DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES

2022-08

McMASTER UNIVERSITY

Department of Economics
Kenneth Taylor Hall 426
1280 Main Street West
Hamilton, Ontario, Canada
L8S 4M4

http://www.mcmaster.ca/economics/

Worst-case Regret in Ambiguous Dynamic Games

Rumen Kostadinov*
Department of Economics, McMaster University

November 2022

Abstract

I study a general model of repeated interactions between long-run players who have no probabilistic beliefs about the environment in which future interactions will take place. I introduce a notion of equilibrium, where at each history players minimise their regret from forgoing an alternative strategy under the worst-case sequence of future games, taking as given the strategies of other players. I derive a recursive characterisation of equilibrium outcomes for fixed discounting, as well as a folk theorem. I demonstrate the tractability of the characterisation in applications to risk-sharing and partnership games.

1 Introduction

Many economic interactions take place in environments that evolve over time. The leading framework for studying dynamic incentives in these settings is the *stochastic game* introduced by Shapley (1953). In a stochastic game agents play a sequence of (stage) games, the evolution of which is governed by a commonly known probability distribution that depends on past games and actions. The goal of this paper is to explore incentives in the absence of such probabilistic considerations, perhaps because the players are unsure about their probabilistic assessments, or because they want to act in a robust manner.

I study a general model of repeated interactions with ambiguity about the environment where they take place. In each period the players know the game they are playing. They simultaneously take actions and then observe the actions of others before moving to the next period. The set of games that might be played in the future is common knowledge, but their relative likelihood is unknown until the beginning of the next period when the game for that period is revealed. Hence, players cannot maximise expected utility as in standard stochastic games. Instead, I assume they minimise worst-case regret – an objective originating in Wald (1950) and Savage (1951) that, despite its popularity in decision theory, has seen little use

*I thank seminar participants at University of Montreal. I gratefully acknowledge financial support from the Social Sciences and Humanities Research Council (SSHRC).

in dynamic problems with multiple agents. In particular, I introduce a notion of Regret Perfect Equilibrium (RPE), where at each history players minimise their payoff loss (regret) from forgoing an alternative strategy under the worst-case sequence of future games, taking as given the strategies of other players. I restrict players to one-shot deviations in order to avoid issues with the time inconsistency of their preferences. This amounts to assuming that players recognise their own time inconsistency and understand that their future selves are not committed to strategies that are optimal from today's perspective.

My work offers a tractable recursive characterisation of RPE outcomes. Traditionally, outcomes in stochastic games are expressed in payoff space. However, payoffs here are contingent on the ambiguous sequence of games, making them harder to interpret. My characterisation focuses on actions instead.

I start by obtaining an upper bound on the set of action profiles that can be played in each game for any fixed discount factor. In a standard stochastic game actions can be supported only if the immediate gain from deviation is smaller than the highest possible gap between continuation equilibrium payoffs (adjusted for discounting). That is, an upper bound can be obtained by assuming the player is rewarded for the equilibrium action with the highest continuation payoff and punished for a deviation with the lowest continuation payoff. Proposition 1 is a similar result for the setting of this paper – it states that an action can be supported if the deviation can be deterred by the largest gap in equilibrium continuation payoffs for *some* environment of future games. While it is an intuitive extension of the result for stochastic games to the worst-case objective, the argument is quite different. In particular, it hinges on the possibility that a deviating player in a worst-case environment regrets another deviation, and not the equilibrium strategy.

Theorem 1 uses the insight from Proposition 1 to provide an upper bound on actions using ideas from Abreu, Pearce, and Stacchetti (1990). The bound is computed through a recursive algorithm that starts from the largest possible payoff gap and obtains progressively smaller gaps from actions that can be supported by previous gaps. The computation is fairly straightforward; even a closed form can be obtained in special cases. This contrasts stochastic games, where computation can be much harder (Yeltekin, Cai, and Judd, 2017; Abreu, Brooks, and Sannikov, 2020).

The main result of the paper, Theorem 2, states that the algorithmic bound can be attained in a RPE assuming the existence of games and actions that can respectively approximate the payoffs of any game and action with arbitrary precision. Two caveats apply. First, the result requires an arbitrarily small increase in the discount factor of each player. Then for any sequence of games there exist RPE where any actions within the algorithmic bound are played along the sequence. It is also possible to implement actions across different sequences of games with some limitations – specific actions must be played in a measure-zero set of games which form worst-case environments necessary for incentive provision.

Theorem 1 and Theorem 2 together provide a tractable algorithmic characterisation of action profiles attainable in a RPE. This characterisation hinges on the richness of actions: The proof of Theorem 2 constructs equilibria where players

who deviate in a manner that maximises their payoff in the current period regret another deviation that secures almost the same current payoff (and yields a higher continuation payoff). Still, a similar method can be used to obtain results on general action spaces. In the case of two actions, an exact characterisation of RPE actions is possible, because a deviator must regret the equilibrium strategy and vice versa. These actions form a lower bound on what is attainable with arbitrary action spaces, while the upper bound from Theorem 1 is also valid in general.

The characterisation for fixed discounting is used to derive a folk theorem in Proposition 4. As players become arbitrarily patient any action profiles can be played, unlike standard folk theorems where individual rationality must be met.

This paper is inspired by the pioneering work of Carroll (2020) who introduced the model of this paper and proposed a different solution concept – Ex-post Equilibrium (XPE), where the strategies must form a Subgame Perfect Equilibrium when restricted to any sequence of stage games. Carroll (2020) obtains a recursive characterisation of XPE actions based on payoff gaps between the best and worst equilibrium actions for each player. The analysis relies on a single long-run player and public randomisation; neither is needed in my characterisation of RPE. Krasikov and Lamba (2022) extend Carroll's approach to characterise a subset of equilibrium outcomes with multiple long-run players.

My characterisation bears some similarities to the payoff-gap recursions of Carroll (2020) and Krasikov and Lamba (2022). This allows for a clear comparison between XPE and RPE. XPE is more conservative because dynamic incentives need to hold regardless of the environment of future games. On the other hand, RPE is more permissive because incentives need to be provided only in a worst-case environment of future games. Any XPE is a RPE, but the set of RPE may be significantly larger. I discuss applications where XPE reduces to playing a Nash Equilibrium in every stage game regardless of the discount factor, which is in stark contrast with the RPE folk theorem mentioned above.

Game-theoretical solution concepts based on worst-case regret minimisation have been proposed by Renou and Schlag (2010) and Halpern and Pass (2012). Both papers focus on static games that are commonly known and assume players minimise regret under some worst-case strategy profiles by other players.[1] In contrast, I consider a dynamic environment of unknown games and a worst-case scenario based on the sequence of future games.

Regret minimisation has also been studied as an objective in repeated play of a fixed, but unknown game. This literature, surveyed in Cesa-Bianchi and Lugosi (2006), obtains asymptotic bounds on players' regret and studies whether regret-minimising strategies converges to an equilibrium of the stage game.

The literature on robust mechanism design has also used worst-case regret minimisation in static settings (Hurwicz and Shapiro, 1978; Bergemann and Schlag, 2008; Guo and Schmaya, 2019, 2022). Dynamics have been notably absent (Carroll, 2019) but interest is growing: Libgober and Mu (2021) study the problem

[1]Halpern and Pass (2012) also consider a repeated prisoner's dilemma but their solution concept is based entirely on the normal form of the game, so there are no considerations for dynamic optimality.

of a monopolist seller who maximises his payoff (not regret) against a worst-case stochastic process by which the buyer learns his value. Libgober and Mu (2022) study the same problem in a setting where the seller cannot commit to a mechanism.

The rest of the paper proceeds as follows. Section 2 formalises the model and defines RPE. Section 3 studies the special case without ambiguity where only a single stage game is possible. Section 4 contains the algorithmic upper bound on equilibrium actions, and Section 5 shows how they can be attained in a RPE. Section 6 contains results on general action spaces, the patient limit, as well as discussion on the relationship between XPE and RPE. Section 7 concludes.

2 Model

There are n players indexed by $i = 1, .., n$ playing a *supergame* consisting of an infinite sequence of (stage) games. Each game belongs to a set Θ which is common knowledge. An *environment* e is a sequence $(e_0, e_1...) \in \Theta^\infty$ of games. At the onset of the supergame the players know nothing about the environment.

The set of actions A_i available to each player i is the same across all games. Let $A := A_1 \times .. \times A_n$ denote the set of all action profiles and $A_{-i} = \times_{j \neq i} A_j$ denote the set of action profiles of players other than i. Player i's payoff from action profile $a \in A$ in game $\theta \in \Theta$ is denoted $u_i(a, \theta)$. The maximum payoff gain for player i if he deviates from this profile is

$$d_i(a, \theta) = \max_{a_i' \in A_i} u_i(a_i', a_{-i}, \theta) - u_i(a, \theta).$$

An action $a_i \in A_i$ is an ε-*best reply* to $a_{-i} \in A_{-i}$ in game θ if $d_i(a_i, a_{-i}, \theta) \leq \varepsilon$. If $\varepsilon = 0$, the action is referred to simply as a *best reply*.

Assumption (A1). Θ and A_i are compact metric spaces and u_i is continuous for each i.

(A1) is a technical assumption maintained throughout the paper. It implies a uniform bound M on the absolute value of payoffs.

2.1 Timing and histories

At the beginning of each period $t = 0, 1, 2..$ players observe the game $\theta_t \in \Theta$ that will be played at time t, but receive no information about the games they will play from $t + 1$ onwards. Then players simultaneously choose actions. The resulting action profile is revelead to each player, concluding the period. The set of time-t histories is

$$H^t := (\Theta \times A)^t \times \Theta$$

with representative element

$$h^t = (\theta_0, a^0, \theta_1, a^1, .., \theta_{t-1}, a^{t-1}, \theta_t).$$

Notice that the space of initial histories is $H^0 = \Theta$.

Example 1 (Risk-sharing (Kocherlakota, 1996))**.** There are two players indexed by $i = 1, 2$. Each period players receive endowments $\theta_1, \theta_2 \in [0, \bar{\theta}]$ which are publicly observed. Then they simultaneously choose a proportion a_i of their endowment to give to the other player. Each player evaluates his net income with a concave, strictly increasing utility function $v : \mathbb{R}_+ \to \mathbb{R}$ with $v(0) = 0$.

In the language of this paper Θ is the set $[0, \bar{\theta}]^2$ of all endowment realisations. Action spaces are $A_1 = A_2 = [0, 1]$ and

$$u_i(a_i, a_j, \theta) = v((1 - a_i)\theta_i + a_j\theta_j),$$

where $j \neq i$. ■

Example 2 (Partnership (McAdams, 2011))**.** Two players $i = 1, 2$ form a partnership. Each period productivity $\theta_i \in [\underline{\theta}, \bar{\theta}]$ for each player i is drawn and publicly observed, where $\underline{\theta} > 0$. Then they simultaneously choose whether to work (w), or shirk (s). Player i produces $2\theta_i$ if he works, and 0 otherwise. Each player receives half of the total output of the partnership less his cost of effort. Shirking has no effort cost, while the cost of working c satisfies $2\underline{\theta} > c > \bar{\theta}$.

In the language of the paper $\Theta = [\underline{\theta}, \bar{\theta}]^2$ and $A_1 = A_2 = \{w, s\}$. The payoffs are given by the following matrix where player 1 chooses the row and player 2 chooses the column.

	w	s
w	$\theta_1 + \theta_2 - c, \theta_1 + \theta_2 - c$	$\theta_1 - c, \theta_1$
s	$\theta_2, \theta_2 - c$	$0, 0$

Table 1: Payoffs in the partnership game

The parametric assumptions imply that shirking is a dominant strategy and (w, w) Pareto dominates (s, s) in every stage game. Thus, the supergame can also be interpreted as a prisoner's dilemma with varying stakes.

■

2.2 Strategies and payoffs

A (pure) strategy for player i is a map $\sigma_i : \cup_{t=0}^{\infty} H^t \to A_i$. Let Σ_i denote the set of strategies for player i. Given a strategy profile $\sigma = (\sigma_1, .., \sigma_n)$, let σ_{-i} denote the strategies of players other than i.

Suppose a strategy profile σ is played. Then the payoff of player i from history h^t depends on the continuation environment $e \in \Theta^{\infty}$ of games to be played from $t + 1$ onwards as follows:

$$U_i(\sigma | h^t, e) = (1 - \delta) \left[u_i \left(\sigma(h^t), \theta_t \right) + \sum_{s=0}^{\infty} \delta^{s+1} u_i \left(\sigma(h^{t+1+s}), e_s \right) \right],$$

where $h^{t+1+s} = (h^{t+s}, \sigma(h^{t+s}), e_s)$ for all $s \geq 0$. The discount factor $\delta \in (0, 1)$ is common to all players.

It will be useful to shorten the notation $U_i(\sigma|h^t, e)$ by appending θ_t, the last element of h^t, to the continuation environment e. For instance, player i's payoff from an initial history $h^0 = \theta$ under continuation environment e is denoted $U_i(\sigma|e')$, where $e' = (\theta, e_0, e_1, ..)$. A related situation arises when a time-t history h^t and a continuation environment e from $t+1$ have been fixed, but the interest is in player i's continuation payoff from $t+1$ after some action profile a_t is played at time t . This payoff is written as $U_i(\sigma|h^t, a_t, e)$.

2.3 Equilibrium

Unlike in standard stochastic games, the objective of the players is not to maximise payoff, but to minimise worst-case regret. Fix a history h^t and a continuation environment $e \in \Theta^\infty$ of games to be played from period $t+1$ onwards. Consider a strategy profile σ. The regret of player i from strategy σ_i is

$$R_i(\sigma|h^t, e) = \sup_{\sigma_i^d \in \Sigma_i^d(\sigma_i, h^t)} U_i(\sigma_i^d, \sigma_{-i}|h^t, e) - U_i(\sigma|h^t, e).$$

This regret is the difference between the best payoff i could obtain from an alternative strategy σ_i^d and the payoff from σ_i given that other players are following σ_{-i}. There is a subset $\Sigma_i^d(\sigma_i, h^t)$ of strategies i is allowed to deviate to. I take these strategies to be one-shot deviations at h^t, returning to σ_i thereafter:

$$\Sigma_i^d(\sigma_i, h^t) = \{\sigma_i^d \in \Sigma_i | \sigma_i(h) \neq \sigma_i^d(h) \Leftrightarrow h = h^t\}$$

When there is no ambiguity, I will identify a one-shot deviation σ_i^d with the deviating action $\sigma_i^d(h^t)$.

Regret is defined in a particular environment e of future games unknown to the players. The equilibrium concept defined below captures the idea that each player minimises regret given the strategies of others, assuming that an adversarial Nature picks a continuation environment that maximises the regret of his chosen strategy.

Definition 1. *A strategy profile σ is a Regret Perfect Equilibrium (RPE) if*

$$\sup_e R_i(\sigma|h^t, e) \leq \inf_{\sigma_i' \in \Sigma_i^d(\sigma_i, h^t)} \sup_e R_i(\sigma_i', \sigma_{-i}|h^t, e) \tag{1}$$

for any player i and history h^t.

The LHS and RHS of (1) are respectively called the equilibrium regret and (lowest) deviation regret of i at h^t.

As noted above, the set of deviant strategies has been restricted to one-shot deviations. This affects both the strategies players contemplate choosing and the strategies they consider counterfactually to compute their regret. In stochastic games the restriction to one-shot deviations is without loss of generality. Here, however, the restriction is with loss because the regret-minimising preferences are

time inconsistent. This is a general hurdle in dynamic models with robust objectives (Carroll, 2019).[2] While these time-inconsistency issues are interesting and present an important challenge, I abstract from them here to maintain tractability and to stay as close as possible to the standard framework of stochastic games. My approach can also be interpreted to assume that players expect to reoptimise their strategies at each future history to minimise worst-case continuation regret.

The restriction to one-shot deviations makes it possible to simplify the regret expressions as follows:

$$R_i(\sigma|h^t,e) = \sup_{a_i\in A_i}(1-\delta)\Big[u_i(a_i,\sigma_{-i}(h^t),\theta_t) - u_i(\sigma(h^t),\theta_t)\Big] \\ +\delta\Big[U_i(\sigma|h^t,a_i,\sigma_{-i}(h^t),e) - U_i(\sigma|h^t,\sigma(h^t),e)\Big]. \tag{2}$$

3 Known environment

Suppose that Θ consists of a single game θ so that there is no ambiguity about the environment $e=(\theta,\theta,..)$ of games that will be played. Then in any RPE σ

$$\sigma_i \in \operatorname*{argmin}_{\sigma_i'\in\Sigma_i^d(\sigma_i,h^t)} R_i(\sigma_i',\sigma_{-i}|h^t,e)$$

for any history h^t and player i. It follows that $R_i(\sigma|h^t,e)=0$, since it is possible to make $R_i(\sigma_i',\sigma_{-i}|h^t,e)$ arbitrarily small by considering strategies $\sigma_i'\in\Sigma_i^d(\sigma_i,h^t)$ that approximately attain the supremum of $U_i(\sigma_i',\sigma_{-i}|h^t,e)$. Hence,

$$\sigma_i \in \operatorname*{argmax}_{\sigma_i'\in\Sigma_i^d(\sigma_i,h^t)} U_i(\sigma_i',\sigma_{-i}|h^t,e),$$

which means that any one-shot deviation from σ_i decreases i's payoff. This is the exact requirement for σ to be a Subgame Perfect Equilibrium (SPE) of an infinitely repeated game θ.

Conversely, if σ is a SPE of an infinite repetition of θ, then $R_i(\sigma|h^t,e)=0$ for every player i and history h^t. It follows that σ is also a RPE. Thus, RPE coincides with SPE in the absence of ambiguity about stage games. This equivalence also holds when the environment e is known but nonstationary.

Since $R_i(\sigma|h^t,e)=0$, it follows from (2) that

$$(1-\delta)d_i(\sigma(h^t),\theta_t) \le \delta\left(\bar{U}_i(e)-\underline{U}_i(e)\right) \tag{3}$$

where $\bar{U}_i(e)$ and $\underline{U}_i(e)$ are respectively the supremum and infimum payoff for player i over all RPE payoffs in environment e. This is a necessary condition familiar from the theory of stochastic games: The maximum payoff i can gain by deviating from his equilibrium strategy does not exceed an *incentive gap* – the difference between i's best and worst continuation equilibrium payoffs (the former used to reward his equilibrium action and the latter used to punish his deviation).

[2]Libgober and Mu (2021, 2022) also face this issue of time inconsistency. They resolve it by assuming the worst-case behaviour of Nature is time-consistent. In my setting this would mean that at every history Nature reselects the worst-case environment of future games to maximise continuation regret.

4 Bounds

In this section I obtain necessary conditions for action profiles played in a RPE.

4.1 Initial estimate

For any $w = (w_1, ..., w_n)$, let

$$A^*(\theta|w) = \{a \in A | (1-\delta) d_i(a, \theta) \leq \delta w_i \ \forall i\}.$$

Action profiles in $A^*(\theta|w)$ are said to be supported in game θ by incentive gap w.

In the setting of Section 3, where the environment is known, a necessary condition for RPE is that the actions at any history are supported by gap $(\bar{U}_1(e) - \underline{U}_1(e), .., \bar{U}_n(e) - \underline{U}_n(e))$, where e is the only possible environment. Proposition 1 below extends this bound to the general case. It states that equilibrium actions must be supported by a gap $w^* = (w_1^*, .., w_n^*)$ consisting of each player's highest difference between best and worst equilibrium payoffs across all environments, that is

$$w_i^* = \sup_e \left(\bar{U}_i(e) - \underline{U}_i(e)\right).$$

Proposition 1. *Let σ be a RPE. Then $\sigma(h^t) \in A^*(\theta_t|w^*)$ at any history h^t.*

The proof of Proposition 1 and other omitted proofs are in the Appendix. Even though Proposition 1 generalises condition (3) from stochastic games, it cannot be obtained using the usual arguments, since RPE and SPE do not coincide in general. The argument, instead, obtains bounds on the equilibrium regret and deviation regret of player i at any history h^t. Suppose that if i does not deviate, he receives the highest continuation payoff $\bar{U}_i(e)$ in any continuation environment e. This can only lower equilibrium regret and increase deviation regret, so the equilibrium condition (1) continues to hold. Notice that i's regret from deviating to an action that maximises his time-t payoff can be no larger than δw_i^* – the largest difference between RPE continuation payoffs. On the other hand his equilibrium regret can be no higher than $(1-\delta) d_i(\sigma(h^t), \theta_t)$, since continuation payoffs on the equilibrium path dominate payoffs following any deviation. If equilibrium regret equals the above bound, then it follows from (1) that

$$(1-\delta) d_i(\sigma(h^t), \theta_t) \leq \delta w_i^*, \tag{4}$$

as required. If the equilibrium regret is lower than the bound by some amount x, the continuation payoffs following a deviation in any environment e cannot be as high as $\bar{U}_i(e)$. This makes it possible to refine the bound on the deviation regret, showing that it decreases by x as well, preserving the desired inequality (4).

4.2 Recursive representation and algorithm

It is now possible to make progress towards a fixed point representation of the optimal incentive gaps w^*. On the one hand, Proposition 1 obtains a superset of

RPE action profiles as a function of w^*. On the other hand, these action profiles imply bounds on RPE payoffs that can be used to estimate w^*. This connection is captured by the following operator $B : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ defined on arbitrary incentive gaps w:

$$Bw = \left(\max_{\theta} \max_{a \in A^*(\theta|w)} u_i(a, \theta) - \min_{a \in A^*(\theta|w)} u_i(a, \theta) \right)_{i=1,..,n}.$$

Let $\bar{a}^i(\theta|w)$ and $\underline{a}^i(\theta|w)$ denote the action profiles in $A^*(\theta|w)$ that respectively maximise and minimise player i's payoff. A game θ is *incentive-optimal* for i given gap w if it maximises the stage-game payoff difference $u_i(\bar{a}^i(\theta|w)) - u_i(\underline{a}^i(\theta|w))$. The operator B computes this payoff difference for each player to obtain a new gap Bw. Note that the new gaps for each player may be computed in different games, since their incentive-optimal games need not coincide.

The operator B is inspired by the recursive operator of Abreu, Pearce, and Stacchetti (1990) which can provide an exact characterisation of equilibrium payoffs in stochastic games (Mailath and Samuelson, 2006). In what follows I use similar arguments to obtain an upper bound on equilibrium actions instead. There is a much closer connection to the recursive operators of Carroll (2020) and Krasikov and Lamba (2022) used to characterise Ex-Post Equilibria; this is discussed in Section 6.3.

Definition 2. *A gap w is self-generating if $Bw \geq w$.*

Lemma 1. *w^* is self-generating.*

Many gaps can be self-generating. The following result obtains the largest self-generating gap $\bar{w}$ using an algorithmic procedure similar to APS.

Lemma 2. *Let $w^0 = (2M, .., 2M)$. Define $w^{k+1} = Bw^k$ inductively for each $k = 0, 1, ...$ Then*

1. *(w^k) is decreasing and converges to some limit $\bar{w}$.*
2. *$B\bar{w} = \bar{w}$.*
3. *$\bar{w} \geq w$ for any self-generating gap $w \leq w^0$.*

Proof. 1. Since payoffs are bounded by M, it follows that $w^1 \leq 2M = w^0$. That the rest of the sequence is monotonically decreasing follows from the monotonicity of B. Hence, (w^k) converges to some $\bar{w} = (\lim_{k\to\infty} w^k_1, .., \lim_{k\to\infty} w^k_n)$.

2. By monotonicity of B it follows that $B\bar{w} \leq Bw^k = w^{k+1}$ for all k. Taking the limit as $k \to \infty$ obtains $B\bar{w} \leq \bar{w}$. It remains to show that $\bar{w}$ is self-generating. To this end, fix a player i and consider sequences $(\theta^k), (\bar{a}^k), (\underline{a}^k)$ such that

$$(Bw^k)_i = u_i(\bar{a}^k, \theta^k) - u_i(\underline{a}^k, \theta^k) \text{ and } \bar{a}^k, \underline{a}^k \in A^*(\theta^k|w^k)$$

for all k. By Assumption (A1) they converge to some respective limits $\theta^\infty, \bar{a}^\infty, \underline{a}^\infty$ along some subsequence. It follows from the continuity of u_i that $\bar{a}^\infty, \underline{a}^\infty \in A^*(\theta^\infty|\bar{w})$. Hence $(B\bar{w})_i \geq \lim_k (Bw^k)_i$. Since i is arbitrary, the result obtains.

3. Let $w \leq w^0$ be a self-generating gap with $w_i > \bar{w}_i$ for some i. By monotonicity of (w^k) there exists $K = \max\{k | w^k \geq w\}$. Hence, there exists j such that $w_j^{K+1} < w_j$. It follows from the monotonicity of B that

$$(Bw)_j \leq (Bw^K)_j = w_j^{K+1} < w_j.$$

Hence, w is not self-generating. A contradiction. □

It is now possible to summarise the results of this section in the following upper bound on the set of equilibrium actions:

Theorem 1. *Let σ be a RPE. Then $\sigma(h^t) \in A^*(\theta_t|\bar{w})$ at any history h^t.*

Proof. Since payoffs are bounded by M, it follows that $w^* \leq w^0$. Lemma 1 and Lemma 2 then imply that $w^* \leq \bar{w}$. The result now follows from Proposition 1, since $A^*(\theta|w)$ is increasing in w. □

4.3 Computation in the risk-sharing game

This section outlines the computation of the gap $\bar{w}$ from Lemma 2 in the risk-sharing supergame from Example 1.

Consider i's best action profile in game θ under gap w. It can be shown that in this profile i gives nothing to $j \neq i$, and j gives i the largest income share supported by gap w:

$$\begin{aligned}&\bar{a}^i(\theta|w) = (0, a_j^i)\\ &\text{where } a_j^i = \max\left\{a_j \in [0,1] \middle| (1-\delta)\Big(v(\theta_j) - v((1-a_j)\theta_j)\Big) \leq \delta w_j\right\}\end{aligned}$$

Similarly, in i's worst action profile he contributes as much as possible, while j gives him nothing:

$$\begin{aligned}&\underline{a}^i(\theta|w) = (a_i^i, 0)\\ &\text{where } a_i^i = \max\left\{a_i \in [0,1] \middle| (1-\delta)\Big(v(\theta_i) - v((1-a_i)\theta_i)\Big) \leq \delta w_i\right\}\end{aligned}$$

Now I find an incentive-optimal game (θ_i, θ_j) for i under gap w. Since j's income affects i's best utility but not his worst, it is without loss of generality to set $\theta_j = \bar{\theta}$ to maximise j's contribution. The income of player i must be the highest amount that he is willing to give to j, leaving i with zero utility in his worst outcome. A lower income for i would decrease his best utility, while leaving his worst utility unchanged. A higher income would also decrease the gap between best and worst utility due to decreasing marginal utility. In summary, there is an incentive-optimal game for i with endowments

$$\theta_i = \min\left\{v^{-1}\left(\frac{\delta}{1-\delta}w_i\right), \bar{\theta}\right\} \text{ and } \theta_j = \bar{\theta}.$$

The recursive operator is then given by

$$(Bw)_i = v\left(\min\left\{v^{-1}\left(\frac{\delta}{1-\delta}w_1\right), \bar{\theta}\right\} + \bar{\theta} - v^{-1}\left(\max\left\{v(\bar{\theta}) - \frac{\delta}{1-\delta}w_2, 0\right\}\right)\right)$$

Since the supergame is symmetric, the algorithm in Lemma 2 boils down to a one-dimensional recursion that is easy to compute. It is also possible to obtain the following closed form for $\bar{w}$ when $v(x) = \sqrt{x}$:

$$\bar{w}_1 = \bar{w}_2 = \min\left\{2v(\bar{\theta})\frac{\delta}{1-\delta}, v(2\bar{\theta}) - v(0)\right\}.$$

When players are patient, $\bar{w}_i$ equals the maximum possible gap, i.e. the incentive-optimal game is $(\bar{\theta}, \bar{\theta})$, i receives all of j's income in his best outcome, and gives all of his income away in his worst outcome. When players are sufficiently impatient, the upper bound on the equilibrium gap is increasing in the maximum income and the discount factor.

5 Attaining the bounds

In this section I demonstrate the attainability of action profiles from Theorem 1 in a RPE. The results use the following assumptions, maintained for the rest of the section.

Assumption (A2). Θ has no isolated points.[3]

Assumption (A3). $A_1, .., A_n$ have no isolated points.

In conjunction with (A1), Assumption (A2) implies that the payoffs from any game can be approximated with arbitary precision in another game. This seems natural in the presence of ambiguity – if players do not know the probability of playing a certain stage game, they likely entertain the possibility of similar games. Since (locally) compact metric spaces with no isolated points are uncountable, (A1) and (A2) also imply that Θ is uncountable.[4] Assumption (A2) holds in Examples 1 and 2 where the set of stage games is a product of intervals.

Assumption (A3) implies the same richness in action spaces; in particular action spaces must be uncountable. This allows for many standard games where the available actions are an interval, such as the risk-sharing game from Example 1, but rules out games with finitely many actions such as the partnership game in Example 2. Games with general action spaces are addressed in Section 6.1.

[3] x is an isolated point of a metric space if any open ball centered on x contains at least one other point.

[4] See Theorem 27.7 in Munkres (2013).

5.1 Outcomes

At the onset, it is unclear what it means to attain actions in a RPE due to the ambiguity about the games that will be played. Following Carroll (2020), I adopt two ways of defining outcomes in the supergame.

First, an analyst may be concerned with actions played in a particular environment, perhaps known to the analyst but not to the players.

Definition 3. *A realisable outcome is a pair (e, a), where e is an environment and $a = (a_t)_{t=0}^{\infty}$ is a sequence of action profiles in A.*

A full outcome, instead, describes actions taken in all possible environments.

Definition 4. *A full outcome is a collection $(a_t)_{t=0}^{\infty}$ of functions $a_t : \Theta^{t+1} \to A$.*

5.2 Approximate implementability

I will study the implementability of realisable and full outcomes on the path of RPE in the following sense.

Definition 5. *A history h^t is on the path of strategy profile σ if $a_s = \sigma(\theta_0, a_0, .., \theta_s)$ for all $s < t$.*

Let $\Gamma(\delta)$ denote the supergame with discount factor δ.

Definition 6. *A realisable outcome (e, a) is (approximately) implementable if for all $\delta' > \delta$ there exists a RPE σ of $\Gamma(\delta')$ such that $\sigma(h^t) = a_t$ for any history h^t on the path of σ such that $(\theta_0, .., \theta_t) = (e_0, .., e_t)$.*

Definition 7. *A full outcome a is (approximately) implementable if for all $\delta' > \delta$, there exists a RPE σ of $\Gamma(\delta')$ and a countable set $X \subseteq \Theta$ such that $\sigma(h^t) = a_t(\theta_0, \theta_1, .., \theta_t)$ for any history h^t on the path of σ.*

Approximate implentability requires that the outcome matches behaviour on the path of equilibria of supergames with higher (but arbitrarily close) discount factors. There is an exception – equilibria implementing a full outcome need not match it in a countable number of games. Since Θ is uncountable, this means that only a zero-measure set of games is excluded from the outcome.[5]

The main result of the paper is that realisable and full outcomes are approximately implementable if and only if the actions at each history are within the bounds from Theorem 1.

Theorem 2.

1. *A realisable outcome is approximately implementable iff $a_t(\theta_0, .., \theta_t) \in A^*(\theta_t|\bar{w})$ for all $(\theta_0, .., \theta_t)$*

[5]Definition 7 can be strengthened so that for any countable set $\Theta^0 \subseteq \Theta$ it requires the existence of equilibria where $X \subseteq \Theta \setminus \Theta^0$. This would not change the results.

2. *A full outcome is approximately implementable iff* $a_t(\theta_0, .., \theta_t) \in A^*(\theta_t|\bar{w})$ *for all* $(\theta_0, .., \theta_t)$.

The proof of the "only if" direction of both parts of Theorem 2 is a direct consequence of the bounds in Theorem 1; it is relegated to the Appendix. The "if" direction is obtained from the equilibrium constructions in the following sections.

5.3 Implementation of full outcomes

Consider any discount factor δ and full outcome a with $a_t(h^t) \in A^*(\theta_t|\bar{w})$ for any h^t. Recall that incentive-optimal games admit the largest range of payoffs for the respective player. Hence, a worst-case environment of incentive-optimal games has the potential to create maximal regret, making it a natural vehicle for the implementation of outcome a. However, there is in general only one incentive-optimal game for each player, which turns out to be too restrictive for the provision of dynamic incentives. Therefore, the equilibrium construction that follows relies on environments consisting of games that are approximately incentive-optimal in the following sense.

Definition 8. *A game* θ^i *is* ε-optimal *for* i *if*

$$u_i(\bar{a}^i(\theta^i|\bar{w}), \theta^i) - u_i(\underline{a}^i(\theta^i|\bar{w}), \theta^i) > \max_{\theta} \Big(u_i(\bar{a}^i(\theta|\bar{w}), \theta) - u_i(\underline{a}^i(\theta|\bar{w}), \theta) \Big) - \varepsilon.$$

It follows from $B\bar{w} = \bar{w}$ (Lemma 2) that

$$\max_{\theta} \Big(u_i(\bar{a}^i(\theta|\bar{w}), \theta) - u_i(\underline{a}^i(\theta|\bar{w}), \theta) \Big) - \varepsilon = \bar{w}_i - \varepsilon.$$

Thus, ε-optimal games for i can support a payoff gap arbitrarily close to $\bar{w}_i$. Assumptions (A1) and (A2) imply that the set of ε-optimal games for i is locally compact and has no isolated points; hence, it is uncountable (cf. footnote 4).

Equilibrium construction

Let $\varepsilon > 0$ and consider countably infinite and pairwise disjoint sets $\Theta^1, .., \Theta^n$ such that each game in Θ^i is ε-optimal games for i.[6] In what follows I construct a strategy profile σ. I will later show that σ forms a RPE in supergames with discount factor $\delta' > \delta$, but for now it is useful to think of σ in the context of $\Gamma(\delta)$.

The construction assigns to each history h^t a set of $\Theta^i(h^t) \subseteq \Theta^i$ of *reward games* for player i. The strategies are fully determined by the reward games as follows:

$$\sigma(h^t) = \begin{cases} \bar{a}^i(\theta_t|\bar{w}) & \text{if } \theta_t \in \Theta^i(h^t) \\ a_t(\theta_0, .., \theta_t) & \text{if } \theta_t \notin \cup_i \Theta^i(h^t) \text{ and } h^t \text{ is on the path of } \sigma \\ \underline{a}^j(\theta_t|\bar{w}) & \text{if } \theta_t \notin \cup_i \Theta^i(h^t) \text{ and } j \text{ was the last player to deviate from } \sigma \end{cases}$$

[6] Such a selection exists because there are uncountably many ε-optimal games for each player.

As long as there are no deviations, the strategies implement the outcome a in each game that is not a reward game for any player. This excludes a countable set $\Theta^1 \cup .. \cup \Theta^n$ of games from the outcome. Off the equilibrium path, the worst profile for the deviator supported by $\bar{w}$ is played. In case h^t exhibits multiple deviations, the player who deviated in the most recent round is punished.[7] Finally, in a reward game for player i the strategies specify the best action profile for i *regardless of past deviations*.

The set of reward games evolves as follows. At each initial history $h^0 \in H^0$ let $\Theta^i(h^0) = \Theta^i$. At each history h^t there are two reward games $\theta_1^i, \theta_2^i \in \Theta^i(h^t)$ and an action $a_i' \in A_i$ that is an ε-best reply to $\sigma_{-i}(h^t)$ for each i. The set of reward games at $t+1$ is unchanged from $\Theta^i(h^t)$ if i does not deviate at h^t. If i deviates, one of θ_1^i, θ_2^i is excluded from future reward games – the former is excluded if, and only if i deviates to a_i'. Formally, the set of reward games for i at $h^{t+1} = (h^t, a, \theta_{t+1})$ is

$$\Theta^i(h^{t+1}) = \begin{cases} \Theta^i(h^t) & \text{if } a_i = \sigma_i(h^t) \\ \Theta^i(h^t) \setminus \{\theta_1^i\} & \text{if } a_i = a_i' \\ \Theta^i(h^t) \setminus \{\theta_2^i\} & \text{if } a_i \neq \sigma_i(h^t), a_i' \end{cases} \tag{5}$$

for any $a \in A$, $\theta_{t+1} \in \Theta$.

Regret bounds

The strategies σ constructed above make use of each player's best and worst action profiles supported by $\bar{w}$ with discount factor δ.[8] However, approximately implementability only requires that these strategies form a RPE when the discount factor is $\delta' > \delta$. Here, I consider any such δ' and obtain bounds on the equilibrium and deviation regret of σ in $\Gamma(\delta')$.

Fix some history h^t and a player i. Consider i's equilibrium regret at h^t. Note that any deviation at h^t gives i a lower continuation payoff from $t+1$ than the payoff he would obtain from following σ *regardless of the continuation environment*. This is because i's deviation shrinks his set of reward games and results in his worst payoff in any non-reward game (assuming other players do not deviate). It follows that i's equilibrium regret can be no larger than his maximum payoff gain at t:

$$\sup_e R_i(\sigma | h^t, e) \leq (1 - \delta') d_i(\sigma(h^t), \theta_t). \tag{6}$$

Suppose i plays a one-shot deviation to a_i' and hence, θ_1^i is excluded from future rewards. When the continuation environment is $e_1^i = (\theta_1^i, \theta_1^i, ..)$ this results in actions $\underline{a}^i(\theta_1^i)$ in each subsequent period. But any other one-shot deviation as well as the equilibrium strategy will result in actions $\bar{a}^i(\theta_1^i)$, since θ_1^i will not be excluded from future rewards. Moreover, at least one of these alternative strategies

[7] RPE relies on unilateral deviations, so the specification of the strategies following multiple deviations in the same period is immaterial.

[8] Note that $\bar{w}$ also depends on δ, since it is obtained from the algorithm in Lemma 2.

plays an ε-best response to $\sigma_{-i}(h^t)$. Thus,

$$R_i(a_i', \sigma_{-i}|h^t, e_1^i) \geq (1-\delta')(-\varepsilon) + \delta'(\bar{w}_i - \varepsilon).$$

Now suppose that i plays a one-shot deviation to $a_i^d \neq a_i'$ and hence, θ_2^i is excluded from future rewards. When the continuation environment is $e_2^i = (\theta_2^i, \theta_2^i, ..)$ this results in actions $\underline{a}^i(\theta_2^i)$ in each subsequent period. If i instead deviates to a_i', actions $\underline{a}^i(\theta_2^i)$ would be played. Thus,

$$R_i(a_i^d, \sigma_{-i}|h^t, e_2^i) \geq (1-\delta')(-\varepsilon) + \delta'(\bar{w}_i - \varepsilon).$$

Hence, deviation regret can be bounded from below as follows:

$$\inf_{\sigma_i' \in \Sigma_i^d(\sigma_i, h^t)} \sup_e R_i(\sigma_i', \sigma_{-i}|h^t, e) \geq \delta'\bar{w}_i - \varepsilon \tag{7}$$

Proof of the "if direction" of Theorem 2 for full outcomes

Let $\delta' > \delta$. For $\varepsilon > 0$ small enough the regret bounds (6) and (7) imply that

$$\sup_e R_i(\sigma|h^t, e) \leq (1-\delta)d_i(\sigma(h^t), \theta_t) \leq \delta\bar{w}_i \leq \inf_{\sigma_i' \in \Sigma_i^d(\sigma_i, h^t)} \sup_e R_i(\sigma_i', \sigma_{-i}|h^t, e) \tag{8}$$

for any history h^t and player i. The second inequality follows from $\sigma(h^t) \in A^*(\theta_t|\bar{w})$. If $\theta_t \in \Theta^i(h^t)$ this is by definition of $\bar{a}^i(\theta_t|\bar{w})$; otherwise, $\sigma(h^t) = a(\theta_0, .., \theta_t)$ which is assumed in $A^*(\theta_t|\bar{w})$.

Hence, for every $\delta' > \delta$ there exists a RPE σ implementing a in the sense of Definition 7 with $X = \Theta^1 \cup .. \cup \Theta^n$.

5.4 Implementation of realisable outcomes

The implementation of a realisable outcome (a, e) is less demanding that the implementation of full outcome because the outcome needs to be matched only in environment e. Hence, this part of Theorem 2 can be proved using the above equilibrium construction for full outcomes. One only needs to ensure that the reward games $\Theta^1, .., \Theta^n$ do not overlap with any games from environment e. Since the sets of ε-optimal games for each player are uncountable, it is straightforward to show that the reward games can be selected in such a manner.

6 Discussion

6.1 General action spaces

The equilibrium construction in Section 5.3 is quite distinct from constructions used in stochastic games. Notably, when a player deviates, what he regrets the most under the worst-case continuation environment is not that he did not follow the equilibrium strategy. Instead, his biggest regret is that he did not deviate to another action that (almost) maximises his immediate payoff and provides the

same (maximal) continuation payoff as the equilibrium strategy. Such a high regret cannot be obtained from the equilibrium strategy unless the latter prescribes a best reply. Hence, this aspect of the construction – that the deviator regrets another deviation – is necessary to attain the bounds from Theorem 1 in RPE. However, it requires that for any deviation, there exists another deviation that plays an ε-best reply. This is exactly the condition imposed by Assumption (A3) on action spaces. This section leverages the tools developed so far to give a partial characterisation of implementability in games with arbitrary action sets, i.e. when Assumption (A3) is dropped.

Lower bound on RPE actions

The characterisation uses analogues of supportable actions A^* and the maximum reward-punishment gap $\bar{w}$ for equilibria where deviators regret the equilibrium strategy the most. Let

$$\underline{A}(\theta|w) = \{a \in A | 2(1-\delta)d_i(a,\theta) \leq \delta w_i \ \forall i\}$$

and let $\underline{w}$ be the gaps obtained by the algorithmic procedure in Lemma 2 with operator $\underline{B}$ given by

$$\underline{B}w = \left(\max_{\theta} \max_{a \in \underline{A}(\theta|w)} u_i(a,\theta) - \min_{a \in \underline{A}(\theta|w)} u_i(a,\theta) \right)_{i=1,..,n}.$$

in place of the operator B.[9] Proposition 2 states that actions in $\underline{A}(\theta|\underline{w})$ are implementable in game θ. Though these actions are a subset of $A^*(\theta|\bar{w})$, they are implementable for any action spaces.

Proposition 2. *Assume (A1) and (A2). Then*

1. *Any realisable outcome (a,e) with $a_t \in \underline{A}(e_t|\underline{w})$ is approximately implementable.*

2. *Any full outcome a with $a(h^t) \in \underline{A}(\theta_t|\underline{w})$ is approximately implementable.*

Proof. The proof follows the arguments from Section 5.3 and Section 5.4. I sketch the argument for full outcomes, omitting some details. Let a be a full outcome with $a(h^t) \in \underline{A}(\theta_t|\underline{w})$ for all h^t and let $\varepsilon > 0$. There exist disjoint, countably infinite sets $\Theta^1, .., \Theta^n$ such that

$$\max_{a \in \underline{A}(\theta^i|\underline{w})} u_i(a,\theta^i) - \min_{a \in \underline{A}(\theta^i|\underline{w})} u_i(a,\theta^i) > \max_{\theta} \left(\max_{a \in \underline{A}(\theta|\underline{w})} u_i(a,\theta) - \min_{a \in \underline{A}(\theta|\underline{w})} u_i(a,\theta) \right) - \varepsilon$$

for all i and $\theta^i \in \Theta^i$. Consider an amendment of the construction in Section 5.3, where the reward games evolve according to

$$\Theta^i(h^{t+1}) = \begin{cases} \Theta^i(h^t) & \text{if } a_i = \sigma_i(h^t) \\ \Theta^i(h^t) \setminus \{\theta_1^i\} & \text{if } a_i \neq \sigma_i(h^t) \end{cases}$$

[9] See Lemma 5 in the Appendix.

instead of (5). That is, all deviations exclude the same game from future rewards. For any $\delta' > \delta$, the bound on i's equilibrium regret in (6) holds by the same argument, but the bound on i's deviation regret in (7) no longer holds. A weaker bound can be obtained by considering a continuation environment $(\theta_1^i, \theta_1^i, ..)$. There, i's biggest regret is that he did not follow the equilibrium strategy and hence

$$\inf_{\sigma_i' \in \Sigma_i^d(\sigma_i, h^t)} \sup_e R_i(\sigma_i', \sigma_{-i}|h^t, e) \geq \delta' \underline{w}_i - (1-\delta') d_i(\sigma(h^t), \theta_t).$$

Similarly to (8), it follows that

$$\begin{aligned} \sup_e R_i(\sigma|h^t, e) &\leq (1-\delta) d_i(\sigma(h^t), \theta_t) \\ &\leq \delta \underline{w}_i - (1-\delta) d_i(\sigma(h^t), \theta_t) \leq \inf_{\sigma_i' \in \Sigma_i^d(\sigma_i, h^t)} \sup_e R_i(\sigma_i', \sigma_{-i}|h^t, e) \end{aligned}$$

for all i and h^t when $\varepsilon > 0$ is sufficiently small. Note that $a(h^t) \in \underline{A}(\theta_t|\underline{w})$ is used to obtain the second inequality. □

Proposition 2 gives a general lower bound on the actions implementable in a RPE. Moreover, the upper bound from Theorem 1 continues to hold because it was derived in the absence of Assumption (A3). Thus, the set of supportable actions in game θ is between $\underline{A}(\theta|\underline{w})$ and $A^*(\theta|\bar{w})$.

Binary action spaces

The above characterisation can be sharpened in games with two actions for each player. In particular, an exact characterisation is obtained when one of the games in Θ is payoff-invariant in the following sense:[10]

Definition 9. *A game θ is trivial if $u_i(a, \theta) = u_i(a', \theta)$ for all i and $a, a' \in A$.*

Trivial games give each player the same payoff no matter the actions played. The risk-sharing example contains a trivial game where both endowments equal zero. A trivial game may also represent a period where players cannot interact for exogenous reasons. For example, a trivial game in the partnership example can be an interruption in activities due to change in business conditions, restructuring, dissolution of the partnership, etc.

Proposition 3. *Assume (A1) and (A2). If $|A_i| = 2$ for all i and there exists a trivial game in Θ, then*

1. *A realisable outcome (a, e) is approximately implementable only if $a_t \in \underline{A}(e_t|\underline{w})$.*
2. *A full outcome a is approximately implementable only if $a(h^t) \in \underline{A}(\theta_t|\underline{w})$.*

The upper bound on implementable actions is obtained because for every deviation there is only one other action that the player can regret. Thus, any deviator's

[10] This assumption was suggested to me by Sam Kapon.

biggest regret *must* be the equilibrium action. This helps obtain the following bound on deviation regret:

$$\inf_{\sigma_i' \in \Sigma_i^d(\sigma_i, h^t)} \sup_e R_i(\sigma_i', \sigma_{-i}|h^t, e) \leq \delta \underline{w} - (1-\delta) d_i(\sigma(h^t), \theta_t).$$

On the other hand, the existence of a trivial game places a lower bound on equilibrium regret. If all future games are trivial, regret equals the maximum payoff gain in the current period. Hence,

$$\sup_e R_i(\sigma|h^t, e) \geq (1-\delta) d_i(\sigma(h^t), \theta_t).$$

It follows that the actions $\sigma(h^t)$ played at any history h^t in a RPE σ are in $\underline{A}(\theta_t|\underline{w})$, which suffices to show Proposition 3.

The same idea can be used to refine the upper bound on implementable actions in other settings. In games with more than two actions the upper bound on deviation regret would be higher, since other deviations may have better immediate payoffs than the equilibrium action. In the absence of a trivial game a weaker lower bound on equilibrium regret can be obtained by considering for each player the game that minimises the largest difference in payoffs across all action profiles. This would be a trivial game if one exists, resulting in a payoff difference of zero. In general, this difference may be higher, necessitating an adjustment of the bound.

Computation in the partnership game

Proposition 3 applies to the partnership supergame in Example 2 augmented with the possibility of a trivial game (where, for instance, each action profile has zero payoff for both players).[11] Thus, approximately implementable outcomes can be characterised via gap $\underline{w}$ obtained by iterative application of $\underline{B}$ starting from $w^0 = (2M, 2M)$.

Consider any gap $w = (w_1, w_2)$ with $0 \leq w_1 = w_2 \leq 2M$. For any player i, action $a_j \in A_j$, and nontrivial game $\theta = (\theta_1, \theta_2)$

$$d_i(s, a_j, \theta) < 0 \qquad \text{and} \qquad d_i(w, a_j, \theta) = c - \theta_j.$$

Hence, shirking can always be supported by gap w (recall that it is a dominant strategy in every stage game). Let

$$\tilde{\theta} = \max\left\{\underline{\theta}, \min\{\theta | 2(1-\delta)(c-\theta) \leq \delta w_1\}\right\}.$$

[11] The introduction of a trivial game may violate Assumption (A2). However, Proposition 3 continues to hold because the argument relies on the existence of games approximating ε-incentive optimal games. Trivial games generate a best-case payoff gap of 0, so they are not ε-incentive optimal for sufficiently small ε except in the degenerate case where every game is trivial.

be the lowest productivity where working can be supported by gap w. It follows that

$$\underline{A}(\theta|w) = \begin{cases} A_1 \times A_2 & \text{if } \theta_1 \geq \tilde{\theta}, \theta_2 \geq \tilde{\theta} \\ \{(s,s)\} & \text{if } \theta_1 < \tilde{\theta}, \theta_2 < \tilde{\theta} \\ \{(s,s),(w,s)\} & \text{if } \theta_1 \geq \tilde{\theta}, \theta_2 < \tilde{\theta} \\ \{(s,s),(s,w)\} & \text{if } \theta_1 < \tilde{\theta}, \theta_2 \geq \tilde{\theta} \end{cases}$$

for every nontrivial game θ. If $\tilde{\theta} > \bar{\theta}$ only shirking is supported. Otherwise, it can be shown that $(\tilde{\theta}, \bar{\theta})$ is an incentive-optimal game for player 1. All action profiles are supported in this game by gap w. Profile (s,w) results in the highest payoff for player 1 equal to $\theta_2 = \bar{\theta}$. Profile (w,s) results in the worst payoff of $\theta_1 - c = \tilde{\theta} - c$. Thus,

$$(\underline{B}w)_1 = (\underline{B}w)_2 = \begin{cases} 0 & \text{if } \tilde{\theta} > \bar{\theta} \\ \bar{\theta} - (\tilde{\theta} - c) & \text{otherwise} \end{cases}$$

It follows that $\underline{w}_1 = \underline{w}_2$. If these gaps are positive, the threshold productivity $\tilde{\theta}$ where working can be supported by $\underline{w}$ satisfies

$$2(1-\delta)(c - \tilde{\theta}) \leq \delta \underline{w}_1 = \delta(\underline{B}\underline{w})_1 = \delta[\bar{\theta} - (\tilde{\theta} - c)].$$

Since $\underline{w}$ is the largest fixed point of $\underline{B}$, $\tilde{\theta}$ is the largest productivity for which the above inequality holds, that is

$$\tilde{\theta} = \max\{\theta | (2 - 3\delta)(c - \theta) \leq \delta\bar{\theta}\}.$$

It is now possible to solve for the implementable outcomes as follows:

- If $\frac{2(1-\delta)}{2-3\delta}\bar{\theta} < c$, then $\tilde{\theta} > \bar{\theta}$ and working cannot be induced in any stage game. Thus, only shirking is implementable.
- If $\delta \geq \frac{2}{3}$, then $\tilde{\theta} = \underline{\theta}$ so working can be induced in every stage game. Thus, any outcome is implementable.
- In the remaining cases any player can be induced to work if, and only if his productivity is no less than $c - \delta\bar{\theta}/(2 - 3\delta)$.

Greater patience and lower cost of effort expand the set of implementable outcomes, as expected. An increase in $\bar{\theta}$, which can be interpreted as greater ambiguity, also has a positive effect on incentives.

6.2 Folk Theorem

There is a long-standing tradition to explore the limits of equilibrium behaviour as the players become arbitrarily patient, i.e. $\delta \to 1$. It is possible to obtain such a result here as well.

Recall the algorithm from Lemma 2 that obtains $\bar{w}$ by repeated application of the operator B. Since B is increasing in δ, the algorithm implies that $\bar{w}$ is also increasing in δ. Thus, as $\delta \to 1$, the immediate gain from deviation vanishes, while the incentive gaps stay bounded away from zero (except in degenerate cases where a player's payoff is unaffected by his own actions and the actions of others). It follows that $A^*(\theta|\bar{w}) \to A$ for all θ as $\delta \to 1$. Under assumptions (A1), (A2) and (A3) Theorem 2 implies that any action profile in any game can be played in a RPE given sufficient patience. The result also holds in general action spaces, since $\underline{A}$ behaves similarly to A^* in the patient limit.

Proposition 4. *Assume (A1) and (A2). Then*

1. *For any realisable outcome (a, e), there exists $\bar{\delta} \in (0, 1)$ such that (a, e) is approximately implementable whenever $\delta > \bar{\delta}$.*

2. *For any full outcome a, there exists $\bar{\delta} \in (0, 1)$ such that a is approximately implementable whenever $\delta > \bar{\delta}$.*

Proposition 4 contrasts folk theorems for stochastic games (Dutta, 1995; Fudenberg and Yamamoto, 2011; Hörner, Sugaya, Takahashi, and Vieille, 2011). In stochastic games players have nontrivial payoff guarantees they can obtain by playing best replies at each stage. These individually rational payoffs place a lower bound on equilibrium payoffs that holds irrespective of patience. Such a lower bound does not exist in my setting because players minimise worst-case regret instead of maximising payoff.

6.3 Comparison to Ex-Post Equilibrium

Carroll (2020) introduced the model studied in this paper, and proposed the following solution concept.

Definition 10. *A strategy profile σ is an Ex-Post Equilibrium (XPE) if the restriction of σ to any environment e forms a SPE of the stochastic game where e_t played at time t with probability 1.*

Every XPE is a RPE because the regret from following the equilibrium strategy at any history is zero for any continuation environment, whereas the regret from deviations is nonnegative. XPE coincides with RPE when there is a single stage game in Θ – both equilibrium notions become equivalent to SPE. But in general, the RPE set may be much larger. For example, consider any XPE of a supergame that admits a trivial game in the sense of Definition 9. Since the strategies form a SPE in a continuation environment of trivial games, they must prescribe stage-game Nash Equilibrium behaviour at every history, regardless of patience. On the other hand, any action profiles in any game are implementable in RPE for high discount factors, as shown in Proposition 4.

It is possible to make a more detailed comparison by specialising to the case of a single long-lived player studied by Carroll (2020). Suppose players $2, .., n$

have discount factor 0, i.e. they are completely myopic. Let $\delta \in (0,1)$ denote the discount factor of the long-lived player 1. For any gap $w \in \mathbb{R}_+$ and game θ, let

$$A^*(\theta|w) = \{a \in A | (1-\delta)d_1(a,\theta) \leq \delta w, d_i(a,\theta) = 0 \ \forall i = 2,..,n\}$$

be the set of supportable actions in θ, accounting for the heterogeneity in discounting. Let

$$B^{XPE}w = \min_{\theta} \max_{a \in A^*(\theta|w)} u_1(a,\theta) - \min_{a \in A^*(\theta|w)} u_1(a,\theta),$$

and let $\bar{w}^{XPE}$ be the (unique) gap obtained by iterative application of B^{XPE} with starting gap $w^0 = 2M$, similarly to Lemma 2. Carroll (2020) shows that the set of XPE actions in any game θ is $A^*(\theta|\bar{w}^{XPE})$.[12]

On the other hand, Theorems 1 and 2 can be adapted to this setting to show that under Assumptions (A1) and (A2) the set of RPE actions in any game θ is $A^*(\theta|\bar{w})$, where

$$Bw = \max_{\theta} \max_{a \in A^*(\theta|w)} u_1(a,\theta) - \min_{a \in A^*(\theta|w)} u_1(a,\theta),$$

and $\bar{w}$ is obtained from the algorithmic procedure in Lemma 2 with A^* and B as defined above.

The differences between XPE and RPE can then be seen by comparing operators B^{XPE} and B. The former obtains the maximum payoff gap for the long-run player in the worst case among all stage games, whereas the latter obtains a best-case maximum gap. XPE is a more conservative solution concept because it requires that dynamic incentives work even in a worst-case environment. RPE is more permissive because dynamic incentives need only work in a single regret-maximising environment.

The comparison is less clear in the of multiple long-lived players. Krasikov and Lamba (2022) use a recursive operator similar to B^{XPE} that recurses on *common gaps* $w = (w_1,..,w_n)$ such that $w_1 = .. = w_n$. In symmetric games their operator becomes a counterpart of my operator B, replacing maximisation over stage games with minimisation as in the case of a single long-lived player. However, this method obtains only a subset of XPE except in special cases.[13]

[12] The results of Carroll (2020) are adapted in several ways. The notions of implementability of full and realisable outcomes in Carroll (2020) are stronger – they require neither the exclusion of games from full outcomes, nor any increase in the discount factor. Carroll (2020) also considers a different recursive operator

$$B^{XPE}w = \min_{\theta}(1-\delta)\left(\max_{a \in A^*(\theta|w)} u_1(a,\theta) - \min_{a \in A^*(\theta|w)} u_1(a,\theta)\right) + \delta w,$$

but my definition obtains the same gap $\bar{w}^{XPE}$ and, consequently, the same characterisation of XPE actions.

[13] Some of these special cases are strongly symmetric equilibria, linear Bertrand models, and high discount factors.

7 Conclusion

This paper extends the celebrated class of preferences for worst-case regret minimisation to dynamic interactions in ambiguous environments. I provide a characterisation of equilibrium actions for fixed discounting. Its tractability allows me to tackle common applications from stochastic games, and it even provides closed-form solutions in some cases. I believe this approach can be fruitful in other applications, even those where the stage game is dynamic such as relational contracting and Stackelberg games. It will be interesting to see how much of this tractability can translate to settings with imperfect monitoring or incomplete information.

References

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

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

Bergemann, D. and K. H. Schlag (2008): "Pricing without priors," *Journal of the European Economic Association*, 6, 560–569.

Carroll, G. (2019): "Robustness in mechanism design and contracting," *Annual Review of Economics*, 11, 139–166.

——— (2020): "Dynamic Incentives in Incompletely Specified Environments," Working Paper `http://individual.utoronto.ca/carroll/onelongrun.pdf`.

Cesa-Bianchi, N. and G. Lugosi (2006): *Prediction, learning, and games*, Cambridge university press.

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

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

Guo, Y. and E. Schmaya (2019): "Robust Monopoly Regulation," Working Paper `https://arxiv.org/abs/1910.04260`.

——— (2022): "Regret-Minimizing Project Choice," Working Paper `https://yingniguo.com/wp-content/uploads/2022/07/Regret-Minimizing-Project-Choice-07062022.pdf`.

Halpern, J. Y. and R. Pass (2012): "Iterated regret minimization: A new solution concept," *Games and Economic Behavior*, 74, 184–207.

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

Hurwicz, L. and L. Shapiro (1978): "Incentive structures maximizing residual gain under incomplete information," *The Bell Journal of Economics*, 180–191.

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

Krasikov, I. and R. Lamba (2022): "Uncertain Repeated Games," SSRN Working Paper `http://dx.doi.org/10.2139/ssrn.4222516`.

Libgober, J. and X. Mu (2021): "Informational robustness in intertemporal pricing," *The Review of Economic Studies*, 88, 1224–1252.

——— (2022): "Coasian Dynamics under Informational Robustness," Working Paper `https://arxiv.org/abs/2202.04616`.

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

McAdams, D. (2011): "Performance and turnover in a stochastic partnership," *American Economic Journal: Microeconomics*, 3, 107–142.

Munkres, J. (2013): *Topology: Pearson New International Edition*, Pearson Education Limited, second ed.

Renou, L. and K. H. Schlag (2010): "Minimax regret and strategic uncertainty," *Journal of Economic Theory*, 145, 264–286.

Savage, L. (1951): "The theory of statistical decision," *Journal of the American Statistical association*, 46, 55–67.

Shapley, L. (1953): "Stochastic games," *Proceedings of the national academy of sciences*, 39, 1095–1100.

Wald, A. (1950): *Statistical decision functions*, New York, Wiley.

Yeltekin, S., Y. Cai, and K. Judd (2017): "Computing equilibria of dynamic games," *Operations Research*, 65, 337–356.

Appendix

Proof of Proposition 1

Fix any i and h^t. Suppose $d_i(\sigma(h^t), \theta_t) > 0$ otherwise, the result is immediate by $w^* \geq 0$. Then there exists $a_i^d \neq \sigma_i(h^t)$ that is a best reply to $\sigma_{-i}(h^t)$ in θ_t. For any

environment e and $\varepsilon > 0$ it follows from (2) that there exists $a_i \in A_i$ such that i's regret from a one-shot deviation to a_i^d satisfies

$$\begin{aligned} R_i(a_i^d, \sigma_{-i}|h^t, e) \leq (1-\delta)\Big[u_i(a_i, \sigma_{-i}(h^t), \theta_t) - u_i(a_i^d, \sigma_{-i}(h^t), \theta_t)\Big] \\ + \delta\Big[U_i(\sigma|h^t, a_i, \sigma_{-i}(h^t), e) - U_i(\sigma|h^t, a_i^d, \sigma_{-i}(h^t), e)\Big] + \varepsilon. \end{aligned}$$

It also follows from (2) that i's regret from his equilibrium action satisfies

$$\begin{aligned} R_i(\sigma|h^t, e) \geq (1-\delta)\Big[u_i(a_i, \sigma_{-i}(h^t), \theta_t) - u_i(\sigma(h^t), \theta_t)\Big] \\ + \delta\Big[U_i(\sigma|h^t, a_i, \sigma_{-i}(h^t), e) - U_i(\sigma|h^t, \sigma(h^t), e)\Big]. \end{aligned}$$

Combining the above inequalities yields

$$\begin{aligned} R_i(a_i^d, \sigma_{-i}|h^t, e) &\leq R_i(\sigma|h^t, e) - (1-\delta)d_i(\sigma(h^t), \theta_t) \\ &\quad + \delta\Big[U_i(\sigma|h^t, \sigma(h^t), e) - U_i(\sigma|h^t, a_i^d, \sigma_{-i}(h^t), e)\Big] + \varepsilon \\ &\leq R_i(\sigma|h^t, e) - (1-\delta)d_i(\sigma(h^t), \theta_t) \\ &\quad + \delta\Big[\bar{U}_i(e) - \underline{U}_i(e)\Big] + \varepsilon \end{aligned}$$

Taking the supremum over e of both sides and using the equilibrium condition at h^t yields

$$0 \leq -(1-\delta)d_i(\sigma(h^t), \theta_t) + \delta w_i^* + \varepsilon$$

The desired result follows by taking $\varepsilon > 0$ arbitrarily small.

Proof of Lemma 1

The proof follows from the following chain of inequalities for all i

$$\begin{aligned} w_i^* &\equiv \sup_e \left(\bar{U}_i(e) - \underline{U}_i(e)\right) \\ &\leq \sup_e (1-\delta) \sum_{t=0}^{\infty} \delta^t \left(\max_{a \in A^*(e_t|w^*)} u(a, e_t) - \min_{a \in A^*(e_t|w^*)} u(a, e_t) \right) \\ &\leq \max_{\theta} \left(\max_{a \in A^*(\theta|w^*)} u(a, \theta) - \min_{a \in A^*(\theta|w^*)} u(a, \theta) \right) = (Bw^*)_i, \end{aligned}$$

where the first inequality follows from Proposition 1.

Proof of Theorem 2 – "only if" direction

For the first part, notice that Theorem 1 and implementability imply that $a_t \in A^*(e_t|\bar{w}(\delta'), \delta')$ for all $\delta' > \delta$. Similarly for the second part, $a(h^t) \in A^*(\theta_t|\bar{w}(\delta'), \delta')$ for all $\delta' > \delta$. Hence, it suffices for both results to show that

$$A^*(\theta|\bar{w}(\delta), \delta) = \bigcap_{\delta' > \delta} A^*(\theta|\bar{w}(\delta'), \delta')$$

for any game θ.

To this end, consider any gap w and sequences $(w_k), (\delta_k)$ such that $w_k \searrow w$ and $\delta_k \searrow \delta$. It follows from the monotonicity and continuity of A^* in the gap and discount factor that $A^*(\theta|w_k, \delta_k) \searrow A^*(\theta|w, \delta)$ for any θ. Hence, $B(w_k|\delta_k) \searrow B(w|\delta)$. It follows from the algorithmic characterisation of Lemma 2 that $\bar{w}(\delta_k) \searrow \bar{w}(\delta)$. By monotonicity and continuity $A^*(\theta|\bar{w}(\delta_k), \delta_k)$ decreases monotonically to $A^*(\theta|\bar{w}(\delta), \delta)$ for θ. The desired result follows because the sequence $\delta_k \searrow \delta$ is arbitrary.

Proof of Proposition 3

The proof is broken up into several lemmas mirroring results in Section 4. The first lemma is a counterpart to Proposition 1.

Lemma 3. *Let σ be a RPE. Then $\sigma(h^t) \in \underline{A}(\theta_t|w^*)$ for all h^t.*

Proof. Consider any i and h^t. Since actions sets are binary, there exists a unique action $a_i^d \in A_i$ distinct from $\sigma_i(h^t)$. Hence, the deviation regret at h^t satisfies

$$\begin{aligned} R_i(a_i^d|h^t, e) &= (1-\delta)\Big[u_i(\sigma(h^t), \theta_t) - u_i(a_i^d, \sigma_{-i}(h^t), \theta_t)\Big] \\ &\quad + \delta\Big[U_i(\sigma|h^t, \sigma(h^t), e) - U_i(\sigma|h^t, a_i^d, \sigma_{-i}(h^t), e)\Big] \\ &\leq -(1-\delta)d_i(\sigma(h^t), \theta_t) + \delta[\bar{U}_i(e) - \underline{U}_i(e)]. \end{aligned}$$

It follows that

$$\inf_{\sigma_i' \in \Sigma_i^d(\sigma_i, h^t)} \sup_e R_i(\sigma_i', \sigma_{-i}|h^t, e) \leq \delta w^* - (1-\delta)d_i(\sigma(h^t), \theta_t). \tag{9}$$

Let e_τ be an environment of trivial games. Then equilibrium regret satisfies

$$\begin{aligned} \sup_e R_i(\sigma|h^t, e) &\geq R_i(\sigma|h^t, e_\tau) \\ &= (1-\delta)\Big[u_i(a_i^d, \sigma_{-i}(h^t), \theta_t) - u_i(\sigma(h^t), \theta_t)\Big] = (1-\delta)d_i(\sigma(h^t), \theta_t). \end{aligned} \tag{10}$$

The result follows by combining (9) and (10). □

The next results are counterparts to Lemma 1 and Lemma 2. They can be proved similarly to their counterparts by replacing A^* with $\underline{A}$ and B with $\underline{B}$.

Lemma 4. $\underline{B}w^* \geq w^*$.

Lemma 5. *Let $w^0 = (2M, .., 2M)$. Define $w^{k+1} = \underline{B}w^k$ inductively for each $k = 0, 1, ...$ Then*

1. *(w^k) is decreasing and converges to some limit $\underline{w}$.*
2. *$\underline{B}\underline{w} = \underline{w}$.*
3. *$\underline{w} \geq w$ for any gap $w \leq w^0$ such that $\underline{B}w \geq w$.*

It follows from Lemmata 3, 4, and 5 that $\sigma(h^t) \in \underline{A}(\theta_t|\underline{w})$ for any RPE σ and history h^t. Proposition 3 then follows by the argument for the "if" direction of Theorem 2.

Proof of Proposition 4

Let $\underline{w}(\delta')$ denote the gap $\underline{w}$ when the discount factor is δ'. Since $\underline{A}$ is increasing in δ, it follows from Lemma 5 that $\underline{w}(\delta)$ is increasing in δ. Fix a player i.

Case 1: Suppose $\underline{w}_i(\delta') = 0$ for all δ'. Consider the algorithmic procedure in Lemma 5. By monotonicity of the sequence (w^k) it must be that $u_i(a,\theta) = u_i(a',\theta)$ for any game θ and action profiles $a, a' \in \underline{A}(\theta|w^0)$. But $\underline{A}(\theta|w^0) = A$ when the discount factor is above $1/2$. Thus, $d_i(a,\theta) = 0$ for all a, θ.

Case 2: Suppose $\delta'\underline{w}_i(\delta') \geq \varepsilon > 0$ for some δ'. Let $\bar{\delta} \geq \delta'$ satisfy $4(1-\bar{\delta})M < \varepsilon$. It follows that

$$2(1-\delta)d_i(a,\theta) \leq 4(1-\delta)M < 4(1-\bar{\delta})M < \varepsilon \leq \delta'\underline{w}_i(\delta') \leq \delta\underline{w}_i(\delta)$$

for any $\delta > \bar{\delta}$, a, and θ.

Hence, in both cases

$$\sup_{a,\theta} 2(1-\delta)d_i(a,\theta) \leq \delta\underline{w}_i(\delta)$$

for all i whenever $\delta > \bar{\delta}$. The proof now follows from Proposition 2.