Uncertain repeated games*

Ilia Krasikov Rohit Lamba

September 2023

Abstract

Multiple long run players play one amongst multiple possible stage games in each period. They observe and recall past play and are aware of the current stage game being played, but are uncertain about the future evolution of stage games. This setup is termed an *uncertain repeated game*. The solution concept requires that a subgame perfect equilibrium be played no matter what sequence of stage games realize. The feasible set of payoffs is then so large and complex that it is not obvious how to frame standard results such as the folk theorem, and further how to construct credible rewards and punishments that work irrespective of the future evolution of games. The main goal of the paper is to build such a language through two different perspectives— one in which the modeler has access to the true stochastic process but not the players and another in which there is simply maximal uncertainty; and then to construct credible dynamic incentives that work generally for uncertain repeated games. A complete characterization of equilibria is presented for large discount factors and various extensions to related models and results are discussed.

1 Introduction

Motivation. The canonical repeated games, and more generally stochastic games, model is instrumental in formalizing the extent of cooperation and conflict that can be credibly sustained in dynamic interactions. In doing so, it makes an arguably restrictive assumption that the exact same stage game is repeated over time or that one amongst several possible stage games is chosen through a commonly understood stochastic process. In this paper, we attempt to expand the scope of this canon along the dimension of robustness to the structure of future play: What is the extent

*Krasikov: Higher School of Economics University Moscow, krasikovis.main@gmail.com; Lamba: Pennsylvania State University, rlamba@psu.edu. We are indebted to Dilip Abreu, V. Bhaskar, Modibo Camara, Gabriel Carroll, Sylvain Chassang, Rohan Dutta, Yuhta Ishii, Elliot Lipnowski, Harry Pei, Antonio Penta, Wolfgang Pesendorfer, Eilon Solan, Maxwell Stinchcombe, Bruno Strulovici, Takuo Sugaya, Balázs Szentes, Caroline Thomas, Thomas Wiseman, Weijie Zhong, John Zhu, and seminar participants at University of Texas Austin, Stanford University, University of Kansas, Princeton University, London School of Economics, Barcelona Summer School in Economics, Essex Workshop on Dynamic Incentives, University of Chicago, Pennsylvania Economic Theory Conference and Stony Brook Conference on Game Theory for their comments and suggestions.

of cooperation and conflict that can be credibly and *robustly* sustained among multiple long-run players when they may not have a common or well-specified understanding of how games evolve in the future.

For example, in their classical study of price competition over a business cycle, Rotemberg and Saloner [1986] showed that forward-looking firms can price above marginal cost when demand shocks are i.i.d. and firms agree on the exact probabilities of booms and recessions. It is well documented that firms can efficiently collude or enter price wars under alternative models of demand shocks (see Haltiwanger and Harrington [1991] and Kandori [1991]) provided that it is commonly understood by all firms. However, in reality, their model of demand shocks might be misspecified; some firms might be better informed than others, dynamically inconsistent, or even uncertain about the future. The natural question then arises: to what extent can collusion and conflict be sustained without making specific assumptions on how the state of the market evolves?

Model. To that end, we consider the following setting, which resembles the standard model of stochastic games with perfect monitoring. A finite set of players with a common discount factor interact repeatedly and play one among finitely many possible stage games in each period. At the beginning of every period, they are informed of the stage game they are about to play, which is chosen independently of past actions, but they do not know the future realization of stage games.[1] All past actions and realized games are observable.

Unlike the standard model of stochastic games, this setup is bereft of any information on how future games evolve over time. Instead, the model allows for an arbitrary notion of how stage games are drawn, in that sense incorporating maximal uncertainty towards the future, we thus call this set-up an *uncertain repeated game*. To discipline the model, we use the following ex-post notion of equilibrium: a strategy profile is said to be an *ex-post perfect equilibrium* (Carroll [2021]) if it constitutes a subgame perfect equilibrium "pointwise", that is, for any possible realizations for future stage games.

Ex-post equilibrium is a fairly demanding criterion, and so any cooperation (or conflict) attained under its guise is robust in its predictive power. It is robust to any misspecifications or disagreement that the players or the outside analyst may have about the underlying stochastic process, and it naturally satisfies a non-regret condition for all players–— no matter the future realization of stage games, no player will individually regret not having deviated at any point.[2]

[1] The fact that the players' actions do not affect the evolution of stage games is common knowledge. Extension to endogenous transitions is possible and discussed in the paper, but at first pass, the attempt is to understand the complexity introduced by exogenous uncertainty.

[2] The idea of ex-post perfect equilibrium is not necessarily steeped in a (direct) model of individual maximization. Instead, ex-post perfect equilibria demand that each player's strategy best responds to the other players' strategies in hindsight. In that sense, as we will argue later, it provides robust 'lower bound' in predicting behavior in stochastic games.

Two perspectives. There are two natural ways to characterize the predictive content of the proposed model, depending on the analyst or the modeler's information. First, suppose that the modeler has access to the true stochastic process that governs the evolution of stage games and that the process is first-order Markov, as standard in stochastic games. This allows us to compute expected continuation payoffs associated with any particular strategy at any point in the game tree. Then, we ask the canonical question that has occupied the repeated games literature for some time: when is an expected payoff vector achievable on-path in an equilibrium? However, in a departure from the literature, the demand for equilibrium is stronger; in particular, we demand it to be ex-post perfect.

One way to think about this problem is to draw a parallel with robust mechanism design, in the sense of Bergemann and Morris [2005]. The designer wants to maximize expected profits, say, but wants to sustain the equilibrium in dominant strategies. Why? For any of the many robustness considerations. It could be that the players do not understand fully how values are distributed, or they may have heterogeneous priors, or they may not be able to do Bayesian updating correctly, etc. Similarly, here, the modeler (or designer) can be interested in sustaining a particular expected payoff on-path but is not sure how well the players understand the draw of stage games, and hence seeks to achieve this expected payoff under the ex-post equilibrium notion.[3]

The second, more abstract and ambitious, approach is to get rid of any references to stochastic processes driving stage games altogether. It could be that the modeler, too, is unsure about the stochastic process governing the evolution of stage games, or we could simply be interested in understanding if a particular sequence of actions that are observed in a data set can ever be rationalized under some realization of stage games. Unfortunately, the standard approach of studying repeated and stochastic games in the payoff space is no longer tractable because the set of feasible payoffs is infinite-dimensional, one for each possible realization of stage games. To overcome this challenge, we revert to the classical approach of studying outcomes (see Abreu [1988]) that describe sequences of action profiles as a function of realized stage games. For example, in the model of Bertrand competition studied in Rotemberg and Saloner [1986], charging the monopoly price in recessions and half of it in booms is an outcome. Here, we seek to characterize the set of outcomes that can arise on-path in some ex-post perfect equilibrium.

In the remarkable recent work, Carroll [2021] takes the second perspective outlined above, but for the case of one long-run player (in the sense of Fudenberg, Kreps, and Maskin [1990], wherein only one player has a positive discount factor). This paper builds on Carroll's elegant analysis. While the broad motivations mentioned above are common to both papers, the first perspective

[3] This question is related to Gershkov, Goeree, Kushnir, Moldovanu, and Shi [2013], which showed the equivalence of Bayesian and dominant strategy incentive compatibility in the standard auction setting with independent private values.

above is novel to this paper; and more generally, we submit that a fuller picture of the limits to cooperation and conflict introduced by uncertainty, and associated costs of robustness, can only be rightfully addressed by studying games with many long-run players. Most applications of the theory of repeated games, e.g., collusion, price wars, risk-sharing, sustainable policy plans, relational contracts and more, necessitate multiple long-run players.[4]

The main results. These two perspectives we just laid out are fairly broad. In this paper, we focus on the case of little discounting and aim to deliver robust folk theorems. Conceptually, these folk theorems can be posed as follows: there is no cost of seeking robustness to the exact understanding of the evolution of future stage games— first in terms of expected payoffs when the modeler has access to the true stochastic process, and second in terms of outcomes when all parties are taken to be maximally uncertain— for large values of the discount factor.

Onto the results. Recollect, an expected payoff is feasible if it can be attained under some well-defined strategy, and it is individually rational if it gives each player at least her minimum threshold value that can be guaranteed when best responding to other players. The first main result shows that every feasible and individually rational expected payoff can be attained by patient players in an ex-post perfect equilibrium, that is, independent of how information is modeled, if each possible stage game satisfies the usual interiority assumption.[5] Strikingly, the result says that the folk theorem is valid, *irrespective* of the extent of misspecficiation, disagreement, asymmetry, time inconsistency, and uncertainty of information on how future stage games are drawn.

To prove the result, we introduce a novel budget mechanism that calibrates the subgame perfect equilibrium strategy profile to satisfy the ex-post criterion. The ex-post criterion picks the tightest incentive constraint, pointwise across all realizations of stage games, even though this tightest constraint may be relevant with small probability under the true stochastic process. The budget mechanism uses this information: it keeps track of players' cumulative incentives to deviate and alters their actions from the benchmark strategy profile if this cumulative count exceeds a pre-determined budget cap. The cap is not binding for frequent stage game realizations, since the original strategy profile is subgame perfect, so the count is within the desired bound in expectation. However, if the tightest incentive constraint occurs for infrequent realizations of stage games, this cap can be violated for the ex-post criterion, and so actions of the players are modified to accommodate this, but expected payoffs are approximately unaltered because of the rarity of these stage game realizations. The technical challenge is in ensuring this budget mechanism blends with the

[4]The multiple long-run player model poses non-trivial technical challenges, some of which are highlighted by Carroll [2021].

[5]Interiority means that there must be some slack in the payoff space to punish and reward players for deviating and participating in others' punishments, respectively. This condition is common in repeated games (Fudenberg and Maskin [1986], Fudenberg et al. [1994]) and stochastic games (Fudenberg and Yamamoto [2011], Hörner, Sugaya, Takahashi, and Vieille [2011]).

standard reward and punishment techniques from repeated games, on-and-off-path.

The second main result establishes the folk theorem for the uncertain repeated game in the language of outcomes for the payoff set is infinite-dimensional and no longer tractable.[6] As is standard, every outcome that is played on-path in some XPE must give each player at least their discounted minmax payoffs irrespective of discounting. Our folk theorem for outcomes establishes the converse statement by showing that patient players can be mixmaxed in all future stage games simultaneously—in the language of Abreu [1988], *a universal penal code exists asymptotically*.[7] As a result, outcomes that deliver payoffs above the discounted minmax are precisely those that can arise on-path in some XPE when players are patient enough.

The folk theorem for outcomes requires another assumption, in addition to the statewise interiority described above. This extra assumption limits how bad it can be for other players to punish a deviator across all stage games.[8] This is a natural condition for without it a universal penal code, that achieves the dynamic minmax is unlikely to exist.

To prove the theorem, we first normalize each stage game by a fixed payoff vector. This allows dynamic incentives to be comparable across various stage game realizations. Second, we restrict attention to strategy profiles in which both on-and-off path continuation payoffs are independent of the future realization of normalized stage games. This second step reduces the dimensionality of the problem to a recursion that can be studied using standard repeated games techniques. However, it must be emphasized that it also imposes a substantial restriction on the class of dynamic incentives under consideration. The nontriviality here rests in the claim that this restricted class of strategy profiles is enough to achieve the dynamic minmax payoff by varying the normalizing vector as $\delta \to 1$, giving us the folk theorem for outcomes.

Finally, to make a direct comparison to Carroll [2021], we show how the technique works for the special class of strongly symmetric ex-post perfect equilibria when focusing on symmetric games. The search for strongly symmetric equilibrium reduces the game to a decision problem, and so a universal penal code can be constructed simultaneously across all realizations of stage games. The restriction on continuation payoffs to be independent of future stage games, thus, does not bite, and a complete characterization of the equilibrium outcomes for all values of the discount factor is provided. Carroll [2021] studies uncertain repeated games with one-long run player; ex-post perfect equilibria there can be formally mapped to strongly symmetric ex-post perfect equilibria in symmetric games with multiple long-run players.

Extensions. In the last part of the paper, we explore the relaxation of the key modeling as-

[6]Recollect an outcome is a sequences of action profiles as a function of realized stage games.

[7]In fact, Carroll [2021] illustrates the difficulty of establishing such a universal penal code for multiple long-run players through a two-period example.

[8]It is satisfied in a large class of games: symmetric games, games with money burning, games which have joint minmax, and more.

sumption in the main results, while keeping ex-post perfectness as the equilibrium criterion.

For two of these, we offer concrete answers through examples. First, we assumed transitions between states are exogenous. Through a simple example, we illustrate how the basic insights go through as long as the dependence of the stochastic process on actions is limited. Second, we considered games in which players observe the current state, say are we in the high or low demand regime. Through a simple example again, we propose a tractable way to incorporate asymmetric observability of the current stage game into the model and show that our budgeting mechanism continues to work.

Next, the monitoring structure here is assumed to be perfect, that is all players observe all other players' actions. The basic construction of the folk theorem here, we conjecture, will go through to an imperfect monitoring world with appropriate modeling assumptions and proof adjustments. It is also natural to ask how to go about characterizing the set of equilibria for a fixed value of the discount factor. This question has been addressed in our companion paper, Krasikov and Lamba [2023], which pins down a subset of ex-post perfect perfect equilibria that through the construction of an outer bound is shown to be tight enough for many of classes of dynamic games.

Finally, we explain that ex-post perfect equilibrium is naturally connected to subgame perfectness under dynamic variational preferences as defined in Maccheroni, Marinacci, and Rustichini [2006a,b] —a strategy profile is ex-post perfect if and only if it is subgame perfect for all (potentially dynamic) ambiguity indices. This explains how ex-post perfect equilibrium provides a lower bound in predictions for a large class dynamic ambiguity averse preferences.

Related literature. This paper is a contribution to the theory of both dynamic decision problems and games, and robust approaches to games of incomplete information. In addition to Carroll [2021], the only other paper we are aware of that rests within the rubric of uncertain repeated games is Kostadinov [2023]. It looks at regret minimization as opposed to ex-post perfectness as the equilibrium criterion and finds results much more permissive than ours or Carroll [2021].

The concept of ex-post perfect equilibrium has been studied in other dynamic models before. Most closely, Fudenberg and Yamamoto [2010] studies repeated games (with a single stage game) where the monitoring structure is fixed but unknown and the equilibrium notion is a version of ex-post perfect; it then establishes a folk theorem for this setup. Beyond repeated and stochastic games, notions of ex-post perfect equilibrium have also been studied in models of dynamic mechanism design (Mirrokni, Leme, Tang, and Zuo [2020]) and in sequential voting (Kleiner and Moldovanu [2017]). The criterion is also related to the belief-free equilibrium, which is used to gain tractability in repeated games with private monitoring (Ely, Hörner, and Olszewski [2005] and Hörner and Lovo [2009]).

Other related papers seek robustness with respect to information in dynamic decision problems and games: Chassang [2013] studies dynamic contracts with limited liability constraints that are robust to underlying stochastic process that maps effort to output; it shows linear contracts do approximately well. Penta [2015] extends some of the ideas of robust mechanism design initiated by Bergemann and Morris [2005] from static to dynamic environments. Libgober and Mu [2023] examines Coasian dynamics under informational robustness— seller does not know the process through which buyer learns her valuation and has ambiguous preferences over this uncertainty; it characterizes the optimal price path in this setting. de Oliveira and Lamba [2023] studies a model of dynamic choice where an analyst tries to rationalize a sequence of choices without having access to the dynamic information seen by the decision maker, and establishes a duality result to pin down these actions.

This paper builds on the classical literature on stochastic games (Shapley [1953] and Solan and Vieille [2015]), and more recent work on folk theorems for stochastic games (Fudenberg and Yamamoto [2011] and Hörner, Sugaya, Takahashi, and Vieille [2011]). Most of the papers in the literature focus on imperfect monitoring of past actions assuming perfect observability of stage games and common knowledge of the stochastic process that drives them. The notable exception is Yamamoto [2019] which studies stochastic games with imperfectly observable stage games. In this paper, the stochastic process of stage games is still commonly known, therefore players' belief about the current stage game can be regarded as a state variable, and the usual notion of subgame perfectness continues to apply.

2 Model and notations

2.1 Primitives

Consider players $i \in N := \{1, \ldots, n\}$ interacting in discrete time $t = 0, 1, \ldots$.

Every period the players face one stage game chosen from a finite set. Formally, a **stage game** is a pair (A, u), where $A = \prod_{i=1}^{n} A_i$ is a finite set of action profiles and $u : A \to \mathbb{R}^n$ is a payoff function. Let $\mathscr{U} := \text{Conv}(u(A))$ be the convex hull of payoffs in the stage game (A, u), and assume, w.l.o.g.,

$$\underline{r}_i := \min_{a \in A} r_i(a) = 0, \text{ where } r_i(a) := \max_{\widetilde{a}_i \in A_i} u(\widetilde{a}_i, a_{-i}) \ \ \forall i \in N. \tag{1}$$

This normalizes the minmax payoff in a stage game for each player to zero. Further, denote by $d(a) := r(a) - u(a)$ the best static deviation gain from a fixed action profile, and extend u to ΔA by linearity, i.e., $u(\beta) := \sum_{a \in A} u(a)\beta(a)$.

There will be multiple possible stage games—the set of possible games Θ is finite and commonly known. Each element $\theta \in \Theta$ specifies a pair $(A(\theta), u(\cdot|\theta))$. Then, an **uncertain repeated**

game is a tuple $(\Theta, (A(\theta), u(\cdot|\theta))_{\theta\in\Theta}, \delta)$, where $\delta < 1$ is the common discount factor. We take $(\Theta, (A(\theta), u(\cdot|\theta))_{\theta\in\Theta})$ as the primitive of the model, and fix it throughout. There is also perfect recall, so whatever is observed by the players is always remembered.

Three additional assumptions are imposed: (i) *current stage game is common knowledge*, i.e., at the beginning of each period, all the players are informed of the stage game, $\theta \in \Theta$, they are about to play; (ii) *perfect monitoring*, i.e., observance of opponents' actions at the end of each period; and (iii) *exogenous transitions*, i.e, action-independent switching between stage games. Potential extensions relaxing these are discussed in Section 6.

Now, we seek a theory of dynamic interactions here whose predictions are robust to how stage games are drawn from the set Θ. To write down payoffs and state results, we will take two perspectives.

In the first, we will allow the modeler to take a call on the objective uncertainty, and set a true transition function for then a well-defined **stochastic game**. Even though the modeler has access to an objective transition function, the players' incentives will still be required to satisfy an ex-post criterion. As is standard in the literature, the true stochastic process is taken to be first-order Markov. We denote it by $\pi \in \mathbb{R}_{++}^{\Theta\times\Theta}$, that is a right-stochastic matrix with the interpretation that $\pi(\theta^t, \theta) = \mathbb{P}(\theta^{t+1} = \theta|\theta^t)$. To simplify exposition, we assume that the initial stage game is drawn from the unique stationary distribution $\mu(\pi) \in \Delta\Theta$.[9]

The second approach will be completely bereft of any stochasticity in the modeling of primitives. Specifically, all statements will be made in terms of a dynamic game fixed by a realization of stage games, $e = (\theta^0, \theta^1, \ldots) \in \Theta^\infty$, which will be termed an **environment**. Here, there is no notion of "expected payoffs", so all results must be stated in terms of outcomes describing sequences of on-path action profiles and associated ex-post payoffs as a function of an environment.

2.2 Strategies and payoffs

At date t the players are symmetrically informed about a public history h^t that includes sequences of past action profiles $(a^0, ..., a^{t-1})$, stage games $(\theta^0, ..., \theta^t)$ and sunspots $(\omega^0, \ldots, \omega^t)$, which are taken to be i.i.d. uniform random variables. This makes our setup analogous to a standard repeated game with perfect monitoring, perfect recall, and access to public correlation device, á la Fudenberg and Maskin [1986], and Mailath and Samuelson [2006] (see Chapters 2 to 6) except that the players face potentially different stage games.

A **(pure) strategy** σ_i for player $i \in N$ prescribes for each history h^t, a pure action $\sigma_i(h^t) \in A_i(\theta^t)$.[10] Given an environment e, the **ex-post payoff** under strategy profile $\sigma := (\sigma_1, \ldots, \sigma_n)$ is

[9]Existence and uniqueness of the stationary distribution follows from Perron–Frobenius Theorem.

[10]The analysis of mixed strategies is almost identical when randomizations chosen by the players is observable

defined inductively as

$$U^{\sigma}(h^t|e) = (1-\delta)u(\sigma(h^t)|\theta^t) + \delta\mathbb{E}_t[U^{\sigma}(\underbrace{h^t, \sigma(h^t), \theta^{t+1}, \omega^{t+1}}_{=h^{t+1}}|e)], \tag{2}$$

where h^t is compatible with e and the expectation $\mathbb{E}_t[\cdot]$ is taken over the $(t+1)$-th sunspot. Slightly abusing notation, denote the payoff at the outset under σ by $U^{\sigma}(e)$. If the stage games evolve according to the transition matrix π, then we can similarly compute the **expected payoff** under σ as

$$U^{\sigma}(h^t|\pi) = (1-\delta)u(\sigma(h^t)|\theta^t) + \delta \sum_{\theta^{t+1}\in\Theta} \mathbb{E}_t[U^{\sigma}(\underbrace{h^t, \sigma(h^t), \theta^{t+1}, \omega^{t+1}}_{=h^{t+1}}|\pi)]\pi(\theta^t, \theta^{t+1}), \tag{3}$$

where again the expectation $\mathbb{E}_t[\cdot]$ is taken over the $(t+1)$-th sunspot. As before, let $U^{\sigma}(\pi)$ be the expected payoff at the outset.

2.3 Equilibrium notions

For completely specified models of dynamic and stochastic games, we use the standard equilibrium notion, namely **Subgame Perfect Equilibrium** (SPE).

Definition 1. *A strategy profile σ is SPE of dynamic game e if for each history h^t that is compatible with e and every alternative strategy profile $\widetilde{\sigma}$,*

$$U^{\sigma}(h^t|e) \geq U^{(\widetilde{\sigma}_i, \sigma_{-i})}(h^t|e) \ \ \forall i \in N; \tag{4}$$

similarly, it is SPE of stochastic game π if for each history h^t and every alternative strategy profile $\widetilde{\sigma}$,

$$U^{\sigma}(h^t|\pi) \geq U^{(\widetilde{\sigma}_i, \sigma_{-i})}(h^t|\pi) \ \ \forall i \in N. \tag{5}$$

The solution concept we shall use for an uncertain repeated game is **Ex-post Perfect Equilibrium** (XPE).

Definition 2. *A strategy profile σ is XPE if it is an SPE of every dynamic game e.*

Two observations are immediate. First, if a strategy profile is XPE, then it is SPE for every stochastic game, but the converse is not true. Moreover, since the one-shot deviation principle holds for SPE of every fixed dynamic game, it also holds for XPE.

Our definition of Ex-post perfect equilibrium follows Carroll [2021], generalizing its functionality from uncertain repeated games with one long-run player to multiple long-run players. It is a natural but demanding concept. It is based on the idea of no regret — for any possible realization of stage games, no player regrets having not deviated. It is also robust to heterogeneous beliefs, misspecifications in modeling information, and potential time inconsistencies in the processing of information by players.

We will study XPE from two complementary perspectives. First, when the modeler is equipped with objective probability on the evolution of stage games and asks the question of when can a certain expected payoff be attained through incentives that are robust to the fine details of the stochastic evolution. This parallels the question from mechanism design — when can a Bayesian objective be attained under ex-post incentive compatibility. And, second, we also explore on-path behavior — what sequences of action profiles can be played on-path irrespective of the specific realization of future stage games? This exercise is useful for understanding the extent of cooperation (or conflict) that can be sustained uniformly across various realizations of dynamic games.

2.4 Comparing XPE and SPE

Before stepping into the results, it is useful to pause and understand the workings on XPE and its contrast from SPE.

To begin, suppose that there are only two periods. The first-period stage game is fixed but there are two possible second-period stage games, say L and R. Here, a deterministic strategy profile prescribes an action profile in period 1, s, and an action profile in period 2, $s_\theta(a)$, for each second-period stage game θ as a function of the action profile in period 1, a. See Figure 1 below for an illustration.

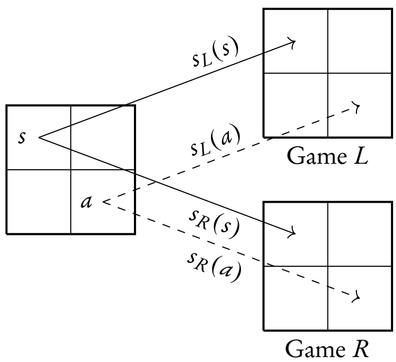


Figure 1: Illustration of a strategic situation and a strategy profile.

Suppose for a moment that both players agree on the probabilities of the second-period stage game. Then, as is standard, a strategy profile (s, s_L, s_R) is subgame perfect if and only if s_θ always selects a static Nash in game θ and for each player,

$$\text{static gain from deviation in period 1} \leq \sum_{\theta=L,R} \mathbb{P}(\theta) \Big(\text{loss from punishment in } \theta\Big). \tag{6}$$

As can be seen from the incentive constraint in (6), an SPE strategy profile must be designed to deter deviations in expectation. In contrast, the stronger ex-post criterion demands deviations to be deterred simultaneously in all second-period stage games replacing (6) with

$$\text{static gain from deviation in period 1} \leq \min_{\theta=L,R} \Big(\text{loss from punishment in } \theta\Big). \tag{7}$$

As a result, no player has a profitable deviation irrespective of her (potentially subjective) probabilistic assessment of L and R.

XPE shares the logic of SPE with a commonly known stochastic process by disallowing profitable unilateral deviations. The core difference between these two criteria, as can be seen from Eq. (6) and Eq. (7), lies in how dynamic losses from punishment are perceived by the players, i.e., $\min_\theta$ versus $\sum_\theta \mathbb{P}(\theta)$. Of course, XPE is a more demanding notion, but it has an appealing feature of robustness to potential misspecifications of the stochastic process and/or additional information the players might have about the second-period stage game.

Furthermore, XPE is compatible with a much larger class of dynamic preferences, namely dynamic variational preferences of Maccheroni et al. [2006a,b]. Such preferences are represented by a convex cost function $C\left(\mathbb{P}(L), \mathbb{P}(R)\right)$, and a player ranks strategy profiles according to the criterion:

$$\inf_{(\mathbb{P}(L),\mathbb{P}(R))} \left(\sum_{\theta \in \Theta} \mathbb{P}(\theta) \Big(\text{discounted utility if } \theta \text{ realizes}\Big) + C\left(\mathbb{P}(L), \mathbb{P}(R)\right) \right). \tag{8}$$

It is not hard to see that an ex-post perfect strategy profile is necessarily subgame perfect when the players have dynamic variational preferences. Conversely, a strategy profile that is subgame perfect for *all* such preferences satisfies the ex-post criterion.[11]

3 Robustness of expected payoffs in stochastic games

In this section, we will ask the question of robustness of expected payoffs. There is an objective first-order Markov process π attached to the evolution of the stage games; however, the equilibrium criterion is ex-post perfectness. To state the result, we will make use of the following definitions of expected equilibrium payoffs.

Definition 3. *$w \in \mathbb{R}^n$ is an expected **equilibrium payoff** of stochastic game π if there exists an SPE σ of π satisfying $U^\sigma(\pi) = w$; and $w \in \mathbb{R}^n$ is an expected **limit equilibrium payoff** if there exists a sequence $(w(\delta))_{\delta<1}$ converging to w such that for each δ, $w(\delta)$ is an equilibrium payoff of π when the players discount the future with δ.*

[11] See Section 6.5 and Supplementary Appendix for details.

Definition 4. *$w \in \mathbb{R}^n$ is an expected* ***robust equilibrium payoff*** *of stochastic game π if there exists an XPE σ of π satisfying $U^{\sigma}(\pi) = w$; and $w \in \mathbb{R}^n$ is an expected* ***robust limit equilibrium payoff*** *if there exists a sequence $(w(\delta))_{\delta<1}$ converging to w such that for each δ, $w(\delta)$ is a robust equilibrium payoff of π when the players discount the future with δ.*[12]

The main result of this section will deliver a complete characterization of robust limit equilibrium payoffs; in particular, a folk theorem type of result. It shows that the under a mild interiority assumption, the set of robust equilibrium payoffs coincides with the set of "feasible and individually rational" expected payoffs, which are expected payoffs that can be obtained by some strategy profiles and guarantee each player at least her expected minmax payoff, normalized here to 0. So, robustness can be guaranteed in the limit as $\delta \to 1$ at no costs — all "feasible and individually rational" can be obtained not only in SPE but also in XPE. To fix ideas, we start with an example.

Example 1. *There are two equally likely stage games, L and R, and two players. Table 1 describes both stage games and depicts their respective feasible payoffs, $\mathscr{U}(L)$ and $\mathscr{U}(R)$. It also illustrates the set of feasible payoffs in the stochastic game when (θ^t) are i.i.d. uniform, that is $\frac{\mathscr{U}(L)+\mathscr{U}(R)}{2}$. In each game, the minmax payoff vector $(0,0)$ is a static Nash equilibrium.*

	x_2	y_2
x_1	$0^\star, 0^\star$	$-1, -1$
y_1	$-1, {}^7\!/_2{}^\star$	${}^1\!/_2{}^\star, {}^1\!/_2$

Game L

	x_2	y_2
x_1	$0^\star, 0^\star$	${}^7\!/_2{}^\star, -1$
y_1	$-1, -1$	${}^1\!/_2, {}^1\!/_2{}^\star$

Game R

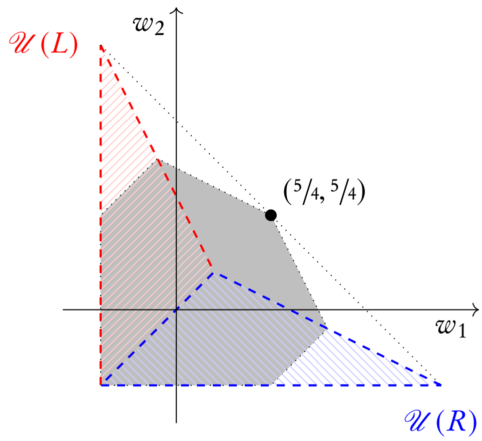


Table 1: Uncertain repeated game for Example 1: "$\star$" indicates the static best-response, $\mathscr{U}(L)/\mathscr{U}(R)$ is a red/blue triangle and $\frac{\mathscr{U}(L)+\mathscr{U}(R)}{2}$ is a gray polygon.

It is easy to pick visually that the maximal symmetric expected payoff in the "average" game, $(\frac{5}{4}, \frac{5}{4})$, can be attained by standard grim trigger type of strategies: play (y_1, x_2) in game L and (x_1, y_2) in game R until no one has deviated. After any deviation, the players switch to playing the static Nash (x_1, x_2) in each game. Clearly, this strategy profile is SPE if and only if $-(1-\delta) + \delta\frac{5}{4} \geq 0$ that gives $\delta \geq \frac{4}{9}$. So, it follows that $(\frac{5}{4}, \frac{5}{4})$ is a limit equilibrium payoff.

[12] The set of robust limit equilibrium payoffs is closed. To see it take a sequence (w^k) in this set that converges to some point w, and, for each k, let $(w^k(\delta))_{\delta<1}$ be robust equilibrium payoffs converging to w^k. Choose (δ^k) so that $\|w^k - w^k(\delta^k)\| \leq \frac{1}{k}$ for all k. Then, $\|w - w^k(\delta^k)\| \leq \|w - w^k\| + \frac{1}{k} \to 0$ as $k \to \infty$, which proves that w is a robust limit equilibrium payoff.

Can this expected payoff also be attained under an XPE for large δ? First of all, note that the grim-trigger strategy profile introduced above will not do it. In the environment $e = L^\infty$, the player 1's ex-post payoff from not deviating and playing y_1 in every period equals -1. It is therefore obvious that player 1 should deviate, no matter the value of δ. Further, since $(\frac{5}{4}, \frac{5}{4})$ can only be written as a convex combination (with equal weights) of two extreme points $(-1, \frac{7}{2})$ and $(\frac{7}{2}, -1)$, there is no other XPE strategy profile that would work. So, $(\frac{5}{4}, \frac{5}{4})$ cannot be attained as an XPE payoff for any $\delta < 1$.

But still, we will now show that $(\frac{5}{4}, \frac{5}{4})$ is indeed a robust limit equilibrium payoff. Intuitively, the grim-trigger strategy profile in which (y_1, x_2) and (x_1, y_2) are played in game L and R, resp., violates conditions for XPE only in a "small" number of environments. This is because frequencies of L and R stage games are identical for almost every environment under π by the strong law of large numbers, and in that average sense, an XPE can be constructed to approximate the expected payoff of $(\frac{5}{4}, \frac{5}{4})$ with arbitrary precision as $\delta \to 1$.

Here goes: fix some large number k and consider a random walk $(b(\theta^{0:t}))$ on $\{0, \pm 1, \ldots, \pm k\}$ that starts at 0, i.e., $b(\emptyset) = 0$. Game L pushes the walk towards the left and R pushes it towards the right, and the two endpoints $\pm k$ are reflecting boundaries, i.e.,

$$b(\theta^{0:t-1}, L) = \max\{b(\theta^{0:t-1}) - 1, -k\} \text{ and } b(\theta^{0:t-1}, R) = \min\{b(\theta^{0:t-1}) + 1, k\}.$$

Now, consider the following strategy profile:

- if $\theta^t = L$, then play (y_1, x_2) when $b(\theta^{0:t-1}) > -k$ and (y_1, y_2) when $b(\theta^{0:t-1}) = -k$;
- if $\theta^t = R$, then play (x_1, y_2) when $b(\theta^{0:t-1}) < k$ and (y_1, y_2) when $b(\theta^{0:t-1}) = k$;
- the players play (x_1, x_2) ad infimum after the first deviation.

Figure 2 visually depicts this strategy profile and associated stage payoffs at each step. An intuitive way to think about this construction is that the random walk represents a budget of how many times a player can obtain a negative payoff if the same stage game keeps arriving.

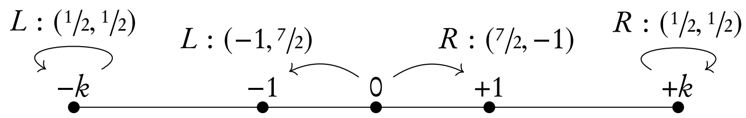


Figure 2: Construction of XPE in Example 1.

This strategy profile, we claim, is an XPE for large values of δ. If δ is close to 1, then the worst case, or the tightest incentive constraint for player 1 is when the random walk is currently at the right corner at $+k$, and then game L occurs in all future periods. So, after $2k$ periods of a stage payoff of -1, the state process reaches $-k$, and player 1 is then guaranteed a payoff of $\frac{1}{2}$. The incentive constraint reads as

$-(1-\delta^{2k}) + \frac{\delta^{2k}}{2} \geqslant 0$ *that implies* $\delta \geqslant \left(\frac{2}{3}\right)^{(2k)^{-1}}$.[13]

Next, it is routine to verify that the random walk admits a unique stationary distribution, which turns out to be uniform over $\{0, \pm 1, \ldots, \pm k\}$. *Taking* $\delta \to 1$, *simple calculations show that this strategy profile then gives each player the following expected payoff:*

$$\frac{5}{4}\frac{2k-1}{2k+1} + 2\frac{1}{2k+1} - \frac{1}{4}\frac{1}{2k+1}. \tag{9}$$

The expression in Eq. (9) converges to $\frac{5}{4}$ *as* $k \to \infty$, *so the point* $(\frac{5}{4}, \frac{5}{4})$ *is a robust limit equilibrium payoff because the set of robust limit equilibrium payoffs is closed, and we are done.*

The gist of this argument is generalizable to any set of stage games that satisfy a standard interiority assumption, to multiple players, and to any irreducible and aperiodic first-order Markov process. A key simplification afforded by the example above is that the minmax payoff is a static Nash in each stage game. No such restriction is required for the more general result. But, it does make the construction more delicate for the punishments too must be robustly incentivized — they too have to be an XPE. This gives us the *robust folk theorem* which states that given a state-wise interiority condition, every feasible and individually rational payoff of a stochastic game can be obtained robustly in an ex-post perfect equilibrium.

The proof leverages the classic approach of Fudenberg and Maskin [1986] in that each deviating player is minmaxed over a sufficiently long time horizon and the other players are eventually rewarded for punishing the deviator. There are two conceptual innovations: history-dependent duration of punishment phases and a state process that limits ex-post losses of players through a budget mechanism, as in Example 1. The former would be required to prove an SPE version of the folk theorem for an arbitrary stochastic game π . The latter, however, is specific to XPE, so that even if the players knew an environment, they would not gain by deviating and triggering punishments.

Careful calibration is required to make the budget mechanism, characterized by the state process, work generally. As we saw in Example 1, XPE in pinned down by the tightest incentive constraint across all environments, given there by $e = L^\infty$. However, for any finite but growing sequence of stages games, the probability of the binding environment gets vanishingly small and so does its impact on expected payoff under the true stochastic process. The budget mechanism uses this extra information to simultaneously achieve ex-post perfectness and the desired expected equilibrium payoff in the limit.

This is done by altering the players' strategy if the realized sequence of stage games is converging

[13] For intermediate values of δ, the worst case might be different. The reader can verify that it is as follows: the walk starts at k and either $e = L^{2k}(RL)^\infty$ or $e = L^\infty$ is the worst environment, which gives player 1 the payoff of $\min\left\{-(1-\delta^{2k}) + \delta^{2k}\frac{-1+\frac{7\delta}{2}}{1+\delta}, -(1-\delta^{2k}) + \frac{\delta^{2k}}{2}\right\}$.

towards the binding environment. And otherwise, the strategy sticks to the benchmark strategy profile which is subgame perfect. Given the subgame perfectness of the original strategy under the true stochastic process, the scenarios where players' strategy needs to be altered are infrequent, and thus the desired expected payoff is asymptotically attained. The technical challenge for the general proof is to calibrate the budget mechanism to work both and on-and-off path, blending it with the standard rewards and punishment construction from repeated games.

Moving onto the formal statement, we first recall two well-known folk theorems for general stochastic games.[14]

Theorem 1 (Theorem 3 in Fudenberg and Yamamoto [2011]; Theorem 2 in Hörner, Sugaya, Takahashi, and Vieille [2011]). *If* $Int(\sum_{\theta\in\Theta}\mathscr{U}(\theta)\mu(\theta|\pi))\cap\mathbb{R}^n_{++}$ *is non-empty, then* w *is a limit equilibrium payoff of* π *if and only if it is an element of*

$$(\textstyle\sum_{\theta\in\Theta}\mathscr{U}(\theta)\mu(\theta|\pi))\cap\mathbb{R}^n_{+}. \tag{10}$$

Here the interiority assumption, $\mathrm{Int}(\sum_{\theta\in\Theta}\mathscr{U}(\theta)\mu(\theta|\pi))\cap\mathbb{R}^n_{++}$, is a direct generalization of the standard version in Fudenberg and Maskin [1986], which in the context of a single repeated game seeks some minimal slack in the payoff set so that each player can be punished and rewarded for punishing others. The condition invoked above seeks a similar slack in expected payoffs. In the context of Example 1, it simply means that the grey-shaded area intersected with the positive orthant should have a non-empty interior.

The result then claims the equivalence between the set of feasible and individually payoffs, $(\sum_{\theta\in\Theta}\mathscr{U}(\theta)\mu(\theta|\pi))\cap\mathbb{R}^n_{+}$, where each player obtains at least her expected minmax payoff; and the set of limit equilibrium payoffs of the stochastic game. Recollect that one direction here is obvious, that all (limit) equilibrium payoffs must be not lower than the expected minmax; the other direction that each feasible and individually rational expected payoff can be asymptotically attained as an SPE constitutes the folk theorem.

The main result of this section shows that there is no cost of robustness in terms of expected equilibrium payoffs under the natural condition of state-wise interiority.

Theorem 2. *Suppose that* $Int(\mathscr{U}(\theta))\cap\mathbb{R}^n_{++}$ *is non-empty for all* $\theta\in\Theta$*. Then, every limit equilibrium payoff of* π*, i.e., an element of the set defined in Eq.* (10)*, is a robust limit equilibrium payoff.*

Proof. See Appendix. □

The result claims that all feasible and individually rational expected payoffs can be asymptotically attained as an ex-post equilibrium of the stochastic game. It establishes an equivalence between

[14] The original results are stated for a more general model with action-dependent transitions and imperfect monitoring of past actions. Theorem 1 here states the special case restricted to our setting.

the sets of limit equilibrium payoffs and robust limit equilibrium payoffs. Robustness in the sense of ex-post perfect equilibrium is a significantly more demanding condition than simply requiring subgame perfectness, and yet for the patient players, there is no loss. In the context of Example 1 again, this means that the part of the grey-shaded area lying in the positive orthant, including payoffs on the two axis, can actually be attained robustly as $\delta \to 1$.

Now, it is important to strengthen the interiority assumption to its state-wise form for we demand a stronger equilibrium criterion. The following example illustrates this point, also suggesting it might be difficult to obtain this theorem under weaker conditions.

Example 2. *Game L is a standard prisoner's dilemma, game R only has one action available for both players which yields them a payoff of 0 each.*

$1, 1$	$-2, 2^\star$
$2^\star, -2$	$0^\star, 0^\star$

Game L

$0^\star, 0^\star$

Game R

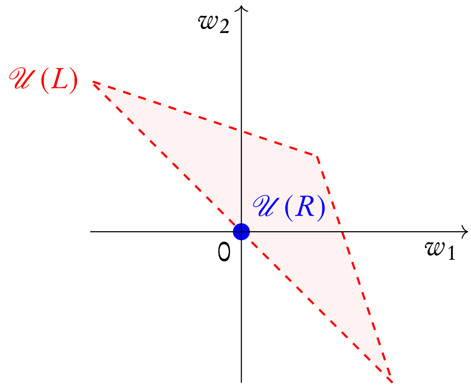


Table 2: Uncertain repeated game for Example 2: "$\star$" indicates the static best-response, $\mathscr{U}(L)$ is a red polygon and $\mathscr{U}(R)$ is a blue dot.

For every positive stochastic matrix π, the associated invariant measure assigns positive probability to game L, i.e., $\mu(L|\pi) > 0$, and the interiority assumption of Theorem 1 holds. So, every payoff vector $w \geq 0$ in $\mathscr{U}(L)\mu(L|\pi)$ is attainable, at least in the limit, as an SPE.

However, when we look at XPE, the interiority assumption from Theorem 2 fails for Game R. In terms of incentives, it is easy to see that since game R has only one action profile, any two strategy profiles give the same payoff in the environment $R^{1:\infty}$. So, there is no room to punish for a deviation in game L if game R realizes forever after. As a result, no matter what π we start with, for any $\delta < 1$, the unique XPE here is to play the static Nash equilibrium in every period.

Theorem 2 delivers an asymptotic characterization of expected payoffs when incentives are required to hold sans any reference to a stochastic process on the evolution of stage games. The modeler though is Bayesian and equipped with a Markov process on how games evolve. This conceptually allows the designer or modeler to have clear objectives in mind, and to that end, it also allows for explicit calculations of expected payoffs. Going forward, we drop stochasticity all together, and seek an environment-wise characterization of actions (and associated ex-post payoffs) that can be chosen by patient players on-path in some XPE.

4 Robustness of on-path behavior in dynamic games

We now embrace complete uncertainty on the evolution of stage games within the context of our model: keeping the set of stages games fixed still, we do not take any call on a stochastic process from which stage games are drawn. The equilibrium concept remains the same, i.e., ex-post perfect equilibrium; however, we now do not have access to π from the earlier section that helped us define expected payoffs. In fact, the set of feasible ex-post payoffs is now given by

$$\left\{ w \in \mathbb{R}^{n \cdot \Theta^\infty} | \exists\, \sigma \text{ s.t. } w(e) = U^\sigma(e) \ \ \forall e \in \Theta^\infty \right\}. \tag{11}$$

This set is infinite-dimensional and, unlike the standard repeated game, varies with δ. We will indirectly characterize this set using the classical language of Abreu [1988] and state results in terms of outcomes.[15]

4.1 The outcomes approach and statement of the result

An **outcome** α is an alternative way to think about strategy profiles. It prescribes for each on-path history $\widetilde{h}^t$, which includes only sequences of stage games and sunspots, $\theta^{0:t}$ and $\omega^{0:t}$, resp., an action profile $\alpha(\widetilde{h}^t) \in A(\theta^t)$.[16] So, a strategy profile induces an outcome that describes on-path behavior, and hence following Eq. (2), we can define the payoff under outcome α as

$$U^\alpha(\widetilde{h}^t|e) = (1-\delta)u(\alpha(\widetilde{h}^t)|\theta^t) + \delta \mathbb{E}_t[U^\alpha(\underbrace{\widetilde{h}^t, \theta^{t+1}, \omega^{t+1}}_{=\widetilde{h}^{t+1}}|e)], \tag{12}$$

where $\mathbb{E}_t[\cdot]$ represents the expectation over the $(t+1)$-th period sunspot. Eq. (12) defines a mapping from outcomes to feasible ex-post payoffs, therefore if we knew which outcomes can occur on-path in some XPE, we would be able to recover the whole set of ex-post payoffs that can be attained in XPE.

Of course, not every outcome can arise on-path in some XPE.

Definition 5. *An* outcome α *is said to be (robustly)* **justifiable** *if there exists an XPE σ such that for each history h^t, if $\alpha(\widetilde{h}^s) = a^s$ for all $s < t$, then $\alpha(\widetilde{h}^t) = \sigma(h^t)$.*

[15] Recently, Panov [2022] further developed this idea to study equilibria in continuous time to bypass the problem of defining strategies in continuous time. Carroll [2021] uses outcomes to describe XPE in uncertain repeated games with one long-run player. We push these ideas in the direction of a general description of equilibrium in terms of outcomes for uncertain repeated games.

[16] Note in our notations, a complete history h^t differs from an on-path history $\widetilde{h}^t$ in that the former also contains a sequence of past action profiles, $a^{1:t-1}$.

That is, an outcome α is justifiable if it agrees on-path with some XPE strategy profile, and robustness is implicit in the definition for it invokes justifiability corresponding to an XPE and not just an SPE using π or for a fixed environment e. In what follows, the word robustness will not be emphasized, but it is clear that we are working with robust justifiability.

It should also be clear that in the space of outcomes, there is an implicit recursion in the notion of justifiability — an outcome is justifiable if and only if upon deviating from its path the players can switch to another justifiable (punishment) outcome. Since each player i can keep deviating ad infimum, she can guarantee herself at least the discounted sum of her stage game minmax payoffs, $(1-\delta)\sum_{t=0}^{\infty}\delta^t \underline{r}_i(\theta^t)$, which equals to 0 for all $e \in \Theta^\infty$ due to our normalization. As a result, if α is justifible, then the player i's dynamic loss from the best static deviation, which gives $d_i(\alpha(\widetilde{h}^t)|\theta^t)$, cannot exceed $\frac{\delta}{1-\delta}\mathbb{E}_t[U_i^\alpha(\widetilde{h}^{t+1}|e)]$. This gives us a simple necessary condition for justifiability, which we term individual rationality keeping with the tradition. The following definition marginally strengthens this condition requiring an additional slack of ε.

Definition 6. *An* outcome α *is is (ex-post) ε-**individually rational** if for every environment $e \in \Theta^\infty$ and each on-path history $\widetilde{h}^t$ that is compatible with e,*

$$(1-\delta)d_i(\alpha(\widetilde{h}^t)|\theta^t) \leqslant \delta(\mathbb{E}_t[U_i^\alpha(\widetilde{h}^{t+1}|e)] - \varepsilon) \ \ \forall i \in N. \tag{13}$$

As discussed, 0-individually rationality is necessary for justifiability irrespective of δ. The main result of this section establishes approximate sufficiency of individual rationality for justifiability for large values of δ. This approximation is captured by ε-individually rationality for arbitrarily small $\varepsilon > 0$, and the result is later strengthened by taking ε to zero. Conceptually, it states each player can be approximately minmaxed in all environments at the same time provided that δ is large enough. To illustrate the main ideas, we start with an example.

Example 3. *There are two games, L and R, and two players. The sets of feasible payoffs, $\mathscr{U}(L)$ and $\mathscr{U}(R)$, are exactly the same as in Example 1 but the minmax payoff vector $(0,0)$ is no longer a static Nash in either game.*

	x_2	y_2	z_2
x_1	$-1,-1$	$0,0^\star$	$-1,-1$
y_1	$0^\star,0$	$-1,-1$	$1/4, 1/4^\star$
z_1	$-1, 7/2^\star$	$1/4^\star, 1/4$	$1/2^\star, 1/2$

Game L

	x_2	y_2	z_2
x_1	$-2/3, -2/3$	$0,0^\star$	$7/2^\star, -1$
y_1	$0^\star,0$	$-1,-1$	$1/4, 1/4^\star$
z_1	$-1,-1$	$1/4^\star, 1/4$	$1/2, 1/2^\star$

Game R

Table 3: Uncertain repeated game for Example 3: "$\star$" indicates the static best-response.

Is it possible for each player to be minmaxed in every period in each stage game on-path? That is, can the payoff vector (0,0) be attained as an XPE no matter the realization of stage games? If it is indeed so, *then we have the complete characterization of justifiable outcomes; that is, every* 0*-individually rational outcome is justifiable, because the outcome that attains the minmax payoff can be used as the punishment. Of course, the answer must depend on* δ*. For* δ *low enough, since either static game does not admit a pure strategy Nash equilibrium, their repetition,* $e = L^{1:\infty}$ *or* $R^{1:\infty}$*, will not have any SPE either, and hence there will be no XPE for the uncertain repeated game as well.*

We claim that the answer is yes, and it can be proven using the automaton approach with just two states, say $\underline{S}$ *and* S*. Consider the following strategy profile:*

- *the play starts at state* $\underline{S}$ *in which the players play* (x_1, x_2) *and stay in this state with probability* $1 - \lambda(\theta^t)$ *if no player deviated;*
- *the players play* (z_1, z_2) *in state* S *and stay in this state with probability one if no player deviated;*
- *any deviation restarts the strategy profile moving it back to state* $\underline{S}$*.*

Clearly, the ex-post payoff under this strategy profile depends on a history only through the current state. Specifically, it delivers each player the ex-post payoff of $\frac{1}{2}$ *when in state* S *and the ex-post payoff* $\underline{w}$ *when in state* $\underline{S}$*, which satisfies the following recursion:*

$$\underline{w}(\theta, e) = \begin{cases} -(1-\delta) + \delta\left((1-\lambda(L))\underline{w}(e) + \lambda(L)\frac{1}{2}\right) & \text{if } \theta = L, \\ -(1-\delta)\frac{2}{3} + \delta\left((1-\lambda(R))\underline{w}(e) + \lambda(R)\frac{1}{2}\right) & \text{if } \theta = R. \end{cases}$$

For this automaton to minmax the players, we must have $\underline{w}(e) = 0$ *for all environments* $e \in \Theta^\infty$*. This uniquely pins down two transition probabilities to be* $\lambda(L) = \frac{2(1-\delta)}{\delta}$ *and* $\lambda(R) = \frac{4(1-\delta)}{3\delta}$*, which are between* 0 *and* 1 *for* $\delta \geq \frac{2}{3}$*. By construction, the best-one shot deviation in state* $\underline{S}$ *gives* 0*, thus no player can profitably deviate in this state. As for the other state, we require* $(1-\delta)\frac{1}{2} + \delta\frac{1}{2} \geq (1-\delta)\frac{7}{2}$*, which gives* $\delta \geq \frac{6}{7}$*.*

In summary, for $\delta \geq \frac{6}{7}$*, there is an XPE in which both players obtain exactly the discounted sum of their stage game minmax payoffs. Hence, an outcome (not necessarily symmetric) is justifiable if and only if it is* 0*-individually rational for large values of* δ*.*

Example 3 illustrates how for large enough values of δ, the players can be robustly minmaxed in all environments at the same time, which allows us to justify all 0-individually rational outcomes. This statement, generalized to all uncertain repeated games, gives us the folk theorem. Since the claim is strong, it needs an extra restrictions on primitives, and in its full generality, requires a bit of slack as captured by ε-individually rationality. As a piece of notation, we use $\bigwedge$ for the coordinate-wise infimum of a bounded Euclidean set.

Theorem 3. *Suppose that* $Int(\mathscr{U}(\theta)) \cap \mathbb{R}^n_{++}$ *is non-empty and* $\bigwedge(\mathscr{U}(\theta) \cap \mathbb{R}^n_+) = 0$ *for all* $\theta \in \Theta$*. Then, for each sufficiently small* $\varepsilon > 0$ *there exists* $\delta^\varepsilon < 1$ *such that for every* $\delta \geq \delta^\varepsilon$*, an outcome* α *is justifiable if it is* ε*-individually rational.*

The first condition is the same state-wise interiority assumption introduced in Theorem 2. The second condition requires that each player can be asymptotically minmaxed in each individual stage game. If this is not the case, i.e., there exists a θ such that the lowest payoff in $\mathscr{U}(\theta) \cap \mathbb{R}^n_+$ for some player is strictly positive for, then asymptotic equivalence between justifiability and individual rationality breaks down even in the repeated game when θ is repeated ad infimum.

The second hypothesis is satisfied in many models in which a player can be punished without hurting the other players too much. It always holds if the players can engage in money-burning[17], and, importantly, it is satisfied in every symmetric uncertain repeated game, see Lemma 2 in Appendix for the proof. The condition also holds in many asymmetric situations, e.g., it is satisfied in Bertrand games, and also Cournot where a player can be punished by taking an action that pushes a market price below her own marginal costs.

To prove Theorem 3, we adopt the recursive method of Abreu, Pearce, and Stacchetti [1990], intersected across all stage games, to identify a class of XPE —it is described in Lemma 1 below. As an illustration of the main idea, suppose payoff matrices in Example 3 are perturbed in an arbitrary way, while still satisfying conditions of Theorem 3. So, a simple symmetric strategy profile based on a two-state automaton may no longer be sufficient to attain the minxmax payoff as an XPE.

As a first step towards developing a general recursion, the payoffs in each stage game are normalised by a vector, $\underline{u} \in \mathbb{R}^{n \cdot \Theta}$. Since the payoff matrices might differ substantially after being perturbed, $\underline{u}$ serves the role of making them comparable again, so that incentives can be constructed that may work across all environments. Then, an important property of strategy profile in Example 3 is imposed—players' continuation ex-post payoffs are independent of the future sequence of stage games. This additional independence restriction dramatically reduces dimensionality of ex-post payoffs and allows for building XPE using powerful recursive techniques. As we vary the initial parametrization, we get a different subset of (normalized) XPE payoffs

So, the defining feature of XPE in the class that we will use to establish the theorem is that at every history h^t, the players' ex-post payoffs depend on a realized environment only through the chosen normalizing vector $\underline{u}$, i.e.,

$$\mathbb{E}_{t-1}[U^{\sigma}(h^t|e)] - (1-\delta)\sum_{\tau=t}^{\infty} \delta^{t-\tau}\underline{u}(\theta^t) \text{ is independent of } \theta^{t:\infty}, \tag{14}$$

where $\mathbb{E}_{t-1}[\cdot]$ is integrating out the t-th period sunspot.

We use this restricted class of XPE to construct punishments, one for each player, so that the players' ex-post payoffs are at most ε in all environments at the same time, provided that δ is large

[17] A game (A, u) with money-burning up to $M \geq 0$ is such that each player is given an additional action that reduces her payoff by exactly M without affecting the other players' payoffs. Clearly, the set of feasible payoffs with money-burning up to $M \geq 0$ satisfies $\mathscr{U}(M) := \mathscr{U} + [-M, 0]^n$, and $\bigwedge(\mathscr{U}(M) \cap \mathbb{R}^n_+) = 0$ holds provided that M is large enough.

enough. Hence, as in Example 3, the minmax payoff vector can be attained, here approximated to ε, on-path in an XPE. Finally, we use this as punishment to conclude that ε-individual rationality is sufficient for justifiability.

4.2 Recursive technique

We refer the reader to Appendix for the complete proof of Theorem 3 and focus here only on the first step.

Lemma 1 (Recursive technique to identify a subset of XPE). *Let $\underline{u} \in \mathbb{R}^{n \cdot \Theta}$. Suppose that $\Gamma \subset \mathbb{R}^n$ is a bounded subinvariant of the recursion $\mathbf{T}_{\underline{u}}$, i.e., $\Gamma \subseteq \mathbf{T}_{\underline{u}}\Gamma$, defined by*[18]

$$\mathbf{T}_{\underline{u}} G = \bigcap_{\theta \in \Theta} Conv\Bigg(\Bigg\{\gamma \in \mathbb{R}^n | \exists a \in A(\theta), (g^i)_{i=0}^n \subset G \ s.t.$$
$$\text{(PK)} \ \ \gamma = (1-\delta)(u(a|\theta) - \underline{u}(\theta)) + \delta g^0,$$
$$\text{(IC)} \ \ (1-\delta)u_i(a|\theta) + \delta g_i^0 \geq (1-\delta) r_i(a|\theta) + \delta g_i^i \ \ \forall i \in N\Bigg\}\Bigg). \tag{15}$$

Then, for every $\gamma \in \Gamma$, there exists an XPE σ satisfying Eq. (14) *and*

$$U^{\sigma}(e) = \gamma + (1-\delta)\sum_{t=0}^{\infty} \delta^t \underline{u}(\theta^t) \ \ \forall e \in \Theta^{\infty}. \tag{16}$$

Proof. See Appendix. □

Lemma 1 describes a way to construct XPE in which the players obtain environment-independent gaps over the normalizing vector $\underline{u}$. A gap here is simply the payoff a player gets over and above the normalizing vector. Instead of recursing on payoffs as in say Abreu et al. [1990], we will recurse over these normalized payoffs, or gaps; and in addition these gaps will be independent of the environment.

Why don't we simply recurse over ex-post payoffs as in the classical approach? Recollect that an attainable payoff can be a function of the full ex-post realization of the environment, $w(e) \in \mathbb{R}^n$ for each $e \in \Theta^{\infty}$. Hence, the set of feasible payoffs, as laid out in Eq. (11), is infinite-dimensional, and the set of equilibrium payoffs too is infinite-dimensional.

So, even though the standard APS recursion is still valid, recursing over subsets of an infinite-dimensional space, it cannot tell us much, in any generality about the structure of equilibria — the geometry of the equilibrium payoff set is hard to compute, even numerically. In this backdrop, environment-independent gaps deliver two objectives: first they reduce dimensionality of

[18] The gs in the recursion depend on θ, the notation is suppressed for clarity.

the problem and make it tractable, and second they provide a robust way to preserve incentives across all future environments.

Figure 3 illustrates the construction. Payoffs from two possible stage games have been shaded in red and blue. Given some set of gaps G, which is a gray circle in the figure, we first apply the standard APS operator to $G + \underline{u}(\theta)$ in stage game θ. This operation produces two sets of payoffs, which are depicted as ellipses around $\underline{u}$, that can be attained in each stage game θ provided that that future continuations are in $G + \underline{u}(\theta)$.

Next, we translate the outputs of the APS recursion by $\underline{u}$ back to the origin and take the intersection over θ to obtain the set of gaps $\mathbf{T}_{\underline{u}}G$ that can be attained with gaps in G. If it happens that G is subinvariant, i.e., a subset of $\mathbf{T}_{\underline{u}}G$ as shown in the figure, then any gap in G can be attained with continuations in G irrespective of what stage games will arrive in the future. This allows us to construct XPE by recurring over elements of G in which realized environments shift payoffs exactly by the discounted sum of normalizing vectors as shown in Eq. (16).

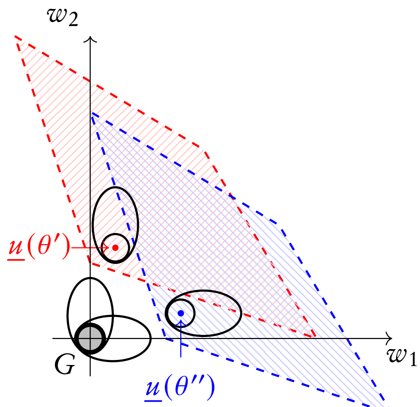


Figure 3: Illustration of recursion in Lemma 1.

In the proof of Theorem 3, we combine this lemma with the classical argument of Fudenberg, Levine, and Maskin [1994] to argue that each player i can be asymptotically minmaxed irrespective of the future stage games by selecting $\underline{u}$ in a way that $\underline{u}$ is positive and $\underline{u}_i(\theta)$ is close to 0 for all $\theta \in \Theta$. By Fudenberg, Levine, and Maskin [1994], a small closed ball around $\underline{u}(\theta)$ is subinvariant with respect to the APS recursion of stage game θ for all large values of δ. Since this holds for all stage games simultaneously, the ball actually satisfies the condition of Lemma 1; hence, there is an XPE in which the players obtain the ex-post payoff of $(1-\delta)\sum_{t=0}^{\infty}\delta^t \underline{u}(\theta^t)$, provided that δ is large enough.

4.3 Exact equivalence

Given that Theorem 3 is an asymptotic result, it is natural to ask if 0-individual rationality is ever equivalent to justifiability for large but fixed values of δ. The following corollary uses an extension of Lemma 1 to give an affirmative answer. Specifically, it lists a set of sufficient conditions under which this is indeed the case.

Corollary 1. *Suppose that for each player i, there exists a tuple $(g^i, a^i, \xi^i) \in \mathbb{R}^n_{++} \times \prod_{\theta \in \Theta} A(\theta) \times \mathbb{R}^{n \cdot \Theta}_{+}$ such that for all $\theta \in \Theta$, $r_i(a^i(\theta)|\theta) = \xi^i_i(\theta) = 0$, and*

(i) g^i belongs to $\mathscr{U}(\theta)$;

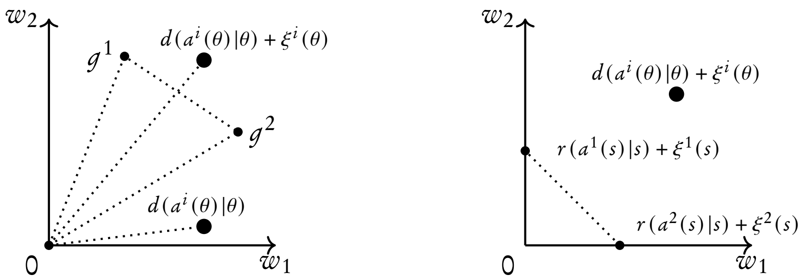
(ii) $d(a^i(\theta)|\theta) + \xi^i(\theta)$ belongs to the conic hull of $\{0, g^1, \ldots, g^n\}$;

(iii) $\lambda \cdot (d(a^i(\theta)|\theta) + \xi^i(\theta)) \geq \max_{(j,s) \in N \times \Theta} \lambda \cdot (r(a^j(s)|s) + \xi^j(s))$ for some $\lambda \in \mathbb{R}^n_+$.

Then, there exists $\bar{\delta} < 1$ such that for every $\delta \geq \bar{\delta}$, an outcome α is justifiable if and only if it is 0-individually rational.

Proof. See Appendix. □

Corollary 1 imposes stronger assumptions than Theorem 3. The first condition assumes that the set $\cap_{\theta \in \Theta}(\mathscr{U}(\theta))$ contains strictly individually rational payoffs. The second condition ensures that each vector $\frac{1-\delta}{\delta}(d(a^i(\theta)|\theta) + \xi^i(\theta))$ belongs to a convex hull of 0 and g's for all sufficiently large δ. Finally, the third condition ensures that each vector $d(a^i(\theta)|\theta) + \xi^i(\theta)$ lies above hyperplanes through the stage game deviation payoffs from the minmax action profiles in the uncertain repeated game.



(a) Condition (ii).

(b) Condition (iii).

Figure 4: Visualizing Conditions (ii) and (iii) of Corollary 1.

Condition (ii) is satisfied if g's can be chosen sufficiently close to the axes, that is the conic hull of 0 and non-negative points in $\cap_{\theta \in \Theta}\mathscr{U}(\theta)$ is all of $\mathbb{R}^n_+$. In other words, it is possible to punish player i without punishing the other players too much, e.g., if money-burning is allowed or there are punishments that do not hurt other players too much, which is typically possible in Cournot with positive costs. Even if this is not the case, the slack variables can be used to relax the constraint, as shown in Figure 4a.

Condition (iii) is not restrictive at all when there are two players.[19] For uncertain repeated games with more than two players, this condition can be seen as a weaker version of the requirement that the players can be minmaxed at the same time. Again, it holds with money-burning and in Cournot with positive costs where the players can jointly select inefficiently large quantities.

5 Symmetric games

A frequently studied special case of repeated interactions is when players are ex ante identical and use strategies that are in a certain sense symmetric. In this section, we explore this special case in the context of our model.

A game (A, u) is said to be **symmetric** if $A_1 = \ldots = A_n$ and for every permutation $\iota : N \to N$, the players' payoffs satisfy

$$u_i(a_1, \ldots, a_i, \ldots a_n) = u_{\iota(i)}(a_{\iota(1)}, \ldots, a_{\iota(i)}, \ldots a_{\iota(n)}) \;\; \forall a \in A, \; \forall i \in N. \tag{17}$$

An uncertain repeated game $(\Theta, (A(\theta), u(\cdot|\theta))_{\theta\in\Theta}, \delta)$ is **symmetric**, if each stage game is symmetric.[20]

A popular notion of symmetric behavior in repeated and stochastic games is strong symmetry, e.g., see Abreu, Pearce, and Stacchetti [1986] and Cronshaw and Luenberger [1994]. Formally, a strategy profile σ is **strongly symmetric** if the players play an identical action at each history h^t, i.e., $\sigma_1(h^t) = \ldots, \sigma_n(h^t)$. We shall refer to strongly symmetric XPE as SSXPE. Of course, the same definitions apply to stochastic and dynamic games, and we will refer to strongly symmetric SPE of these games as SSSPE.

It is not hard to see the model of SSXPE is mathematically isomorphic to the model of XPE with one long-run player who faces an infinite sequence of short-lived players, which was studied in Carroll [2021]. We refer the reader to the Supplementary Appendix in which we formally establish this equivalence and focus here on substantive aspects of SSXPE.

Strong symmetry substantially simplifies the analysis because the players' ex-post payoffs are identical at every history. It turns out that if the set of SSXPE is non-empty, then the set of ex-post SSXPE payoffs admits the largest and smallest elements, which can be identified using Lemma 1 in a closed form. Specifically, the operator $\mathbf{T}_{\underline{u}}$, when adapted to strongly symmetric strategy profiles, coincides with the recursion in Carroll [2021], provided that the normalization is chosen appropriately.[21]

[19] For each player i, let the action profile $a^i(\theta)$ to be the joint minmax in game θ and set $\xi^i \equiv 0$. The reader can verify that Condition (iii) is then always satisfied.

[20] This definition is fairly standard in the literature, e.g., see Section 5 of Dasgupta and Maskin [1986].

[21] Krasikov and Lamba [2023] obtain the same result by deriving certain bounds on XPE payoffs and showing that their bounds simplify to Eq. (19) in the context of SSXPE.

We need to introduce some auxiliary notations to simplify the statement of the next result. Let $\overline{m}(g,\theta)$ and $\underline{m}(g,\theta)$, respectively, be the maximal and minimal payoffs player 1 can obtain in stage game θ when the players play an identical action and cannot gain more than $\frac{\delta}{1-\delta}g \in \mathbb{R}$ from the best one-shot deviation. Formally:

$$\begin{cases} \overline{m}(g|\theta) & := \sup_{a_1 \in A_1(\theta)} \quad u_1(a_1,\ldots,a_1|\theta) \;\text{ s.t. }\; \frac{1-\delta}{\delta}d_1(a_1,\ldots,a_1|\theta) \le g, \\ \underline{m}(g|\theta) & := \inf_{a_1 \in A_1(\theta)} \quad r_1(a_1,\ldots,a_1|\theta) \;\text{ s.t. }\; \frac{1-\delta}{\delta}d_1(a_1,\ldots,a_1|\theta) \le g. \end{cases} \tag{18}$$

Clearly, $\overline{m}(g,\theta) = -\infty$ and $\underline{m}(g,\theta) = \infty$ whenever the constraint set in Eq. (18) is empty; otherwise, both are finite.

Theorem 4 (Carroll [2021]). *Let g^* be the largest fixed point of*

$$g \in [-\infty,\infty) \mapsto \inf_{\theta\in\Theta}(\overline{m}(g|\theta) - \underline{m}(g|\theta)). \tag{19}$$

If it is negative, then there is no SSXPE; otherwise,

$$(1-\delta)\sum_{t=0}^{\infty}\delta^t\overline{m}(g^*|\theta^t) \ge U_1^{\sigma}(e) \ge (1-\delta)\sum_{t=0}^{\infty}\delta^t\underline{m}(g^*|\theta^t) \;\; \forall e \in \Theta^\infty,\; \forall \text{ SSXPE } \sigma, \tag{20}$$

where both bounds can be attained in SSXPE.

Proof. See Appendix. □

Theorem 4 makes two claims. First, the theorem provides the necessary and sufficient condition for existence of some SSXPE. Second, assuming that at least one SSXPE exists, the theorem asserts existence of uniformly best/worst SSXPE, and Eq. (20) identifies their respective payoffs.

These conclusions immediately imply analogs of our general statements for strongly symmetric strategies. To keep our discussion crisp, we shall focus on first perspective, viz. robustness of expected payoffs. The reader interested in second perspective—robustness of behavior in dynamic games—is encouraged to look up Corollary 3 in Appendix.

Corollary 2. *Suppose $\overline{m}(\infty|\theta) - \underline{m}(\infty|\theta) > 0$ for all $\theta \in \Theta$. Then, there exists $\overline{\delta} < 1$ such that for every $\delta \ge \overline{\delta}$, the set of limit SSXPE payoffs of π coincides with the set of limit SSSPE payoffs of π and equals*

$$\left[\sum_{\theta\in\Theta}\underline{m}(\infty|\theta)\mu(\theta|\pi), \sum_{\theta\in\Theta}\overline{m}(\infty|\theta)\mu(\theta|\pi)\right]. \tag{21}$$

Proof. See Appendix. □

Corollary 2 shows that robustness of expected payoffs attainable in strongly symmetric strategies is costless provided that the players can be incentivized in those strategies uniformly across stage games, i.e., $\overline{m}(\infty|\theta) - \underline{m}(\infty|\theta) > 0$ for all $\theta \in \Theta$.[22] This conclusion follows directly from the fact the players' have perfectly aligned preferences in strongly symmetric strategy profiles, which effectively turns the uncertain repeated game into a decision problem. For example, under the condition of the corollary, the best SSXPE payoff is attainable by playing a pure action profile $(a_1(\theta), \ldots, a_1(\theta))$ that maximizes $u_1(a_1, \ldots, a_1|\theta)$ in stage game θ, for each $\theta \in \Theta$. This is not necessarily the case in asymmetric XPE, as shown in Example 1.

Furthermore, even in symmetric situations with no uncertainty, strongly symmetric strategies may be restrictive. First, such strategies could disallow punishing the players all the way to their minmax payoff vectors, which happens when $\underline{m}(\infty|\theta) > \underline{r}(\theta) = 0$ for some stage game $\theta \in \Theta$. Second, the best joint payoff vectors could be not attainable in SSXPE, i.e., $\overline{m}(\infty|\theta) < \max_{(w_1,\ldots,w_1)\in\mathscr{U}(\theta)} w_1$ for some stage game $\theta \in \Theta$. Example 6 in Appendix illustrates both points.

6 Extensions and discussion

In this paper, we examine robustness of perfect equilibria in stochastic and dynamic games with respect to the evolution of future stage games when players are patient. We made three key assumptions: monitoring of past actions is perfect, the present stage game is commonly known by all players, and state transitions are independent of players' actions. Moreover, we characterized the main results for limit discounting. These assumptions play an important role, and it should not be expected that costs of robustness are always negligible when any of them is not satisfied. In what follows we briefly discuss the implications of relaxing these assumptions, and end the section with a discussion on non-Bayesian modeling of beliefs that is hitherto absent from the uncertain repeated games framework.

6.1 Imperfect observability of states

Let's start with imperfect observability of (θ^t). A simple and hopefully intuitive model is to split θ into $n+1$ components $\theta = (\theta_0, \theta_1, \ldots, \theta_n)$ so that θ_0 is publicly known to all players but θ_i is privately observed by player i. Then, assume the following: each player i's set of actions $A_i(\theta)$ in stage game θ depend only on (θ_0, θ_i), which ensures that this player cannot learn the other players' information about θ when deciding on her own action. In addition, either assume that all players directly learn the stage game at the end of the period when actions have been taken or that

$$\{(u_i(a|\theta'), a) | a \in A(\theta')\} \cap \{(u_i(a|\theta''), a) | a \in A(\theta'')\} = \emptyset \ \ \forall(\theta', \theta'') \in \Theta^2,\ \forall i \in N,$$

[22]It is easy to see that if this condition fails, then, for all values of $\delta < 1$, any SSXPE is such that the players play a static Nash equilibrium at every history.

which implies that observing a utility-action profile pair at the end of the period is sufficient to perfectly learn the stage game.

Conceptually, a model of this sort can be studied using our techniques but with the additional requirement that each player i's strategy $\sigma_i(h^t)$ can depend only on (θ_0^t, θ_i^t) at every history h^t. The main observation here is that imperfect observability shrinks the set of equilibrium payoffs that are attainable in a stochastic game with no uncertainty but costs of robustness, in demanding that incentive satisfy ex-post perfectness, might still be insignificant in the limit. We illustrate these ideas through an example.

Example 4. *Consider Example 1 but assume that player 1 alone observes the stage game, L or R, but player 2 does not, and must play an identical action in both stage games. Clearly, the set of feasible expected payoffs is given by*

$$Conv\left(\left\{\frac{u(a(L)|L)+u(a(R)|R)}{2}|a \in A(L)\times A(R) \text{ s.t. } a_2(L)=a_2(R)\right\}\right), \tag{22}$$

which is not the same as $\frac{\mathcal{U}(L)+\mathcal{U}(R)}{2}$. *The maximal expected symmetric payoff vector now is* $(\frac{3}{4},\frac{3}{4})$. *This payoff can be attained when player* 1 *selects the second/first row in game L/R and player* 2 *randomizes between her actions with equal probabilities.*

We now show that it is possible to asymptotically attain $(\frac{3}{4},\frac{3}{4})$ *in an XPE using the construction similar to Example 1. Consider a random walk* $(b(\theta^{0:t}))$ *as described in the example before, and define a strategy profile as follows:*

- *if* $b(\omega^{0:t-1}) \neq \pm k$, *then player* 1 *selects* y_1/x_1 *in game L/R and player* 2 *randomizes with equal probabilities;*
- *if* $b(\omega^{0:t-1}) = \pm k$, *then the players play* (y_1, y_2) *that gives* $(\frac{1}{2},\frac{1}{2})$ *irrespective of* θ^t.
- *the players play* (x_1, x_2) *ad infimum after the first deviation.*

By exactly the same argument as in Example 1, this strategy profile is an XPE for large values of δ. *As* $\delta \to 1$, *it gives each player*

$$\frac{3}{4}\frac{2k-1}{2k+1}+\frac{1}{2}\frac{2}{2k+1}, \tag{23}$$

which converges to $\frac{3}{4}$ *as* $k \to \infty$.

6.2 Endogenous transitions

Let's now discuss a way to incorporate action-dependent state transitions into the model. Start with a fixed set of deterministic transitions Φ. Each element ϕ of Φ specifies a next period state as a function of this period stage game and actions chosen by the players, i.e., $(a,\theta) \mapsto \phi(a|\theta) \in \Theta$. A specific state transition is privately drawn by nature from Φ according to some $\kappa \in \Delta\Phi$ supported on Φ. Integrating over ϕ, we obtain that states evolve stochastically as follows:

$$\pi(a,\theta,s) := \mathbb{P}(\theta^{t+1} = s|\theta^t = \theta, a^t = a) = \sum_{\phi\in\Phi} \mathbb{1}(\phi(a|\theta) = s)\kappa(\phi). \tag{24}$$

Clearly, any stochastic game driven by state transitions π can be unpacked in this way. An appropriate notion of XPE immediately follows: a strategy profile is ex-post perfect if at every history h^t, each player would not regret not deviating even if she knew the exact sequence of future transition functions, i.e., $(\phi^{t+1}, \phi^{t+2}, \ldots)$. We again use an example to illustrate the workings of this model.

Example 5. *There are two stage games, L and R. Game L is a Prisoner's dilemma from Example 2, and game R is the same but every player's payoff is higher by 2. There are four possible transitions functions $\Phi = \{\phi_1, \phi_2, \phi_3, \phi_4\}$ such that $\phi_1 \equiv L$, $\phi_2 \equiv R$, $\phi_3(a|\theta) = L$ iff $\theta = R$ and $a_1 = a_2$, $\phi_4(a|\theta) = R$ iff $\theta = L$ and $a_1 = a_2$. Suppose that $\kappa(\phi_1) = \kappa(\phi_2) = \frac{1}{3}$ and $\kappa(\phi_3) = \kappa(\phi_4) = \frac{1}{6}$, then we obtain the stochastic game similar to the one in an earlier working paper version of Abreu, Brooks, and Sannikov [2020].*

	x_2	y_2
x_1	$1, 1$ [1/3]	$-2, 2^\star$ [1/2]
y_1	$2^\star, -2$ [1/2]	$0^\star, 0^\star$ [1/3]

Game L

	x_2	y_2
x_1	$3, 3$ [1/3]	$0, 4^\star$ [1/2]
y_1	$4^\star, 0$ [1/2]	$2^\star, 2^\star$ [1/3]

Game R

Table 4: Stochastic game for Example 5: "$\star$" indicates the static best-response, $[\cdot]$ is the probability of staying in stage game θ, i.e., $\pi(a,\theta,\theta)$ for each $a \in A(\theta)$.

We now show that there is an XPE in which the players play (x_1, x_2) on-path in every period provided δ is large enough. To see it consider the grim-trigger strategy in which the players start playing (y_1, y_2) after the first deviation.

Let $w(\phi^{0:\infty}|\theta) \in \mathbb{R}$ be each player's ex-post on-path payoff provided that the date 0 stage game is θ and the future stage games are determined according to the sequence $\phi^{0:\infty}$. Clearly, w can be expressed recursively as follows:

$$w(\phi^{0:\infty}|\theta) = (1-\delta)\mathbb{1}_{\{\theta=L\}} + 3\mathbb{1}_{\{\theta=R\}} + \delta w(\phi^{1:\infty}|\phi^0(x_1,x_2|\theta)). \tag{25}$$

Note that, since $\phi(x_1,x_2|\theta) = \phi(y_1,y_2|\theta)$ for all $(\theta,\phi) \in \Theta\times\Phi$, the players' ex-post payoff in the punishment phase is given by $w-1$. The net deviation gain from (x_1,x_2) is 1 in each stage game for *both players. As a result, no player has a profitable one-shot deviation on-path if and only if*

$$1 + w(\phi^{1:\infty}|\phi^0(x_1, x_2|\theta)) - w(\phi^{1:\infty}|\phi^0(y_1, x_2|\theta)) \geq \frac{1-\delta}{\delta}. \tag{26}$$

Similarly, no player has a profitable one-shot deviation in the punishment phase if and only if

$$w(\phi^{1:\infty}|\phi^0(x_1, x_2|\theta)) - w(\phi^{1:\infty}|\phi^0(y_1, x_2|\theta)) \geq -2\frac{1-\delta}{\delta}. \tag{27}$$

To see that Eq. (26) *and Eq.* (27) *are satisfied, for each pair of states* $(\theta, s) \in \Theta^\infty$, *define* $\Delta(\theta, s) := \inf_{\phi^{0:\infty} \in \Phi^\infty}(w(\phi^{0:\infty}|\theta) - w(\phi^{0:\infty}|s))$. *By construction,* $\Delta(\theta, \theta) = 0$ *for each* $\theta \in \Theta$, *and, by Eq.* (25),

$$\begin{cases} \Delta(L, R) = -2(1-\delta) + \delta \min\{\Delta(L, L), \Delta(R, R), \Delta(R, L)\} = -2(1-\delta), \\ \Delta(R, L) = 2(1-\delta) + \delta \min\{\Delta(L, L), \Delta(R, R), \Delta(L, R)\} = 2(1-\delta)^2, \end{cases} \tag{28}$$

which proves that Eq. (26) *and Eq.* (27) *hold for large values of* δ.

6.3 Imperfect public monitoring

We assumed that each player observes her opponents' actions and so rewards and punishments can be directly designed on the basis of this observability. This is obviously a strong assumption, and the literature typically relaxes it by assuming that a public signal is observed as a function of the actions of the players — this is called imperfect public monitoring. For example, in a repeated Cournot setting, instead of observing the quantities set by the opponents, the players observe a signal of the market price (see Green and Porter [1984]).

We conjecture that imperfect public monitoring can be incorporated into our analysis, subject to suitable identifiability conditions, that is, enough statistical information can be inferred from the public signals about the players' actions. In particular, the recursive technique of Lemma 1 can be directly generalized by relaxing the perfect observability assumption. So, we conjecture that under the hypothesis of Theorem 3 and standard rank conditions on monitoring technology (Fudenberg et al. [1994]), each player can be robustly minmaxed for large values of δ. We also conjecture that a version of Theorem 2 can be established. This is an immediately worthwhile direction for future work.

6.4 Results for fixed δ

A natural question to consider next is also the characterization of XPE for fixed values of δ: What is the set of robust equilibrium payoffs, in the sense of Definition 4, for a fixed discount factor in general stochastic games? And, what is the set of justifiable outcomes, in the sense of Definition 5, for a fixed δ in an uncertain repeated game?

We partially address the second question in a related paper, Krasikov and Lamba [2023]. The recursion characterized in Lemma 1 is valid for any fixed value of δ. It calculates a fixed point over environment-independent continuation values after normalizing each stage game with a vector of payoffs, $\underline{u}(\theta) \in \mathbb{R}^n \ \forall\theta \in \Theta$. This gives us a subset of XPE outcomes and by choosing the normalizing vectors smartly, we can ensure the subset is large enough. An outer bound on XPE outcomes is then constructed to get a grip on how well the lower bound performs in getting us close enough to the entire set of XPE outcomes. In fact for large values of δ and for strongly symmetric XPE, the two bounds collapse.

6.5 Modeling of preferences and beliefs

On the question of beliefs and dynamic preferences, the uncertain repeated games model takes an extreme stand, by taking no stand at all. Departing from non-Bayesian dynamics, one could potentially rely on multiple priors or the ambiguous preference approach, in the sense of Gilboa and Schmeidler [1989], or its general formulations such as Maccheroni, Marinacci, and Rustichini [2006b]. Several excellent models have been proposed on how to take this approach to dynamic settings, the main choice being how to update on multiple priors; see, for example, Epstein and Schneider [2003], Maccheroni, Marinacci, and Rustichini [2006a], and Hanany and Klibanoff [2007], among many others.

The notion of XPE can be thought of as a lower bound on predictions that can be obtained by fully specified models of this sort — a strategy profile satisfies the ex-post criterion if and only if it is subgame perfect simultaneously for all possible dynamic ambiguity indices (see Supplementary Appendix). Thus, it seems to be an appropriate notion to use when a modeler or an analyst is uncertain about players' preferences and beliefs. Of course, there are situations where some additional information about preferences is common knowledge, and robustness built in XPE may be 'excessive'. We think that it would be interesting to further explore uncertainty in repeated games of the kind studied here, allowing for a smaller family of dynamic variational preferences.

7 Final remark

The main goal of the paper here is theoretical— set out to offer a general model of uncertain repeated games, provide a clear enough language to state standard results such as the folk theorem, and construct equilibrium payoffs and outcomes. However, we believe that a variety of applications seem imminent. For example, what kind of collusion (Green and Porter [1984]), price wars (Rotemberg and Saloner [1986]), sustainable policy plans (Chari and Kehoe [1990]), risk sharing (Kocherlakota [1996]) or relational contracts (Levin [2003]) are robustly sustainable, without making strong assumptions about how the future unfolds.

8 Appendix

8.1 Appendix: Robustness of expected payoffs in stochastic games

Proof of Theorem 2. Clearly, it suffices to show that every w in the interior of $\sum_{\theta\in\Theta}\mathscr{U}(\theta)\mu(\theta|\pi)\cap\mathbb{R}^n_+$ is a robust limit equilibrium payoff. Our proof is constructive, and it is divided into three major parts. First, we will introduce additional objects and define a parametric family of strategy profiles. Then, we will show that strategy profiles in this family form an XPE for all large values of δ. Finally, we will establish that they can be used to approximate w with arbitrarily good precision as $\delta \to 1$.

Preliminaries. To begin, fix $x \in \prod_{\theta\in\Theta}\mathrm{Int}(\mathscr{U}(\theta))$ satisfying $w = \sum_{\theta\in\Theta}x(\theta)\mu(\theta|\pi) > 0$ and $\underline{x} \in \prod_{\theta\in\Theta}(\mathrm{Int}(\mathscr{U}(\theta))\cap\mathbb{R}^n_{++})$. Both points exist due to the hypothesis of the theorem. Select also minmaxing action profiles $(a^i)_{i\in N} \subset \prod_{\theta\in\Theta}A(\theta)$, one for each player, i.e., $r_i(a^i(\theta)|\theta) = 0$ for all $\theta\in\Theta$ and $i\in N$. We shall use $(x, \underline{x}, (a^i)_{i\in N})$ to define a parametric family of strategy profiles that will be used to establish the theorem. There are three additional components that are used to build a strategy profile.

Ingredient 1. Definitions of x and $\underline{x}$ imply that, for every sufficiently small $\varepsilon > 0$, we can find $(\beta^i, \underline{\beta}^i)_{i=0}^n \subset \prod_{\theta\in\Theta}\Delta A(\theta)\times\prod_{\theta\in\Theta}\Delta A(\theta)$ such that the following holds for every stage game-player pair $(\theta, i)\in\Theta\times N$:

$$u_i(\beta^j(\theta)|\theta) = \begin{cases} x_i(\theta)+\varepsilon & j = 0, \\ x_i(\theta)+\varepsilon+\left(2\mathbb{1}_{\{j\neq i\}}-1\right)\varepsilon^2 & j = 1,\dots,n, \end{cases} \tag{29}$$

$$u_i(\underline{\beta}^j(\theta)|\theta) = \begin{cases} \underline{x}_i(\theta)+\varepsilon & j = 0, \\ \underline{x}_i(\theta)+\varepsilon+\left(2\mathbb{1}_{\{j\neq i\}}-1\right)\varepsilon^2 & j = 1,\dots,n, \end{cases} \tag{30}$$

In words, in each stage game θ, the correlated action profiles $\beta^0(\theta)$ and $\underline{\beta}^0(\theta)$ deliver $x(\theta)$ and $\underline{x}(\theta)$ plus some small slack of $\varepsilon > 0$, and other correlated action profiles $\beta^j(\theta)$ and $\underline{\beta}^j(\theta)$ in addition punish player j and reward other players by even smaller amount, ε^2.

Ingredient 2. Let $\nu > 0$ be sufficiently small so that $\lambda^i(\theta) := \left(\varepsilon-\varepsilon^2-u_i(a^i(\theta)|\theta)\right)\cdot\nu \in (0,1)$ for all $\theta\in\Theta$ and $i\in N$. In our construction, λ capture probabilities of ending punishment phases, and thus ν is a proxy of duration of such phases. The exact value of ν will be chosen later as a function of ε.

Ingredient 3. Finally, define inductively a (Markov) random walk with a reflecting barrier

$b : \cup_{t\geq 0}\Theta^t \to [-\varepsilon^{-1}, \varepsilon^{-1}]^n$ as follows: for each player $i \in N$, set

$$b_i(\theta^{0:t}) = \min\left\{\max\left\{b_i(\theta^{0:t-1}) + x_i(\theta^t), -\varepsilon^{-1}\right\}, \varepsilon^{-1}\right\}, \tag{31}$$

where $b_i(\emptyset) = 0$. The walk starts at 0 and then changes by $x(\theta)$ each time game θ arrives and the boundary conditions are not binding. If the boundary condition binds for some coordinate i, then the walk takes the smallest step for that coordinate needed to reach the barrier.

We now define a strategy profile, which is parametrized by (ε, ν), and it also implicitly depends on $(\beta^i, \underline{\beta}^i)_{i=0}^n$. There are $2n + 1$ phases: the on-path phase, n punishment and n reward phases, one for each player. Any deviation by player i triggers the player i's punishment phase, whereas deviations by multiple players are ignored.

On-path phase. The game starts in this phase and stays in it until the first individual deviation. The strategy profile prescribes to play according to $\beta^0(\theta^t)$ if $b(\theta^{0:t})$ is above the lower barrier for all players and $\underline{\beta}^0(\theta^t)$, otherwise.

Player i's punishment phase. In this phase, the players minimax player i by playing $a^i(\theta)$ in stage game θ. The game stays in this phase with probability $1 - \lambda^i(\theta)$ and transitions to the player $-i$'s reward phase with the complementary probability.

Player $-i$'s reward phase. Similar to the on-path phase, the players are required to play according to $\beta^i(\theta^t)$ if $b(\theta^{0:t})$ is above the lower barrier for all players and $\underline{\beta}^i(\theta^t)$, otherwise.

Verifying conditions for XPE. We claim that the constructed strategy profile is ex-post perfect for all large values of δ provided that (ε, ν) are small enough. Since the one-shot deviation principle applies to XPE, because it applies to SPE of every dynamic game, it is sufficient to establish that there is no profitable one-shot deviation.

We first define the stream of ex-post payoffs in each of $2n+1$ phases. Given some environment $e \in \Theta^\infty$ and a current position of the walk $\underline{b} \in [-\varepsilon^{-1}, \varepsilon^{-1}]^n$, let $(U^{0,t}(e, \underline{b}))_{t=0}^\infty$ be the players' ex-post payoff in the on-path phase starting at $b(\emptyset) = \underline{b}$, when sunspots are averaged out, i.e.,

$$U^{0,t}(e, \underline{b}) = (1-\delta)\Big(\varepsilon + x(\theta^t)\mathbb{1}_{\{\min_{i\in N} b_i(\theta^{0:t}) > -\varepsilon^{-1}\}} + \underline{x}(\theta^t)\mathbb{1}_{\{\min_{i\in N} b_i(\theta^{0:t}) = -\varepsilon^{-1}\}}\Big) + \delta U^{0,t+1}(e, \underline{b}). \tag{32}$$

By construction, the player $j(\neq i)$'s ex-post payoffs in the player $-i$'s reward phase are given by $(U_j^{0,t}(e, \underline{b}))_{t=0}^\infty$ plus $\left(2\mathbb{1}_{\{i\neq j\}} - 1\right)\varepsilon^2$; as for the player i's punishment phase starting at $b(\emptyset) = \underline{b}$, the player j's ex-post payoffs $(U_j^{i,t}(e, \underline{b}))_{t=0}^\infty$ satisfy

$$U_j^{i,t}(e, \underline{b}) = (1-\delta)u_j(a^i(\theta^t)|\theta^t) + \delta\lambda^i(\theta^t)\Big(U_j^{0,t+1}(e, \underline{b}) + \left(2\mathbb{1}_{\{i\neq j\}} - 1\right)\varepsilon^2\Big) + \delta(1-\lambda^i(\theta^t))U_j^{i,t+1}(e, \underline{b}). \tag{33}$$

In what follows, we examine one-shot deviations in each of the phases.

On-path phase. Fix player $i \in N$. Combine Eq. (32) and Eq. (33) to obtain the following recursive expression:

$$U_i^{0,t}(e,\underline{b}) - U_i^{i,t}(e,\underline{b}) \geq (1-\delta)\Big(\varepsilon + x_i(\theta^t)\mathbb{1}_{\{\min_{j\in N} b_j(\theta^{0:t}) > -\varepsilon^{-1}\}} - u_i(a^i(\theta^t)|\theta^t)\Big) + \delta\lambda^i(\theta^t)\varepsilon^2 + \\ + \delta(1-\lambda^i(\theta^t))\Big(U_i^{0,t+1}(e,\underline{b}) - U_i^{i,t+1}(e,\underline{b})\Big), \quad (34)$$

where the inequality is due to $\underline{x}_i(\theta^t) \geq 0$. Eq. (34) implies

$$\liminf_{\delta\to 1}\Big(U_i^{0,t}(e,\underline{b}) - U_i^{i,t}(e,\underline{b})\Big) \geq \lambda^i(\theta^t)\varepsilon^2 + (1-\lambda^i(\theta^t))\liminf_{\delta\to 1}\Big(U_i^{0,t+1}(e,\underline{b}) - U_i^{i,t+1}(e,\underline{b})\Big). \tag{35}$$

Iterating forward Eq. (35), we obtain that for all sufficiently large $\delta < 1$,

$$U_i^{0,t}(e,\underline{b}) - U_i^{i,t}(e,\underline{b}) \geq \frac{\min_{\theta\in\Theta}\lambda^i(\theta)}{\max_{\theta\in\Theta}\lambda^i(\theta)}\varepsilon^2 > 0. \tag{36}$$

Since the lower bound is independent of $(e,\underline{z},t)$ and the players' stage game payoffs are bounded, no profitable one-shot deviation exists in the on-path phase when the players are patient enough.

Player i's punishment phase. For player $j \neq i$, the argument is similar to the on-path phase. Specifically, the analog of Eq. (35) reads as

$$\begin{aligned}\liminf_{\delta\to 1}\Big(U_j^{i,t}(e,\underline{b}) - U_j^{j,t}(e,\underline{b})\Big) &\geq \liminf_{\delta\to 1}\Big(U_j^{0,t}(e,\underline{b}) - U_j^{j,t}(e,\underline{b})\Big) + \liminf_{\delta\to 1}\Big(U_j^{i,t}(e,\underline{b}) - U_j^{0,t}(e,\underline{b})\Big) \\ &\geq \lambda^j(\theta^t)\varepsilon^2 + (1-\lambda^j(\theta^t))\liminf_{\delta\to 1}\Big(U_j^{0,t+1}(e,\underline{b}) - U_j^{j,t+1}(e,\underline{b})\Big) + \\ &+ \lambda^i(\theta^t)\varepsilon^2 + (1-\lambda^i(\theta^t))\liminf_{\delta\to 1}(U_j^{i,t+1}(e,\underline{b}) - U_j^{0,t+1}(e,\underline{b})).\end{aligned} \tag{37}$$

As a result, for for all large $\delta < 1$,

$$U_j^{i,t}(e,\underline{b}) - U_j^{j,t}(e,\underline{b}) \geq \frac{\min_{\theta\in\Theta}\lambda^j(\theta)}{\max_{\theta\in\Theta}\lambda^j(\theta)}\varepsilon^2 + \frac{\min_{\theta\in\Theta}\lambda^i(\theta)}{\max_{\theta\in\Theta}\lambda^i(\theta)}\varepsilon^2 > 0. \tag{38}$$

We now look at player i's incentives in her punishment phase. This is the trickiest case, and we will need this phase to be long enough, i.e., ν is small, so that this player has no profitable deviation.

To begin, let us identify when player i cannot gain from deviating once and miss transitioning to the reward phase at time t. Since this player is minmaxed in her punishment phase, i.e.,

$r_i(a^i(\theta)|\theta) = 0$ for each stage game $\theta \in \Theta$, she cannot gain from any one-shot deviation if and only if

$$\delta \lambda^i(\theta^t) \frac{U_i^{0,t+1}(e, \underline{b}) - U_i^{i,t+1}(e, \underline{b}) - \varepsilon^2}{1-\delta} \geq -u_i(a^i(\theta^t)|\theta^t). \tag{39}$$

Subtract ε^2 in Eq. (34) and iterate it forward to obtain

$$\begin{aligned}\frac{U_i^{0,t+1}(e, \underline{b}) - U_i^{i,t+1}(e, \underline{b}) - \varepsilon^2}{1-\delta} \geq \sum_{\tau=t+1}^{\infty} \left(\varepsilon - \varepsilon^2 - u_i(a^i(\theta^\tau)|\theta^\tau)\right) \prod_{l=t}^{\tau-1} \delta(1-\lambda^i(\theta^l)) + \\ + \sum_{\tau=t}^{\infty} x_i(\theta^\tau) \mathbb{1}_{\{\min_{j\in N} b_j(\theta^{0:\tau}) > -\varepsilon^{-1}\}} \prod_{l=t}^{\tau-1} \delta(1-\lambda^i(\theta^l)).\end{aligned} \tag{40}$$

We shall bound from below each term on the right-hand side of Eq. (40). Recall that $\varepsilon - \varepsilon^2 - u_i(a^i(\theta)|\theta) = \lambda^i(\theta)\nu^{-1}$ for each stage game $\theta \in \Theta$. Then, observe that the first term depends only on e, and it is at least as large as

$$\begin{aligned}\Delta' &:= \inf_{\widetilde{e} \in \Theta^\infty} \sum_{t=0}^{\infty} \lambda(\widetilde{\theta}^t) \cdot \nu^{-1} \prod_{\tau=0}^{t-1} \delta(1-\lambda^i(\widetilde{\theta}^\tau)) = \\ &= \min_{\theta\in\Theta} \lambda^i(\theta) \cdot \nu^{-1} + \delta(1-\lambda^i(\theta))\Delta' = \\ &= \min_{\theta} \left(\frac{\lambda^i(\theta)}{1-\delta(1-\lambda^i(\theta))} \right) \cdot \nu^{-1}.\end{aligned} \tag{41}$$

The second term is a bit more subtle. To begin, let $B \geq \max_{(i,\theta)\in N\times\Theta} \max_{a\in A(\theta)} |u_i(a|\theta)|$ be the bound on stage game payoffs. By definition, $\delta(1-\lambda^i(\theta^t))$ lies between $\delta(1-\max_{\theta\in\Theta} \lambda^i(\theta))$ and $\delta(1-\min_{\theta\in\Theta} \lambda^i(\theta))$; similarly, $x_i(\theta^t)\mathbb{1}_{\{\min_{j\in N} b_j(\theta^{0:\tau}) > -\varepsilon^{-1}\}}$ is an element of $[-B, B]$. It follows that the second term in Eq. (40) is not lower than the value of the following auxiliary program:

$$\begin{aligned}&\inf_{(z,\widetilde{\delta},b)} \sum_{t=0}^{\infty} z_t \mathbb{1}_{\{b_t > -1/\varepsilon\}} \prod_{\tau=0}^{t-1} \widetilde{\delta}_\tau \quad \text{subject to} \\ &b_t = \min\left\{\max\left\{b_{t-1} + z_t, -\varepsilon^{-1}\right\}, \varepsilon^{-1}\right\}, \quad b_{-1} \in [-\varepsilon^{-1}, \varepsilon^{-1}], \\ &\widetilde{\delta}_t \in [\delta(1-\max_{\theta\in\Theta} \lambda^i(\theta)),\ \delta(1-\min_{\theta\in\Theta} \lambda^i(\theta))],\ z_t \in [-B, B] \ \ \forall t.\end{aligned} \tag{42}$$

In this problem, $\widetilde{\delta}_t$ and (b_t, z_t) are proxies for $\delta(1-\lambda^i(\theta^t))$ and $(b_i(\theta^t), x_i(\theta^t))$ in the second term on the right-hand side of (40). However, these auxillary variables $(z, \widetilde{\delta}, b)$ are much less constrained, hence the infimum in (42) is a lower bound on that term.

We now study the infimum in (42). For large values of δ and small values of (ε, ν), it is optimal to start at $b_{-1} = \varepsilon^{-1}$ and maximally frontload negative payments, i.e., $z_t = \min\{b_{t-1} + \varepsilon^{-1}, -B\}$ and $\widetilde{\delta}_t = \delta(1-\min_{\theta\in\Theta} \lambda^i(\theta))$. Since it takes at most $\frac{2\varepsilon^{-1}}{B} - 1$ steps to reach the right endpoint from the left one with increments of B, the value of this auxiliary problem is at least as large as

$$\Delta'' := \frac{1 - \left[\delta(1 - \min_{\theta\in\Theta} \lambda^i(\theta))\right]^{\frac{2\varepsilon^{-1}}{B}-1}}{1 - \delta(1 - \min_{\theta\in\Theta} \lambda^i(\theta))}(-B). \tag{43}$$

To conclude the argument, combine the two bounds devised in Eq. (41) and Eq. (43) to obtain the asymptotic lower bound on the player i's dynamic loss from punishments due to her one-shot deviation at time t:

$$\liminf_{\delta\to 1} \delta\lambda^i(\theta^t)\frac{U_i^{0,t+1}(e, \underline{b}) - U_i^{i,t+1}(e, \underline{b}) - \varepsilon^2}{1-\delta} \geq \liminf_{\delta\to 1} \delta\lambda^i(\theta^t)(\Delta' + \Delta'') =$$
$$= \left(\varepsilon - \varepsilon^2 - u_i(a^i(\theta^t)|\theta^t)\right)\left(1 + \underbrace{\frac{1 - \left[1 - \min_{\theta\in\Theta}(\varepsilon - \varepsilon^2 - u_i(a^i(\theta^t)|\theta)) \cdot \nu\right]^{\frac{2\varepsilon^{-1}}{B}-1}}{\min_{\theta\in\Theta}(\varepsilon - \varepsilon^2 - u_i(a^i(\theta^t)|\theta))}(-B)}_{=o(\nu)}\right). \tag{44}$$

Since the expression in the second line of Eq. (44) converges to $\varepsilon - \varepsilon^2 - u_i(a^i(\theta^t)|\theta^t) > -u_i(a^i(\theta^t)|\theta^t)$ as ν goes to 0, player i has no incentive to deviate in her punishment phase for all large values of δ whenever ν is sufficiently small.

Player $-i$'s reward phase. Let $j \neq i$. Recall that the player j's payoff is larger than her payoff in the on-path phase by exactly ε^2 irrespective of the environment and value of b. It follows that player j cannot profitably deviate in the player $-i$'s reward phase, provided that she cannot do so in the on-path phase.

Finally, observe that, since the players' payoffs are bounded, Eq. (44) implies that player i has no incentives to deviate in the reward phase provided that she cannot profitably deviate in the punishment phase.

Approximating expected payoffs. We have established a strategy profile parametrized by (ε, ν), where both parameters are small and ν is chosen so that the term in Eq. (44) is larger than $-u(a^i(\theta)|\theta)$ for every stage game-player pair. As shown in the previous section, this strategy profile is ex-post perfect for all large values of δ. In what follows we shall show that it can approximate w in expectation.

For each $i \in N$, let T_i be the (stopping) time when $b_i(\theta^{0:t})$ reaches either of the endpoints, i.e., $\pm\varepsilon^{-1}$. Since (θ^t) is a stationary reversible Markov chain, for each t, the distributions of $(\theta^0, \ldots, \theta^t)$ and $(\theta^t, \ldots, \theta^0)$ coincide. By Phatarfod et al. [1971], the probability that $b_i(\theta^{0:t})$ takes a strictly negative value at time t is the same as the probability that the (Markov) random walk with increments $(x_i(\theta))_{\theta\in\Theta}$ and absorbing barriers at $\pm\varepsilon^{-1}$ is strictly negative at time t, that is

$$\mathbb{P}\left(b_i(\theta^{0:t}) < 0\right) = \mathbb{P}\left(b_i(\theta^{0:t\wedge T_i}) < 0\right) \quad \forall t. \tag{45}$$

Using Eq. (45), we now are going to bound from above the probability of hitting the lower boundary in the long-run. Suppose first that $x_i(\theta) \geq 0$ for all $\theta \in \Theta$. Since $\sum_{\theta\in\Theta} x_i(\theta)\mu(\theta|\pi) > 0$,

$$\lim_{t\to\infty} \mathbb{P}\left(b_i(\theta^{0:t}) = -\varepsilon^{-1}\right) \leq \lim_{t\to\infty} \mathbb{P}\left(b_i(\theta^{0:t\wedge T_i}) < 0\right) = 0. \tag{46}$$

Suppose next that $x_i(\theta) < 0$ for some $\theta \in \Theta$. For $\rho \in \mathbb{R}$, let $\xi_i(\rho)$ be the Perron's root of the matrix $\left(\pi(\theta,s)e^{x_i(s)\rho}\right)_{(\theta,s)\in\Theta\times\Theta}$ and denote the corresponding eigenvector by $v_i(\rho)$. By Theorems 1 and 3 in Miller [1961] and Lemma 1 in Sadowsky [1989],

(i) $\xi_i(\rho)$ is analytic and strictly convex;

(ii) $\frac{\partial}{\partial\rho}\xi_i(0) = \sum_{\theta\in\Theta} x_i(\theta)\mu(\theta|\pi) > 0$ and $\lim_{|\rho|\to\infty}\xi_i(\rho) = \infty$;

(iii) there are $0 < \underline{v}_i \leq \overline{v}_i < \infty$ such that $v_i(\rho) \in [\underline{v}_i, \overline{v}_i]$ for all $\rho \in \mathbb{R}$.

Let ρ_i be the point of a minimum of $\xi_i(\rho)$, which is well-defined and negative due to Properties (i) and (ii). By Lemma 3 in Sadowsky [1989], T_i is finite with probability one. Then, Property (iii) and Lemma 5 in Sadowsky [1989] imply

$$\lim_{t\to\infty} \mathbb{P}\left(b_i(\theta^{0:t}) = -\varepsilon^{-1}\right) \leq \lim_{t\to\infty} \mathbb{P}\left(b_i(\theta^{0:t\wedge T_i} < 0\right) \leq \frac{\overline{v}_i}{\underline{v}_i} e^{\rho_i \varepsilon^{-1}}. \tag{47}$$

It follows from Eq. (46) and Eq. (47) that there exist $c > 0$ and $\overline{\rho} < 0$, independent of ε, such that

$$\lim_{t\to\infty} \mathbb{P}\left(b_i(\theta^{0:t}) = -\varepsilon^{-1}\right) \leq c e^{\overline{\rho}\varepsilon^{-1}} \quad \forall i \in N. \tag{48}$$

To sum up, the long-run probability with which the original random walk, i.e., one with reflecting barriers at $\pm\varepsilon^{-1}$, hits the lower boundaries can be made arbitrarily small as $\varepsilon \to 0$.

We now use the aforementioned bound on the long-run probability of hitting $-\varepsilon^{-1}$ to show that the player i's expected payoff can be made arbitrarily close to w_i as $\delta \to 1$ and $\varepsilon \to 0$. First of all, remark that $\mathbb{E}[x_i(\theta^t)] = \sum_{\theta\in\Theta} x_i(\theta)\mu(\theta|\pi) = w_i$. Therefore, since the payoffs are bounded uniformly across players and stage games by some $B > 0$, the expected value of $U_i^{0,0}(e,0)$, which is defined in Eq. (32), satisfies

$$\left|\mathbb{E}\left[U_i^{0,0}(e,0)\right] - w_i - \varepsilon\right| \leq \mathbb{E}\left[(1-\delta)\sum_{t=0}^{\infty}\delta^t \mathbb{1}_{\{\min_{j\in N} b_j(\theta^{0:t}) = -\varepsilon^{-1}\}}\right](2B). \tag{49}$$

Take $\delta \to 1$ in Eq. (49) and use the bound in Eq. (48) to obtain

$$\limsup_{\delta\to 1}\left|\mathbb{E}\left[U_i^{0,0}(e,0)\right]-w_i-\varepsilon\right|\le\left(1-\left[1-ce^{\overline{\rho}/\varepsilon}\right]^n\right)(2B). \tag{50}$$

Since the set of robust limit equilibrium payoffs is closed, the result follows from letting $\varepsilon\to 0$; of course, in that process ν has to be suitably adjusted as discussed in the previous section so that the strategy profile is ex-post perfect for all large values of δ. □

8.2 Appendix: Robustness of on-path behavior in dynamic games

Proof of Lemma 1. First of all, note that the operator $\mathbf{T}_{\underline{u}}$ is monotone, and it maps bounded subsets of $\mathbb{R}^n$ to bounded subsets of $\mathbb{R}^n$. By the standard argument (e.g., see Abreu, Pearce and Stacchetti [1990]), $\mathbf{T}_{\underline{u}}$ admits the largest bounded subinvariant, say $\Gamma(\underline{u})$, which includes Γ. We shall show that every element of

$$\mathscr{G}(x):=\left\{w\in\mathbb{R}^{n\cdot\Theta^\infty}|\exists\gamma\in\Gamma(x)\text{ s.t. }\gamma=w(e)-(1-\delta)\sum_{t=0}^{\infty}\delta^t\underline{u}(\theta^t)\ \forall e\in\Theta^\infty\right\} \tag{51}$$

is an (ex-post) XPE payoff, which would then imply Eq. (16).

Indeed, as is well-known, the whole set of (ex-post) XPE equilibrium payoffs is the largest (uniformly) bounded subinvariant of the following APS recursion:

$$\begin{aligned}\mathbf{F}\mathscr{V}=\Big\{&w\in\mathbb{R}^{n\cdot\Theta^\infty}|\forall\theta\in\Theta,\ \exists\beta(\theta)\in\Delta A(\theta),(v^i(\theta))_{i=0}^n\subset\mathscr{V}^{A(\theta)}\text{ s.t.}\\ &w(\theta,e)=(1-\delta)u(\beta(\theta)|\theta)+\delta\textstyle\sum_{a\in A(\theta)}v^0(e|a,\theta)\beta(a|\theta)\ \ \forall(\theta,e)\in\Theta^\infty,\\ &\left(v_i^0(e|a,\theta)-v_i^i(e|a,\theta)-\tfrac{1-\delta}{\delta}d_i(a|\theta)\right)\beta(a|\theta)\ge 0\ \ \forall a\in A(\theta),\ \forall(\theta,e)\in\Theta^\infty,\ \forall i\in N\Big\}.\end{aligned} \tag{52}$$

Substitute $\mathscr{G}(x)$ into Eq. (52) to obtain

$$\begin{aligned}\mathbf{F}\mathscr{G}(x)=\Big\{&w\in\mathbb{R}^{n\cdot\Theta^\infty}|\forall\theta\in\Theta,\ \exists\beta(\theta)\in\Delta A(\theta),(g^i(\theta))_{i=0}^n\subset\Gamma(\underline{u})^{A(\theta)}\text{ s.t.}\\ &w(\theta,e)-(1-\delta)(\underline{u}(\theta)+\delta\textstyle\sum_{t=0}^\infty\delta^t\underline{u}(\theta^t))=(1-\delta)(u(\beta(\theta)|\theta)-\underline{u}(\theta))+\\ &\qquad+\delta\textstyle\sum_{a\in A(\theta)}g^0(a,\theta)\beta(a|\theta)\ \ \forall(\theta,e)\in\Theta^\infty,\\ &\left(g_i^0(a,\theta)-g_i^i(a,\theta)-\tfrac{1-\delta}{\delta}d_i(a|\theta)\right)\beta(a|\theta)\ge 0\ \ \forall a\in A(\theta),\ \forall\theta\in\Theta,\ \forall i\in N\Big\},\end{aligned} \tag{53}$$

which gives

$$\mathbf{F}\mathscr{G}(x)\supseteq\mathscr{G}(x). \tag{54}$$

$\square$

Proof of Theorem 3. To begin, let $\Gamma(\underline{u})$ be the largest bounded subinvariant of $\mathbf{T}_{\underline{u}}$, which exists by the argument in the proof of Lemma 1. We claim that 0 is an element of $\Gamma(\underline{u})$ for every $\underline{u} \in \prod_{\theta\in\Theta}(Int(\mathscr{U}(\theta)) \cap \mathbb{R}^n_{++})$ when δ is large enough.

Indeed, let $G \subset \mathbb{R}^n$ be a closed ball centered at 0 that satisfies

$$G + \underline{u}(\theta) \subseteq Int(\mathscr{U}(\theta)) \cap \mathbb{R}^n_{++} \;\; \forall\theta \in \Theta. \tag{55}$$

It is easy to see that, for each $\theta \in \Theta$, the set $G + \underline{u}(\theta)$ is decomposable on tangent hyperplanes in the terminology of Fudenberg et al. [1994]. As a result, Theorem 4.1 and Lemma 4.2 in this paper imply existence of $\delta(\underline{u}) < 1$ such that for every $\delta \geq \delta(\underline{u})$,

$$G \supseteq \mathbf{T}_{\underline{u}} G. \tag{56}$$

By definition, $\Gamma(\underline{u})$ is the largest subinvariant, thus $G \subseteq \Gamma(\underline{u})$.

Constructing robust punishments. Let $\varepsilon > 0$, and note that the conditions of Theorem 3 imply that there are $(\underline{u}^i)_{i\in N} \subset \prod_{\theta\in\Theta}(Int(\mathscr{U}(\theta)) \cap \mathbb{R}^n_{++})$ such that

$$\underline{u}^i_i(\theta) \leq \varepsilon \;\; \forall\theta \in \Theta, \; \forall i \in N. \tag{57}$$

It follows from the above argument that for $\delta \geq \delta^\varepsilon := \max_{i\in N} \delta(\underline{u}^i)$, there are XPE $(\sigma^i)_{i\in N}$ satisfying

$$U_i^{\sigma^i}(e) \leq \varepsilon \;\; \forall e \in \Theta^\infty, \; \forall i \in N. \tag{58}$$

Verifying sufficiency of ε-individual rationality. Having established that the players can be robustly minmaxed for large values of δ, the proof of Theorem 3 is rather immediate. Note that the one-shot deviation applies to XPE because it applies to SPE of every dynamic game e. Therefore, an outcome α is justifiable if and only if for every on-path history $\widetilde{h}^t$ there exist XPE $(\sigma^i(\widetilde{h}^t))_{i=1}^n$ such that no player i can gain by the best-one shot deviation given that she will be punished by $\sigma^i(\widetilde{h}^t)$, that is

$$U_i^\alpha(\widetilde{h}^t|e) \geq (1-\delta) r_i(\alpha(\widetilde{h}^t)|\theta^t) + \delta U_i^{\sigma^i(\widetilde{h}^t)}(\theta^{t+1:\infty}) \;\; \forall\theta^{t+1:\infty}, \; \forall i \in N. \tag{59}$$

The fact that ε-individually rational outcomes can satisfy Eq. (59) for all large values of δ directly follows from the claim in Eq. (58). $\square$

Proof of Corollary 1. The proof is based on a simple extension of Lemma 1. Specifically, we shall use a recursive argument and construct XPE in which $\mathbb{E}_{t-1}[U^\sigma(h^t|e)]$ is independent of $\theta^{t+1:\infty}$. Contrast this to Eq. (14) where the normalized payoff is also required to be independent of this period stage game θ^t.

Define $\mathcal{G}$ as follows:

$$\mathcal{G} := \left\{ w \in \mathbb{R}^{n\cdot\Theta^\infty} | w(\theta,\gamma) - \gamma(\theta) = 0 \ \ \forall(\theta,e) \in \Theta^\infty \text{ for some } \gamma \in \prod_{\theta\in\Theta} G(\theta) \right\}, \text{ where } \forall\theta\in\Theta,$$
$$G(\theta) := \operatorname{Conv}\left(\left\{ g^1,\dots,g^n, (1-\delta)(r(a^1(\theta)|\theta) + \xi^1(\theta)),\dots,(1-\delta)(r(a^n(\theta)|\theta)+\xi^n(\theta)) \right\} \right). \tag{60}$$

The following figure illustrates the construction of $G(\theta)$. Note that, due to Conditions (ii) and (iii), every vector $\frac{1-\delta}{\delta}(d_i(a^i(\theta)|\theta) + \xi^i(\theta))$ is an element $\bigcap_{s\in\Theta} G(s)$ for all large values of δ.

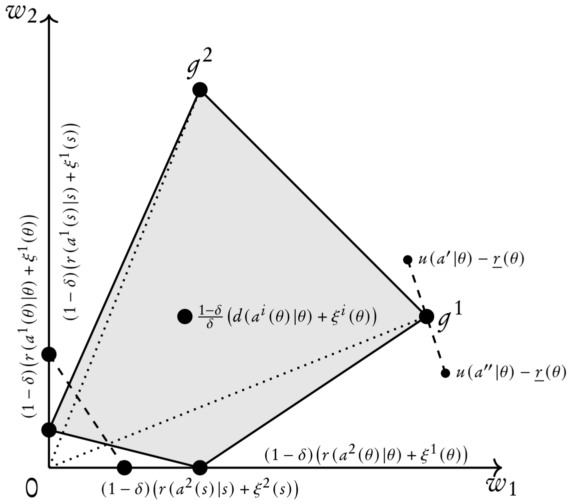


Figure 5: Illustrating the construction of $G(\theta)$.

We claim that the set $\mathcal{G}$ is a subinvariant with respect to the APS recursion $\mathbf{F}$, which is defined in Eq. (52), provided that δ is large enough. Since $\mathcal{G}$ is convex, it suffices to show every extreme point of this set is an element of $\mathbf{F}\mathcal{G}$.

- Consider $w(\theta,e) = g^i$. Condition (i) ensures that g^i can be attained in every stage game θ by some $\beta^i(\theta) \in \Delta A(\theta)$. So, in terms of Eq. (52), set $\beta(\theta) = \beta^i(\theta)$ and $v^0(e|a,\theta) = g^i$ and $v^j(e|a,\theta) = (1-\delta)(r(a^j(\theta^1)|\theta^1) + \xi^j(\theta^1))$ for $j = 1,\dots,n$. Since g^i is strictly positive and the players' payoffs are bounded, no profitable deviation exists for all large values of δ.

- Consider $w(\theta,e) = (1-\delta)(r(a^i(\theta)|\theta) + \xi^i(\theta^1))$. Again, in terms of Eq. (52), set $\beta(\theta)$ to put all mass on $a^i(\theta)$, $v^0(e|a,\theta) = \frac{1-\delta}{\delta}(d(a^i(\theta)|\theta) + \xi^i(\theta^1))$ and $v^j(e|a,\theta) = (1-\delta)(r(a^j(\theta^1)|\theta^1) - \xi^j(\theta^1))$ for $j = 1,\dots,n$. By construction, no player can profitably deviate as $\xi^i(\theta) \geq 0$. As shown in Figure 5, $v^0(\cdot|a,\theta)$ is an element $\bigcap_{s\in\Theta} G(s)$, thus a feasible continuation payoff, for all large values of δ.

□

8.3 Appendix: Symmetric games

Proof of Theorem 4. Let g_* be the largest incentive gap between two SSXPE that can be sustained irrespective of future stage games, that is

$$g_* := \sup_{\sigma',\sigma''} \inf_{e\in\Theta^\infty} (U_1^{\sigma'}(e) - U_1^{\sigma''}(e)) \text{ s.t. } \sigma', \sigma'' \text{ are SSXPE}, \tag{61}$$

where $g_* = -\infty$ if no SSXPE exists.

We now show that $g_* \leq g^*$, where g^* is defined in Theorem 4. By definition, in any SSXPE σ at each history we must have $\frac{1-\delta}{\delta} d(\sigma(h^t)|\theta^t) \leq g_*$; otherwise, there is some continuation environment e in which a deviation from $\sigma(h^t)$ cannot be deterred in SSXPE. It follows that the players' stage game payoff $u_1(\sigma(h^t)|\theta^t)$ at this history h^t is not higher than $\overline{m}(g_*|\theta^t)$, thus

$$U_1^\sigma(e) \leq (1-\delta)\sum_{t=0}^{\infty}\delta^t \overline{m}(g_*|\theta^t). \tag{62}$$

Similarly, since each player can deviate from $\sigma(h^t)$ obtaining $r_1(\sigma(h^t)|\theta^t)$, her ex-post payoff satisfies

$$U_1^\sigma(e) \geq (1-\delta)\sum_{t=0}^{\infty}\delta^t \underline{m}(g_*|\theta^t). \tag{63}$$

Combine Eq. (62), (63) and the definition of g_* to obtain

$$g_* \leq \inf_{\theta\in\Theta}(\overline{m}(g_*|\theta) - \underline{m}(g_*|\theta)), \tag{64}$$

which proves that $g_* \leq g^*$.

Attaining the bound on the incentive gap. Clearly, if g^* is negative, then g_* is negative as well; thus, no SSXPE exists. So, suppose that g^* is non-negative. We now show this bound is tight and construct two SSXPE satisfying Eq. (20).

To begin, take some symmetric normalizing vector $\underline{u}$, and note that $\mathbf{T}_{\underline{u}}$ applied to $G = \{g \in [\underline{g}, \overline{g}]^n | g_1 = \ldots = g_n\}$, where $\overline{g} \geq \underline{g}$, with an additional restriction that $a_1 = \ldots = a_n$ gives the set $\{\gamma \in [\underline{\gamma}, \overline{\gamma}]^n | \gamma_1 = \ldots = \gamma_n\}$, where

$$\overline{\gamma} = \inf_{\theta\in\Theta}(1-\delta)(\overline{m}(\overline{g}-\underline{g}|\theta) - x_1(\theta)) + \delta\overline{g}, \quad \underline{\gamma} = \sup_{\theta\in\Theta}(1-\delta)(\underline{m}(\overline{g}-\underline{g}|\theta) - x_1(\theta)), \tag{65}$$

provided that $\overline{\gamma} \geq \underline{\gamma}$.

We shall devise two choices of $\underline{u}_1$ so that $G = \mathbf{T}_{\underline{u}} G$ and use this fact in conjunction with Lemma 1 to establish existence of SSXPE in Eq. (20).

- First, let $(\overline{g}, \underline{g}) = (g^*, 0)$ and select the normalization so that $\underline{\gamma}$ is guaranteed to be 0 in Eq. (65), i.e., for each $\theta \in \Theta$, set $\underline{u}_1(\theta) = \underline{m}(g^*|\theta)$. By definition of g^*, $\overline{\gamma} = g^*$ in Eq. (65); therefore, by Lemma 1, there is an SSXPE in which the players obtain the payoff of $(1-\delta)\sum_{t=0}^{\infty} \delta^t \underline{m}(g^*|\theta^t)$.

- Second, let $(\overline{g}, \underline{g}) = (0, -g^*)$ and select the normalization so that $\overline{\gamma}$ is guaranteed to be 0 in Eq. (65), i.e., for each $\theta \in \Theta$, set $\underline{u}_1(\theta) = \overline{m}(g^*|\theta)$. The reader can verify that $\underline{\gamma} = -g^*$ in Eq. (65); therefore, by Lemma 1, there is an SSXPE in which the players obtain the payoff of $(1-\delta)\sum_{t=0}^{\infty} \delta^t \overline{m}(g^*|\theta^t)$.

□

Proof of Corollary 2. Let $\overline{\delta} < 1$ be such that

$$\frac{1-\overline{\delta}}{\overline{\delta}} \max_{a_1 \in A_1(\theta), \theta \in \Theta} d_1(a_1, \dots, a_1|\theta) \leq \min_{\theta \in \Theta}(\overline{m}(\infty|\theta) - \underline{m}(\infty|\theta)). \tag{66}$$

Such $\overline{\delta}$ exists because the term on the right-hand side is positive. It is immediate from the definition of g^* that for all $\delta \geq \overline{\delta}$, the gap g^* equals to the term on the right-hand side in Eq. (66); moreover, $\overline{m}(\infty|\theta) = \overline{m}(g^*|\theta)$ and $\underline{m}(\infty|\theta) = \underline{m}(g^*|\theta)$. The corollary follows from Theorem 4, specifically Eq. (20). □

An outcome α is **strongly symmetric** if at each public history, the players take an identical action, i.e., $\sigma_1(\widetilde{h}^t) = \dots = \sigma_n(\widetilde{h}^t)$. We say that a strongly symmetric outcome α is **robustly justifiable in SSXPE** if there is an SSXPE in which the players' on-path actions are specified by α.

Corollary 3. *Suppose $\overline{m}(\infty|\theta) - \underline{m}(\infty|\theta) > 0$ for all $\theta \in \Theta$. Then, there exists $\overline{\delta} < 1$ such that for every $\delta \geq \overline{\delta}$, a strongly symmetric outcome α is robustly justifiable in SSXPE if and only it is 0-individually rational in the uncertain repeated game $(\Theta, (A(\theta), u(\cdot|\theta) - \underline{m}(\infty|\theta))_{\theta\in\Theta}, \delta)$.*

Proof. By the argument in the proof of Corollary 2, there exists $\overline{\delta} < 1$ such that for all large values of $\delta \geq \overline{\delta}$, the worst SSXPE exists and it gives exactly $(1-\delta)\sum_{t=0}^{\infty} \delta^t \underline{m}(\infty|\theta^t)$ to the players. So, a strongly symmetric outcome α is robustly justifiable in SSXPE if and only if

$$U_1^{\alpha}(\widetilde{h}^t|e) \geq (1-\delta) r_i(\alpha(\widetilde{h}^t)|\theta^t) + \delta(1-\delta) \sum_{\tau=t+1}^{\infty} \delta^{\tau-(t+1)} \underline{m}(\infty|\theta^{\tau}), \tag{67}$$

which is exactly 0-individually rationality in $(\Theta, (A(\theta), u(\cdot|\theta) - \underline{m}(\infty|\theta))_{\theta\in\Theta}, \delta)$.

□

References

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

D. Abreu, D. Pearce, and E. Stacchetti. Optimal cartel equilibria with imperfect monitoring. *Journal of Economic Theory*, 39(1):251–269, 1986.

D. Abreu, D. Pearce, and E. Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58(5):1041–1063, 1990.

D. Abreu, B. Brooks, and Y. Sannikov. Algorithms for stochastic games with perfect monitoring. *Econometrica*, 88(4):1661–1695, 2020.

D. Bergemann and S. Morris. Robust mechanism design. *Econometrica*, 73(6):1771–1813, 2005.

G. Carroll. Dynamic incentives in incompletely specified environments. University of Toronto, 2021.

V. V. Chari and P. J. Kehoe. Sustainable plans. *Journal of Political Economy*, 98(4):783–802, 1990.

S. Chassang. Calibrated incentive contracts. *Econometrica*, 81(5):1935–1971, 2013.

M. B. Cronshaw and D. G. Luenberger. Strongly symmetric subgame perfect equilibria in infinitely repeated games with perfect monitoring and discounting. *Games and Economic Behavior*, 6(2): 220–237, 1994.

P. Dasgupta and E. Maskin. The existence of equilibrium in discontinuous economic games, part i (theory). *The Review of Economic Studies*, 53(1):1–26, 1986.

H. de Oliveira and R. Lamba. Rationalizing dynamic choices. Fundação Getúlio Vargas-EESP and Pennsylvania State University, 2023.

J. C. Ely, J. Hörner, and W. Olszewski. Belief-free equilibria in repeated games. *Econometrica*, 73 (2):377–415, 2005.

L. G. Epstein and M. Schneider. Recursive multiple-priors. *Journal of Economic Theory*, 113(1): 1–31, 2003.

D. Fudenberg and E. Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

D. Fudenberg and Y. Yamamoto. Repeated games where the payoffs and monitoring structure are unknown. *Econometrica*, 78(5):1673–1710, 2010.

D. Fudenberg and Y. Yamamoto. The folk theorem for irreducible stochastic games with imperfect public monitoring. *Journal of Economic Theory*, 146(4):1664–1683, 2011.

D. Fudenberg, D. M. Kreps, and E. S. Maskin. Repeated games with long-run and short-run players. *The Review of Economic Studies*, 57(4):555–573, 1990.

D. Fudenberg, D. Levine, and E. Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1039, 1994.

A. Gershkov, J. K. Goeree, A. Kushnir, B. Moldovanu, and X. Shi. On the equivalence of bayesian and dominant strategy implementation. *Econometrica*, 81(1):197–220, 2013.

I. Gilboa and D. Schmeidler. Maxmin expected utility with non-unique prior. *Econometrica*, 18 (2):141–153, 1989.

E. J. Green and R. H. Porter. Noncooperative collusion under imperfect price information. *Econometrica*, 52(1):87–100, 1984.

J. Haltiwanger and J. E. Harrington, Jr. The impact of cyclical demand movements on collusive behavior. *The RAND Journal of Economics*, 22(1):89–106, 1991.

E. Hanany and P. Klibanoff. Updating preferences with multiple priors. *Theoretical Economics*, 2: 261–298, 2007.

J. Hörner and S. Lovo. Belief-free equilibria in games with incomplete information. *Econometrica*, 77(2):453–487, 2009.

J. Hörner, T. Sugaya, S. Takahashi, and N. Vieille. Recursive methods in discounted stochastic games: An algorithm for $\delta \to 1$ and a folk theorem. *Econometrica*, 79(4):1277–1318, 2011.

M. Kandori. Correlated demand shocks and price wars during booms. *The Review of Economic Studies*, 58(1):171–180, 1991.

A. Kleiner and B. Moldovanu. Content-based agendas and qualified majorities in sequential voting. *American Economic Review*, 107(6):1477–1506, 2017.

N. R. Kocherlakota. Implications of efficient risk sharing without commitment. *The Review of Economic Studies*, 63(4):595–609, 1996.

R. Kostadinov. Worst-case regret in ambiguous dynamic games. McMaster University, 2023.

I. Krasikov and R. Lamba. Uncertain repeated games for fixed discounting. Higher School of Economics and Pennsylvania State University, 2023.

J. Levin. Relational incentive contracts. *American Economic Review*, 93(3):835–857, 2003.

J. A. Libgober and X. Mu. Coasian dynamics under informational robustness. University of Southern California and Princeton University, 2023.

F. Maccheroni, M. Marinacci, and A. Rustichini. Dynamic variational preferences. *Journal of Economic Theory*, 128(1):4–44, 2006a.

F. Maccheroni, M. Marinacci, and A. Rustichini. Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, 74(6):1447–1498, 2006b.

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

H. Miller. A generalization of wald's identity with applications to random walks. *The Annals of Mathematical Statistics*, pages 549–560, 1961.

V. Mirrokni, R. P. Leme, P. Tang, and S. Zuo. Non-clairvoyant dynamic mechanism design. *Econometrica*, 88(5):1939–1963, 2020.

M. Panov. Agreements of continuous-time games. Higher School of Economics, 2022.

A. Penta. Robust dynamic implementation. *Journal of Economic Theory*, 160:280–316, 2015.

R. Phatarfod, T. Speed, and A. Walker. A note on random walks. *Journal of Applied Probability*, 8 (1):198–201, 1971.

J. J. Rotemberg and G. Saloner. A supergame-theoretic model of price wars during booms. *American Economic Review*, 76(3):390–407, 1986.

J. S. Sadowsky. A dependent data extension of wald's identity and its application to sequential test performance computation. *IEEE transactions on information theory*, 35(4):834–842, 1989.

L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Science*, 39(10):1095–1100, 1953.

E. Solan and N. Vieille. Stochastic games. *Proceedings of the National Academy of Science*, 112(45): 13743–13746, 2015.

Y. Yamamoto. Stochastic games with hidden states. *Theoretical Economics*, 14:1115–1167, 2019.

9 Supplementary appendix

Relationship between XPE and SPE under dynamic variational preferences.

In this section, we elaborate on the connection between the criterion of ex-post perfectness and SPE when players have variation preferences of Maccheroni, Marinacci, and Rustichini [2006a,b]. The similar point in the context of justifiability of outcome has been made in Carroll [2021].

According to this model, the player i's conditional preferences over strategy profiles after observing $\theta^{0:t}$ are described by the so-called ambiguity index $\pi \in \Delta\Theta^\infty \mapsto C_i(\pi|\theta^{0:t}) \in \mathbb{R}_+ \cup \{\infty\}$. The ambiguity index captures the player i's costs of the choice of probability over future stage games, and it is required to be closed, lower-semy continuous and grounded, i.e., finite for some probabilities. [23] Then, player i evaluates strategy profiles according to the following criterion:

$$\inf_{\pi\in\Delta\Theta^\infty}\Big(\mathbb{E}^\pi[U_i^\sigma(h^t|e)] + C_i(\pi|\theta^{0:t})\Big). \tag{68}$$

Now, let σ be ex-post perfect. We claim that player i plays best response at each history irrespective of her dynamic ambiguity index. To see it, assume, by contradiction, that there exists some ambiguity index C_i, history h^t and another strategy $\widetilde{\sigma}_i$ so that it is profitable for player i to deviate from σ_i

$$\inf_{\pi\in\Delta\Theta^\infty}\Big(\mathbb{E}^\pi[U_i^\sigma(h^t|e)] + C_i(\pi|\theta^{0:t})\Big) < \inf_{\pi\in\Delta\Theta^\infty}\Big(\mathbb{E}^\pi[U_i^{(\widetilde{\sigma}_i,\sigma_{-i})}(h^t|e)] + C_i(\pi|\theta^{0:t})\Big). \tag{69}$$

It follows that there exists some probability $\widetilde{\pi}$ so that

$$\begin{aligned}\mathbb{E}^{\widetilde{\pi}}[U_i^\sigma(h^t|e)] + C_i(\widetilde{\pi}|\theta^{0:t}) < \inf_{\pi\in\Delta\Theta^\infty}\Big(\mathbb{E}^\pi[U_i^{(\widetilde{\sigma}_i,\sigma_{-i})}(h^t|e)] + C_i(\pi|\theta^{0:t})\Big) \leq \\ \leq \mathbb{E}^{\widetilde{\pi}}[U_i^{(\widetilde{\sigma}_i,\sigma_{-i})}(h^t|e)] + C_i(\widetilde{\pi}|\theta^{0:t}),\end{aligned} \tag{70}$$

which contradicts ex-post perfectness of σ.

Conversely, we claim that σ is ex-post perfect if each player plays a best-response at every history for all possible dynamic variational preferences she might have. To see, by way of contradiction, assume that there exists some player i that has a profitable one-shot deviation in some environment e at some history h^t. Define this player ambiguity index C_i as follows: for each $s = 0, 1, \ldots$, set $C_i(\pi|\widetilde{\theta}^{0,s})$ is zero if $\widetilde{\theta}^{0,s} = \theta^{0:s}$ and if π assigns probability of one to events of the form $\{\theta^0\}\times\ldots\times\{\theta^\tau\}\times\Theta^\infty$, and $\Delta\Theta^\infty$ for all $\tau \geq s$; otherwise, $C_i(\pi|\widetilde{\theta}^{0,s})$ is B, where B is large.[24]

[23] Here, Θ^∞ is endowned with the natural algebra generated by sets of the form $\{\theta^0\}\times\ldots\times\{\theta^t\}\times\Theta^\infty$, and $\Delta\Theta^\infty$ stays for the space of finitely-additive probability measures equipped with the weak-* topology as a subset of the topological dual of the space of simple functions endowned with the sup norm.

[24] This ambiguity index is lower-semi continuous, because $\Delta\Theta$ is Hausdorff, hence singletons are closed. Maccheroni et al. [2006a] in addition requires the effective domain of $C_i(\cdot|\widetilde{\theta}^{0:s})$ to contain some non-degenerate elements that assign positive probabilities to each possible realization of $\theta^{s+1:\tau}$ for all $\tau \geq s+1$. Our proposed ambiguity index trivially satisfies this condition.

Since the players' stage game payoffs are bounded, at h^t, for any strategy player i's strategy $\widetilde{\sigma}_{-i}$, the infimum in

$$\inf_{\pi\in\Delta\Theta^\infty}\left(\mathbb{E}^\pi[U_i^{(\widetilde{\sigma}_i,\sigma_{-i})}(h^t|\widetilde{e})]+C_i(\pi|\widetilde{\theta}^{0:t})\right). \tag{71}$$

is attained by π that picks e with probability one. So, the infimum in Eq. (71) equals to $U_i^{(\widetilde{\sigma}_i,\sigma_{-i})}(h^t|e)$. As a result, since σ_i is not an ex-post best response at h^t, the player i can profitably deviate at this history when her preferences are given by Eq. (71).

Symmetric games satisfy condition for Theorem 3.

Lemma 2. *Consider a symmetric game (A,u). If $Int(\mathscr{U})\cap\mathbb{R}^n_{++}$ is nonempty, then $\bigwedge(\mathscr{U}\cap\mathbb{R}^n_+)=0$.*

Proof. The result is immediate if $\mathscr{U}\cap\{w\in\mathbb{R}^n|w_1=0\}$ contains a nonnegative point, so suppose that it does not. Let w be an arbitrary point in this set, and denote $\beta\in\Delta A$ that gives w, i.e., $w=\sum_{a\in A}u(a)\beta(a)$. Denote the set of pertumations of players by $\mathcal{I}$, then, by symmetry,

$$\sum_{a\in A}\frac{\sum_{\iota\in\mathcal{I}:\iota(1)=1}u_i(a_{\iota(i)},a_{\iota(-i)})}{\sum_{\iota\in\mathcal{I}:\iota(1)=1}}\beta(a)=\begin{cases}0 & i=1,\\ \frac{1}{n-1}\sum_{j=2}^n w_j & i\neq 1.\end{cases} \tag{72}$$

It follows from Eq. (72) that $(0,x,\dots,x)$ is an element of $\mathscr{U}$ with $x:=\frac{1}{n-1}\sum_{j=2}^n w_j$. Note that $x<0$ due to our assumption that no nonnegative point in $\mathscr{U}\cap\{w\in\mathbb{R}^n|w_1=0\}$ exists.

The interiority hypothesis implies existence of a point $\widetilde{w}\in\mathscr{U}\cap\mathbb{R}^n_{++}$. Repeating the same construction as in Eq. (72) but now averaging over all permutations, we obtain that $(\widetilde{x},\widetilde{x},\dots,\widetilde{x})$ with $\widetilde{x}:=\frac{1}{n}\sum_{i=1}^n\widetilde{w}_i$ is an element of $\mathscr{U}$. By construction, $\widetilde{x}>0$, which implies that 0 is in $\mathscr{U}$ as a convex combination of $(\widetilde{x},\widetilde{x},\dots,\widetilde{x})$, $(0,x,\dots,x)$, $(x,0,\dots,x)$, ..., $(x,x,\dots,0)$. □

Equivalence of XPE with one long-run player and SSXPE with several long-run players.

Consider an uncertain repeated game $(\Theta,(A(\theta),u(\cdot|\theta))_{\theta\in\Theta},\delta)$ in which player 1 is long-lived and the other players are short-lived. Let σ be an XPE. Clearly, the short-lived players must be playing a static best-response at each history h^t, that is $\sigma(h^t)\in A^*(\theta^t)$, which is defined by

$$A^*(\theta):=\{a\in A(\theta)|d_i(a|\theta)=0\ \forall i\neq 1\}\quad\forall\theta\in\Theta. \tag{73}$$

Define a symmetric uncertain repeated game $(\Theta,(\widetilde{A}(\theta),\widetilde{u}(\cdot|\theta))_{\theta\in\Theta},\delta)$ with two long-lived players as follows:

$$\widetilde{A}_i(\theta):=A^*(\theta),\quad u_i(\widetilde{a}|\theta):=\widetilde{u}_1(\widetilde{a}_i|\theta)\mathbb{1}_{\{\widetilde{a}_1=\widetilde{a}_2\}}+r_1(\widetilde{a}_{-i}|\theta)\mathbb{1}_{\{\widetilde{a}_1\neq\widetilde{a}_2\}}\ \text{ for } i=1,2. \tag{74}$$

Since the long-lived player cannot profitably deviate in the original game and two players in the new game are effectively replicas of him, (σ, σ) is a SSXPE of this symmetric uncertain repeated game.

Conversely, let σ be a SSXPE of $(\Theta, (A(\theta), u(\cdot|\theta))_{\theta\in\Theta}, \delta)$ in which every player is long-lived. Define an uncertain repeated game $(\Theta, (\widetilde{A}(\theta), \widetilde{u}(\cdot|\theta))_{\theta\in\Theta}, \delta)$ with long-lived player 1 and short-lived player 2 as follows: $\widetilde{A}(\theta) = A(\theta)$, and

$$\widetilde{u}_1(a|\theta) := u_1(a|\theta), \quad \widetilde{u}_2(a|\theta) := -\mathbb{1}_{\{a_1 \neq a_2\}}. \tag{75}$$

It is always in player 2's interest to match player 1's action. Since σ is a SSXPE of the original game, player 1 cannot profitably deviate provided that player 2 will match her action; thus, (σ_1, σ_2) is an XPE of the new game.

Example 6. *Consider a repeated game with three players, each has two actions. As usual, player is choosing a row, player 2 is selecting a column and player 3 is choosing a table.*

	a_3': a_2'	a_3': a_2''		a_3'': a_2'	a_3'': a_2''
a_1'	$1/2, 1/2, 1/2$	$0^\star, 2^\star, 0^\star$	a_1'	$0^\star, 0^\star, 2^\star$	$5/2^\star, -1, -1$
a_1''	$2^\star, 0^\star, 0^\star$	$-1, -1, 5/2^\star$	a_1''	$-1, 5/2^\star, -1$	$1/2, 1/2, 1/2$

Table 5: Repeated game for Example 6: "$\star$" indicates the static best-response.

The reader can verify that the interiority assumption holds, i.e., $\mathscr{U} \cap \mathbb{R}^n_{++}$ is a convex hull of $\{w | \exists i \in N : w_i \in \{1/2, 2\}, w_j = 0\ \forall j \neq i\}$. Note that $\overline{m}(\infty) = \frac{1}{2} < \frac{2}{3}$; and, since $\overline{m}(\infty) = 2 > \frac{1}{2}$, no SPE in strongly symmetric strategies exists irrespective of δ. On the other hand, the best feasible symmetric payoff vector equals to $(\frac{2}{3}, \frac{2}{3}, \frac{2}{3})$, and this can be attained as a SPE by randomizing in a history-independent way between three static Nash equilibria with equal probabilities. Each player i's minmax payoff can also be attained as a SPE by playing the respective static Nash equilibrium.