $\delta \geq \delta_0$, a (ε, M)-version G of $\hat{G}$, and a completion σ of s in G_δ.

We need to show that deviations after a history h_T on the Main path or on some Reward Path cannot be profitable. Let σ_i' be a strategy of player i that deviates after h_T and reverts to σ_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

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

We have:

$$\begin{aligned}
\sum_{t=1}^{R'} \mathrm{E}_{\sigma,h+T}\, g_j(h_{T+t}) &\geq R'\,(x_i - r - \varepsilon) - K\sqrt{R'} \\
\sum_{t=R'+1}^{P+R+1} \mathrm{E}_{\sigma,h+T}\, g_j(h_{T+t}) &\geq (P+R-R'+1)(x_i - \frac{r}{2} - \varepsilon) - B \\
\sum_{t=1}^{P+1} \mathrm{E}_{\sigma',h+T}\, g_j(h_{T+t}) &\leq 1 + P(V_i(\alpha,\beta)+\varepsilon) \\
&\leq 1 + P(x_i - 2r + \varepsilon) \\
\sum_{t=P+2}^{P+R+1} \mathrm{E}_{\sigma,h+T}\, g_j(h_{T+t}) &\leq R(x_i - r + \varepsilon) + K\sqrt{R}
\end{aligned}$$

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

$$\begin{aligned}
(1+\varepsilon)\sum_{t=1}^{P+R+1} x_t &\geq -\Big(K(\sqrt{R}+\sqrt{R'}) + 2\varepsilon(P+R+1) + 1 - x_i + B\Big) + \frac{3}{2}Pr \\
&\geq -\Big(2K\sqrt{R} + 2\varepsilon(P+R+1) + 2 + B\Big) + \frac{3}{2}Pr \\
&\geq \frac{1}{2}Pr + M.
\end{aligned}$$

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

$$\sum_{t=1}^{\infty} \delta^t \, \mathrm{E}_{\sigma,h_T} \, g(h_{T+t}) > \sum_{t=1}^{\infty} \delta^t \, \mathrm{E}_{\sigma',h_T} \, g(h_{T+t}).$$

Hence a deviation from σ 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 λ 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 λP. 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 (ε, M)-effective by Lemma 4, and from Lemma 5 the code is (ε, M)-robust.

CNRS - École Polytechnique, CREST and London School of Economics