Dynamic Incentives in Incompletely Specified Environments

Gabriel Carroll, Stanford University

gdc@stanford.edu

July 20, 2020

Abstract

Consider a repeated interaction where it is unknown which of various stage games will be played each period. This framework captures the logic of intertemporal incentives even though numeric payoffs to any strategy profile are indeterminate. A natural solution concept is ex post perfect equilibrium (XPE): strategies must form a subgame-perfect equilibrium for any realization of the sequence of stage games. When (i) there is one long-run player and others are short-run, and (ii) public randomization is available, we can adapt the standard recursive approach to determine the maximum sustainable gap between reward and punishment. This leads to an explicit characterization of what outcomes are supportable in equilibrium, and an optimal penal code that supports them. Any non-XPE-supportable outcome fails to be an SPE outcome for some (possibly ambiguous) specification of the stage games. Unlike in standard repeated games, restrictions (i) and (ii) are crucial.

Thanks to (in random order) Drew Fudenberg, Andrzej Skrzypacz, Paul Milgrom, and Takuo Sugaya for discussions, as well as seminar audiences at UCSD, Columbia, and Rice.
The author is supported by an NSF CAREER award.

1 Introduction

This paper studies a model of repeated interactions in which it is unknown what stage game will be played in each period. The stage game may vary from one period to the next, and there is no prior over the process determining its evolution.

Repeated games are the classic theoretical paradigm for studying how, and to what extent, non-myopic behavior can be incentivized by the promise of future rewards. Typically, the analyst models an interaction by writing down some game and asks what outcomes can be supported in equilibrium. The standard theory offers, at least in principle, a recipe for answering this question: first, use the recursive analysis of Abreu, Pearce and Stacchetti (1990) (henceforth APS) to determine the worst possible equilibrium payoff for each player (which may be strictly worse than just repeatedly playing a static Nash equilibrium); then, the possible outcome paths are identified as those for which deviations can be deterred using this worst equilibrium as punishment (Abreu, 1988).

However, the canonical model makes the rigid assumption that players play exactly the same game over and over. In reality, the nature of interactions between players may vary over time, and the players may not be able to precisely describe or agree on its future evolution. We might expect that the basic principle of dynamic incentives—that a player can be induced to forgo a short-run temptation of, say, 3 payoff units if he can be promised a reward of $3/\delta$ in the future—should remain valid even when the future evolution of the environment cannot be fully specified. It is therefore natural to explore such incompletely-specified models to try to express this intuition. And once we have formulated such a model, it is natural to ask more generally to what extent the analytical tools from the standard theory carry over.

An important feature of our setting is that a strategy profile does not determine a numeric payoff for each player, since the payoffs depend on what stage games arise. Much of the standard toolkit for studying repeated games, such as the APS recursive characterization, works in payoff space. Thus, another perspective on the analysis here is that it explores how much of this toolkit can be developed without reference to payoff levels.

Our main focus here will be on repeated interactions with a fixed discount factor $\delta < 1$, and the following four features:

(i) One long-run strategic player interacts with a series of short-run players (as in e.g. Fudenberg, Kreps and Maskin (1990)).

(ii) A public randomization device is available.

(iii) Attention is restricted to pure strategies (conditional on the public randomization).

(iv) Actions are perfectly observed.

We will show how, in this setting, the recursive method can be adapted and used to characterize what outcomes are supportable in equilibrium.

Feature (i) is of course restrictive, but it still gives plenty of scope for studying the structure of dynamic incentives. As usual in the literature, this formulation allows multiple interpretations: the short-run players may be different individuals in each period, they may be long-lived but completely impatient individuals, or they may be a continuum of players in the same role and whose individual deviations are not detectable. This assumption can be varied somewhat; for example, the techniques developed here should also be applicable with two long-run players, one of whom has full commitment. The crucial feature is that there is only one player who needs to be given dynamic incentives. As will be explained below, the analysis does not extend to multiple such players.

Assumption (ii) is also crucial. This is in contrast to the usual setting of repeated games, where allowing for public randomization is mostly a technical convenience (for example, the APS recursive analysis can be carried out with or without public randomization). Here, this assumption is a substantive necessity, as described below.

Assumptions (iii) and (iv) are made mostly for ease of exposition; many of the ideas here could be developed without them. Admittedly, many of the interesting applications of repeated games involve imperfect monitoring, but it seems clearest to develop the conceptual machinery first in the simpler setting considered here.

The interaction assumed here has the following structure: There is a known set of stage games that may arise. Each period, the players all observe what stage game comes up. They then choose their actions and receive their payoffs. The long-run player maximizes the δ-discounted sum of stage payoffs. Although the process governing the stage games in each period is unspecified, it is common knowledge that future stage games do not depend on players' past actions (unlike in stochastic games).

We will adopt a solution concept based on the idea that the long-run player can be induced to forgo a short-run gain if it is promised that future play will compensate him enough to deter him from deviating, no matter what later stage games may come up. Formally, this solution concept, *ex-post perfect equilibrium* (XPE), requires that the strategies should form a subgame-perfect equilibrium for every sequence of stage games that may be realized. We wish to characterize what outcomes can be supported by XPE strategies.

Note that this solution concept is not rooted in a model of individual maximization.[1] In principle, one could demand that if players have uncertainty about future stage

[1]In particular, XPE is strictly stronger than just requiring maximization in the worst case over future

games, they should have beliefs (and higher-order beliefs) about them; even if they are not Bayesian, they should have preferences over acts on future stage games, specifying how they trade off a better outcome in one possible future environment against a worse outcome in another, and they should maximize those preferences. One possible interpretation of XPE is that it gives players a simple way to coordinate on self-enforcing strategies, without needing to think about their beliefs (or each other's). An alternative interpretation is that the players do know the stage game process, but the *analyst* is uncertain, and she would like to offer a simple description of strategies that demonstrate that some amount of non-myopic behavior is supportable, where "simplicity" is operationalized by requiring that behavior should not depend on the players' knowledge about the future.

The key analytic technique is to adapt a version of the recursive characterization from APS to our setting. We characterize the set of values of w such that it is possible to find two XPE profiles, one "reward" and one "punishment," such that the reward gives the long-run player a payoff at least w more than the punishment no matter what stage games are realized. Thus, instead of recursing on continuation values, we recurse on the reward-punishment gap.

This leads to our first main result, an explicit characterization of the outcomes that can be supported in XPE. In the leading case where on-path behavior does not condition on the public randomization, there is an especially intuitive description of such outcomes: they are the ones in which, at each period, the "debt" owed to the long-run player for forgoing short-run gains in past periods never accumulates beyond the maximum sustainable reward-punishment gap. A special case of this result applies when only one stage game is possible, in which case the result characterizes the SPE outcomes of a traditional repeated game with only one long-run player; this description does not seem to exist in the literature and may be independently worthwhile.

Although this result is cast as describing the supportable outcomes for a given specification of uncertainty, we can equivalently view it as characterizing the extent of uncertainty that is consistent with sustaining a particular outcome. For example, in the "product-choice game" traditionally used as an example of an interaction between long-run and short-run players, we can use this to ask: what must the consumers believe about the firm's future opportunities in order to be persuaded that the firm has enough incentives to provide high quality in the present period?

A second result, which falls out of the proof of the first, is the existence of an optimal penal code that can be used as punishment to support any XPE outcome, as in Abreu

stage games.

(1988). This optimal penal code gives the long-run player his worst outcome among all XPE, *no matter* what stage games are realized. These two results together illustrate how classic ideas from repeated games successfully carry over to our framework.

Because, as observed above, the XPE solution concept is not based on individual maximization, one might well ask about its relevance for positive prediction. In particular, it is clear that any outcome that is supportable in XPE can be achieved no matter what players believe about the stage game process. But might the same be true for outcomes that are *not* supportable in XPE? In fact, if players have standard Bayesian beliefs and maximize expected utility, there may indeed be outcomes that are not supportable in XPE, yet are supportable in SPE no matter what these beliefs are. However, if we broaden the possible preferences to allow for ambiguity aversion, this is no longer the case: any outcome that is not supportable in XPE is not even supportable in SPE under some specification of preferences. This is our third result, and it gives a reason why characterizing the XPE outcomes is relevant even for an analyst who views SPE, rather than XPE, as the correct description of behavior.

As mentioned above, our recursive analysis relies both on having a single long-run player and on the availability of public randomization. Dropping either of these assumptions would lead the analysis of this paper to break down. Although it is unclear how to formulate a theorem that no recursive characterization exists, we can show concretely that the theory fails to carry over by demonstrating that optimal penal codes can fail to exist when either assumption is dropped. Since the optimal penal code plays a central role in the proof of the description of supportable outcomes, this suggests that such a description in general, if one can be given, would have to look quite different. Section 6 provides the relevant counterexamples, and offers some discussion of why both assumptions are crucial to the recursive technique. This contrasts with standard repeated games, where either or both assumptions can be dropped without trouble.

The rest of the paper proceeds linearly: first an illustrative example, then the model, analysis, results, and discussion. Literature will be discussed as it comes up.

2 Illustration

This extended example demonstrates the central questions and novel features of our setting.

Example 2.1. We begin with a typical specification of the "product-choice" game often

used to illustrate long-run / short-run player models (e.g. Mailath and Samuelson, 2006, Section 1.5). Player 1 (the long-run player) is a firm, who can produce low-quality or high-quality products in each period; in each period, player 2 (the short-run player) is a buyer who can buy either an expensive or a cheap product, without seeing the quality in advance. The products are priced at 6 and 0 respectively; an expensive product is worth 8 to the buyer if high quality and 0 if low, while a cheap one is always worth 0. For the firm, producing high quality costs 2 while low quality is costless. This gives rise to the net payoffs shown in the matrix G at the left side of Figure 1. In the repeated game, this stage game is played each period, past actions are observed, and the firm maximizes the δ-discounted sum of stage payoffs.

G:

	e	c
h	$4,2$	$-2,0$
l	$6,-6$	$0,0$

G':

	e	d	c
h	$4,2$	$1,1$	$-2,0$
l	$6,-2$	$3,1$	$0,0$
s	$4,-6$	$1,-3$	$-2,0$

G'':

	e	d	c
h	$5,2$	$3,\frac{4}{3}$	$-1,0$
l	$6,-\frac{2}{3}$	$4,\frac{4}{3}$	$0,0$
s	$3,-6$	$1,-4$	$-3,0$

Figure 1: Variants of the product-choice game.

Given the pure-strategy restriction and the fact that the short-run players must be best-responding, only he or lc can ever be played in equilibrium. The unique stage Nash equilibrium is lc, but if $\delta \geq 1/3$, the "cooperative" outcome he can be sustained in equilibrium by the threat of reversion to lc if player 1 ever deviates.

As is well-known, however, more complex punishments can often support cooperation more effectively than Nash reversion, and the middle game G' in the figure illustrates this. For a story behind this game, imagine that each buyer may be either a "discerning" type who values the products as before, or an "undiscerning" type who values the expensive good at 8 regardless of its quality. The firm does not know the buyer's type; each type has probability 1/2. In addition, we give the firm an extra "sabotage" action which is also costly (it costs 2) but makes both products worth 0 to both buyer types. The firm now has three actions: (h)igh quality, (l)ow, or (s)abotage; and the buyer has three: (e)xpensive, (d)ifferentiate (i.e. buy expensive iff the undiscerning type), and (c)heap. The payoffs are as shown.

Three action profiles satisfy the buyer's best-response constraint, namely he, ld, sc. The unique stage Nash is ld. The cooperative outcome he can be sustained by the threat of reversion to stage Nash only for $\delta \geq 2/3$: otherwise, the firm's short-run gain of 2 by deviating to l is too tempting relative to the loss of 1 in future periods. However,

as long as $\delta \geq 1/3$, the cooperative outcome can be supported by the "carrot-and-stick" punishment wherein, if the firm ever deviates, then *sc* is played for one period, followed by a return to *he* in subsequent periods. If the firm deviates when *sc* is supposed to be played, then we again specify punishing with *sc* for one period (and then returning to *he*), and so forth. This works because, both in the cooperation and in the punishment phases, the short-run gain of 2 from deviating is outweighted by the loss of 6 next period (resulting from playing *sc* instead of *he*). Note also that if $\delta < 1/3$, then cooperation can never be sustained. Indeed, the firm can always guarantee itself at least 0 by playing l in every period. Since the equilibrium payoff can never be above 4 (due to the buyer's best-response constraint), the punishment for a deviation cannot exceed 4 in all subsequent periods. When $\delta < 1/3$, a short-run gain of 2 cannot be deterred by such a punishment, and so the only equilibrium outcome is the stage Nash *ld* forever.

The right game G'' is similar, but with a few changes: high quality now costs 1 rather than 2; sabotage costs 3; and the undiscerning type now arrives with probability 2/3. Suppose this game is played each period. Again, *he*, *ld* or *sc* must be played. Here, the carrot-and-stick strategies support the *he* outcome for $\delta \geq 3/8$. For $\delta < 3/8$, only stage Nash is possible: First note that since the long-run player can never get more than 5, and can be assured at least 0 by playing l, no possible future punishment can offset a short-run deviation gain of 3. Hence, *sc* can never arise in equilibrium, as the incentive to deviate to l is too strong. So the only profile available as punishment is the stage Nash *ld*, but this is insufficient to deter deviations from *he* when $\delta < 1/2$.

Thus, in repeated game G', the cooperative outcome is supportable iff $\delta \geq 1/3$, and in G'' it is supportable iff $\delta \geq 3/8$.

Now we turn to the focus of this paper: suppose that each period, G' or G'' is to be played, but it is not known in advance which one. One of the two games arrives, the players observe it and then make their moves.

Note that the carrot-and-stick strategies now support *he* as long as $\delta \geq 1/2$. Indeed, in either game and in either phase (cooperation or punishment), the short-run gain from deviating is at most 3, whereas the loss the following period from playing *sc* instead of *he* will be at least 6, which outweighs the gain. Thus, we can conclude confidently that these strategies form an equilibrium, even though we cannot compute the payoffs without knowing the realized sequence of games. Conversely, for $\delta < 1/2$, these strategies are not assured to work: If, when time for punishment comes, the players find themselves in G'' but expect to be in G' the following period, then the firm's short-run gain of 3 from deviating from *sc* outweighs the loss of 6 next period. Consequently, in the cooperative

phase, the firm cannot be trusted to play h when it should, since it may not be possible to sustain punishment the next period.

In fact, if $\delta < 3/7$, only stage Nash is sustainable without knowing the stage games in advance. To see this, suppose G'' is to be played today but the firm anticipates G' in all future periods. The maximum punishment inflictable in each future period is 4 (since the firm will get at most 4 on-path, and can assure itself at least 0 by playing l); hence, for $\delta < 3/7$, sc cannot be played today, since deviation cannot be adequately punished. This means that there is no way to get sc to be played when G'' arrives. Then, he can never be sustained in either game (nor, for that matter, sc in G') because the buyer worries that the firm may expect G'' in all future periods, so that the largest possible punishment is a loss of 1 in future periods (from he to ld), not enough to deter the short-run deviation. Notice, in particular, that for $\delta \in [3/8, 3/7)$, the good outcome he can never arise in equilibrium of this game with uncertainty, even though it was supportable both in the repeated game with G' played every period and with G'' played every period.

Finally, what about $\delta \in [3/7, 1/2)$? It turns out that he can be supported in equilibrium (both when G' arrives and when G'' arrives), even though the carrot-and-stick punishments no longer do the trick. As we shall see later, it can be supported with more complex punishments that randomize the timing of the return to the cooperative phase (thus using the public randomization) in order to make the overall punishment more severe, and therefore better able to discourage the most tempting deviations (namely, deviations from sc in G'').

△

In the following sections, we study the general version of the questions we have explored here: Given a set of stage games that may arrive, and a discount factor, how can the analyst figure out what outcomes are supportable in equilibrium (and how they can be supported)?

3 Model

We proceed by first developing the model in terms of a standard repeated-game setup. The notation will largely follow Mailath and Samuelson (2006), suitably adapted for our framework of uncertainty. We then introduce some adaptations to notation that will be a bit more convenient for our focus on a single player's long-run incentives.

3.1 Standard formulation

There are $n \geq 2$ players. Player 1 is a long-run player, with a discount factor $\delta \in (0,1)$, and the others are short-run players. As usual, we can be agnostic as to whether player $i > 1$ in period t is physically the same person (or persons) as player i in period t', but it is notationally simpler to use the same label i for both. There is a nonempty set $\mathcal{G}$ of possible *stage games*. In any stage game $G \in \mathcal{G}$, we denote the set of actions available to player i as $A_i(G)$. We assume that actions are labeled so that $A_i(G)$ and $A_i(G')$ are disjoint for $G \neq G'$; this makes the definitions up front slightly cumbersome but will simplify notation later. Write $A(G) = \times_{i=1}^n A_i(G)$. Also, write $A_i = \cup_{G \in \mathcal{G}} A_i(G)$ for the set of all actions that i can ever play, and likewise $A = \cup_G A(G)$. Then, player i's stage payoff function is simply written $u_i : A \to \mathbb{R}$. We assume a uniform bound M on the possible stage payoffs: $|u_i(a)| \leq M$ for all i, a. All these objects are exogenously given primitives. We assume that each $A_i(G)$ is a compact metric space, and that $u_i(a)$ is continuous on $A(G)$ for each G. (Finite action sets are a special case.) We equip A_i and A with their disjoint union topologies.

In the repeated game, in each period $t = 0, 1, 2, \ldots$, the players observe the realized stage game $G^t \in \mathcal{G}$, as well as the public randomization signal $\omega^t \sim U[0,1]$, and then they simultaneously choose actions. Thus, a history at time t consists of the stage games, public random signals, and actions at past dates, together with the stage game and random signal at the present date. So the set of time-t histories is

$$H^t = \left(\cup_{G \in \mathcal{G}} \left(\{G\} \times [0,1] \times A(G)\right)\right)^t \times (\mathcal{G} \times [0,1])$$

with representative element

$$h^t = (G^0, \omega^0, a^0;\ G^1, \omega^1, a^1;\ \ldots;\ G^{t-1}, \omega^{t-1}, a^{t-1};\ G^t, \omega^t).$$

We focus on pure strategies; thus, a strategy for player i is a measurable function $s_i : \cup_{t=0}^{\infty} H^t \to A_i$, such that $s_i(h^t) \in A_i(G^t)$ whenever the history h^t ends in G^t. A strategy profile takes the form $s = (s_1, \ldots, s_n)$, or can be equivalently written $s : \cup_t H^t \to A$, with the corresponding restriction $s(h^t) \in A(G^t)$. It will sometimes be useful to abbreviate a finite history of random signals $(\omega^0, \ldots, \omega^t)$ by $\omega^{0,\ldots,t}$, and to write $\mathbb{E}^t[\cdots]$ for the time-t expectation operator (i.e. the expectation conditional on signals $\omega^{0,\ldots,t}$).

We refer to a realization of the sequence of stage games as an *environment*, $E = (G^0, G^1, \ldots)$. A history h^t is *consistent* with the environment E if the stage games ap-

pearing at all periods $0, 1, \ldots, t$ in h^t are the same as those specified in E. Given a strategy profile s, an environment E, and a history $h^t = (G^0, \omega^0, a^0; \ldots; G^t, \omega^t)$ that is consistent with E, we define subgame payoffs as follows. For any realization path $(\omega^{t+1}, \omega^{t+2}, \ldots)$ for the subsequent random signals, we can recursively define the action profiles $a^{t'} = s(G^0, \omega^0, a^0; \ldots; G^{t'}, \omega^{t'})$ for each $t' \geq t$. Then, player 1's subgame payoff at h^t is the (normalized) discounted sum of stage payoffs

$$U_1(s|E, h^t) = (1-\delta)\mathbb{E}\left[\sum_{t'=t}^{\infty} \delta^{t'-t} u_1(a^{t'})\right],$$

where the expectation is over the (future) public randomization. Player i's payoff, for each $i > 1$, is simply

$$U_i(s|E, h^t) = u_i(a^t).$$

Given environment E, strategy profile s is a *subgame-perfect equilibrium* (SPE) for E if, for each player i, each history h^t consistent with E, and each alternative strategy s_i',

$$U_i(s|E, h^t) \geq U_i(s_i', s_{-i}|E, h^t). \tag{3.1}$$

The usual arguments for the one-shot deviation principle apply: it suffices to have (3.1) hold for all h^t consistent with E and all s_i' that differ from s_i only at the history h^t.

We can also define player 1's continuation payoff in environment E, following a history h^t consistent with E and an action profile a^t, as

$$U_1(s|E, h^t, a^t) = (1-\delta)\mathbb{E}\left[\sum_{t'=t+1}^{\infty} \delta^{t'-(t+1)} u_1(a^{t'})\right],$$

where, again, the expectation is over the public random signals $(\omega^{t+1}, \omega^{t+2}, \ldots)$, and the future actions are determined by beginning from h^t followed by a^t and then playing according to s. This quantity is not part of the definition of SPE, but it is relevant to player 1's incentives to deviate: (3.1) is satisfied for one-shot deviations by player 1 at h^t if and only if

$$(1-\delta)u_1(s(h^t)) + \delta U_1(s|E, h^t, s(h^t)) \geq (1-\delta)u_1(a_1', s_{-1}(h^t)) + \delta U_1(s|E, h^t, (a_1', s_{-1}(h^t)))$$

for all deviations a_1'. Similarly, we can define

$$U_1(s|E) = (1-\delta)\mathbb{E}\left[\sum_{t=0}^{\infty}\delta^t u_1(a^t)\right],$$

the expected payoff from the beginning of the game in environment E.

Strategy profile s is an *ex-post perfect equilibrium* (XPE) if it is an SPE for every environment.[2] Later, we will indicate sufficient conditions on primitives to ensure that an XPE exists.

3.2 More convenient notation

We can apply a standard simplification for games with short-run players (e.g. Fudenberg, Kreps and Maskin, 1990): For each $G \in \mathcal{G}$, let $A^*(G)$ be the set of action profiles at which no short-run player wishes to deviate,

$$A^*(G) = \{a \in A(G) \mid u_i(a) \geq u_i(a_i', a_{-i}) \text{ for all } i > 1, a_i' \in A_i(G)\}.$$

Evidently, the constraints (3.1) for the short-run players are satisfied iff $s(h^t) \in A^*(G^t)$ for all histories h^t (consistent with the environment E).

With this in mind, we can now dispense with explicit consideration of the short-run players' incentives, focusing only on the long-run player. We accordingly drop the player subscript for payoffs: henceforth, we write u and U rather than u_1 and U_1 unless there is ambiguity.

We can summarize the above as

Lemma 3.1. *Strategy profile s is an XPE if and only if both the following conditions hold:*

1. *for every history h^t, with stage game G^t arising at time t, we have $s(h^t) \in A^*(G^t)$;*
2. *for every environment E, every history h^t consistent with E, and every possible deviation s_1' by player 1 that differs from s_1 only at history s_1', we have $U(s|E, h^t) \geq U(s_1', s_{-1}|E, h^t)$.*

[2]The terminology is inspired by that of Fudenberg and Yamamoto (2010), who study a repeated game in which the stage game is fixed over time but unknown; their equilibrium concept requires subgame-perfection for each such game. Some literature has used the name "belief-free equilibrium" for related concepts, e.g. Ely, Hörner and Olszewski (2005); Hörner and Lovo (2009).

Notice that the set of XPE has a recursive structure: s is an XPE if it meets conditions (1)–(2) at every period-0 history and each continuation strategy profile starting from date 1 is an XPE.

In addition, when $a \in A(G)$, let us write $\widehat{u}(a) = \max_{a_1' \in A_1(G)} u(a_1', a_{-1})$ for the stage payoff that would result from the myopically optimal deviation from a. (Here and henceforth, we take "myopically optimal deviation" to mean "conforming" when 1's action is already a best reply in the stage game.) Clearly $\widehat{u}(a) \geq u(a)$, and $\widehat{u}$ is again continuous on $A(G)$. Although it makes no difference formally, a conceptual reframing may be helpful: rather than think of action profiles as consisting specifically of an action by each player, and contemplating explicit deviations by player 1, we may think of action profiles (that may arise in equilibrium) in a stage game G simply as abstract objects belonging to a set $A^*(G)$, and focus on $\widehat{u}(a)$ as the quantity relevant to player 1's incentive to deviate.

Finally, if $E = (G^0, G^1, G^2, \ldots)$, it will be useful to write $E^{-t} = (G^t, G^{t+1}, \ldots)$, the continuation environment starting in period t, and to further abbreviate E^{-1} as simply E^-.

4 Analysis

4.1 Recursive technique

Player 1 can be dissuaded from a deviation that earns a short-term gain of g only if doing so reduces the continuation payoff by at least $\frac{1-\delta}{\delta} g$ in every possible environment. This suggests trying to find the largest "gap" $w \geq 0$ such that there exist two XPE's, say $\overline{s}$ and $\underline{s}$, such that $U(\overline{s}|E) - U(\underline{s}|E) \geq w$ for every environment E; doing so then lets us rule out some action profiles because deviation cannot be prevented.

We adapt the recursive machinery from APS to describe the set of such values w. Their B operator for n-player games maps subsets of $\mathbb{R}^n$ to subsets of $\mathbb{R}^n$. Here, we are concerned only with one long-run player, so the recursion is done on subsets of $\mathbb{R}$. Moreover, public randomization makes our set convex, hence an interval, and its lower bound is zero. So we only need to keep track of the upper bound, i.e. a single number.

With this in mind, we first define, for any $w \geq 0$ and $G \in \mathcal{G}$,

$$A^*(G, w) = \left\{ a \in A^*(G) \mid \widehat{u}(a) - u(a) \leq \frac{\delta}{1-\delta} w \right\}.$$

Now define

$$B(w;G) = (1-\delta)\left(\max_{a\in A^*(G,w)} u(a) - \min_{a'\in A^*(G,w)} \widehat{u}(a')\right) + \delta w. \tag{4.1}$$

(If $A^*(G,w)$ is empty, then take $B(w;G) = -\infty$. Note that as long as $A^*(G,w)$ is nonempty, it is closed, and the max and min exist by continuity.)

Intuitively, this $B(w;G)$ represents the largest possible gap in 1's payoff between two different strategy profiles, given that G is played at date 0, the date-0 incentive constraints must be satisfied, and all continuation payoffs starting from date 1 must lie within an interval of width w. Indeed, these last two requirements together imply that both profiles must specify an action in $A^*(G,w)$ at date 0. Moreover, the payoff from following the "bad" strategy profile cannot be less than the payoff from a date-0 deviation; thus the payoff gap between the good and bad strategy profiles is at most the gap between conforming to the good profile and deviating from the bad profile. Decomposing this gap into its period-0 component and its continuation component produces the two terms on the right side of (4.1).

The above argument sketches why the expression in (4.1) is an upper bound on the payoff gap between two strategy profiles, and suggests how to attain it: Normalizing the interval of allowable continuation payoffs to $[0,w]$, specify that the "good" profile begins with the a attaining the max in (4.1) and promises a continuation payoff of w if 1 conforms; the "bad" profile begins with the a' attaining the min and promises a continuation payoff of 0 if 1 deviates. To ensure the correct gap in the *on-path* payoffs, the continuation payoff after conforming in the "bad" profile should be set so that 1 is indifferent between initially conforming and deviating. This can indeed be done (the fact that $a' \in A^*(G,w)$ ensures that this continuation is at most w). Note that public randomization is essential for this, as it ensures that the set of allowable continuation payoffs is an interval. We will revisit this point in Section 6.

Now define

$$B(w) = \inf_{G\in\mathcal{G}} B(w;G).$$

This is the maximum payoff gap that can be guaranteed regardless of what stage game arrives in the initial period, given that continuation payoffs lie in an interval of width w.

Notice that $B(w;G)$ is strictly increasing in w at a rate of at least δ (the first term of (4.1) is weakly increasing because $A^*(G,w)$ is increasing in w, and the second term is clearly increasing at rate δ). Therefore, $B(w)$ is as well.

We now adopt an assumption that will be maintained for the rest of the paper:

Assumption 4.1. *There exists $w \geq 0$ such that $B(w) \geq w$.*

As we shall see, this assumption will imply that an XPE exists (and in fact, the converse is also true).

As an aside, either of the following sufficient conditions on primitives implies that Assumption 4.1 is satisfied:

1. For every $G \in \mathcal{G}$, there exists $a \in A^*(G)$ such that $\widehat{u}(a) = u(a)$ (i.e. a stage Nash equilibrium).

   (This ensures the assumption holds with $w = 0$.)

2. There exists $\epsilon > 0$ such that, for every $G \in \mathcal{G}$, there exist $a, a' \in A^*(G)$ with $u(a) \geq \widehat{u}(a') + \epsilon$, and $\delta \geq \frac{2M}{2M+\epsilon}$.

   (In this case, $A^*(G, \epsilon) = A^*(G)$ for all G, and then $B(\epsilon; G) \geq \epsilon$ for all G, so we can take $w = \epsilon$.)

However, rather than adopt either of these, we will just make Assumption 4.1 directly.

Let w^* be the largest value such that $B(w) \geq w$. It is straightforward that this maximum indeed exists, and that in fact $B(w^*) = w^*$.

This w^* is the limiting value of a recursion. To show this, we need a continuity argument (our analogue to Theorem 5 of APS):

Lemma 4.2. *The functions $B(w; G)$ and $B(w)$ are right-continuous in w.*

(The proof of this result, and all others not given in the text, are in Appendix A.)

With this property, one can readily show that starting with a value of w large enough to be an upper bound for w^*, for example any $w_0 > 2M$ (note that indeed $w > 2M$ implies $B(w; G) < w$ for each G), and then iterating B gives a decreasing sequence that converges to w^*. However, for technical reasons, it will be useful to take a slightly different sequence, one in which w_{k+1} is strictly above $B(w_k)$. Specifically:

Lemma 4.3. *Define a sequence as follows: $w_0 > 2M$, $w_1 \in (B(w_0), w_0)$, and for $k = 2, 3, \ldots$, put $w_k = (B(w_{k-1}) + B(w_{k-2}))/2$. Then:*

1. *$w_0 > w_1 > w_2 > \cdots$;*

2. *$w_k > B(w_{k-1})$ for $k \geq 1$;*

3. $w_k \to w^*$.

We can now show that there is no way to guarantee a payoff gap between two different XPE's of more than w^*. In fact, a stronger statement is true: For any $\epsilon > 0$, we can find an "adversarial" environment such that, in this environment, even if any SPE is allowed, the largest and smallest attainable payoffs differ by less than $w^* + \epsilon$.

Lemma 4.4. *Given any $\epsilon > 0$, there exists a finite T and a sequence of stage games $G^0, G^1, \ldots, G^T \in \mathcal{G}$ with the following property: For any environment E that begins with stage games $G^0, \ldots, G^T$, and any two SPE's $\overline{s}$ and $\underline{s}$ for this environment,*

$$U(\overline{s}|E) - U(\underline{s}|E) < w^* + \epsilon.$$

The proof uses the sequence from Lemma 4.3. We show by induction that there is an adversarial environment that prevents the payoff gap from exceeding w_k. In particular, since $w_k > B(w_{k-1})$, we can choose a stage game G such that $B(w_{k-1}; G) < w_k$. Then, if G is played in the initial period, and subsequent periods feature the sequence of stage games that prevents a gap of more than w_{k-1} (which exists by the induction hypothesis), then the total payoff gap cannot exceed w_k.

Proof of Lemma 4.4. For each $k = 1, 2, \ldots$, let $\overline{G}_k \in \mathcal{G}$ be such that $B(w_{k-1}; \overline{G}_k) < w_k$; this exists by Lemma 4.3 part 2. We will show that in any environment that begins with the stage games $\overline{G}_k, \overline{G}_{k-1}, \ldots, \overline{G}_1$ (in that order), the payoffs from any two SPE's differ by less than w_k. Since $w_k \to w^*$, the lemma then follows, by taking k large enough relative to ϵ.

We prove the statement by induction on k. The base case $k = 0$ is trivial, since in any environment at all, the payoffs of any two action profiles within a stage differ by at most $2M < w_0$, and therefore the same is true for the payoffs of any two SPE's. Now suppose the statement holds for $k - 1$. Consider an environment E beginning with $\overline{G}_k, \overline{G}_{k-1}, \ldots, \overline{G}_1$.

Let s be any SPE. Let a^0 be the action profile played at some date-0 history $h^0 = (\overline{G}_k, \omega^0)$, and a_1' be player 1's myopically optimal deviation; the incentive constraint reads

$$(1-\delta)u(a^0) + \delta U(s|E, h^0, a^0) \geq (1-\delta)u(a_1', a_{-1}^0) + \delta U(s|E, h^0, (a_1', a_{-1}^0))$$

or, rearranging,

$$(1-\delta)(u(a_1', a_{-1}^0) - u(a^0)) \leq \delta(U(s|E, h^0, a^0) - U(s|E, h^0, (a_1', a_{-1}^0))).$$

The left side is $(1-\delta)(\widehat{u}(a^0) - u(a^0))$, while the right side is δ times the difference of two SPE payoffs in the continuation environment E^-, and so is less than δw_{k-1} by the induction hypothesis. Hence, a^0 must lie in $A^*(\overline{G}_k, w_{k-1})$. That is, only action profiles in $A^*(\overline{G}_k, w_{k-1})$ can be played at date 0 in SPE.

Now let $\overline{s}, \underline{s}$ be two different SPE's. The payoff from $\overline{s}$ is

$$\mathbb{E}[(1-\delta)u(a^0) + \delta U(\overline{s}|E, \overline{G}_k, \omega^0, a^0)]$$

(where the expectation is over the random signal ω^0 and resulting action profile a^0)

$$\leq (1-\delta) \max_{a \in A^*(\overline{G}_k, w_{k-1})} u(a) + \delta \sup_{s' \text{ is SPE for } E^-} U(s'|E^-).$$

Likewise, the payoff from $\underline{s}$ is at least the payoff from deviating to the myopically action a_1' in date 0, which is

$$\mathbb{E}[(1-\delta)\widehat{u}(a^0) + \delta U(\underline{s}|E, \overline{G}_k, \omega^0, (a_1', a_{-1}^0))]$$

(note that a^0 is now determined by $\underline{s}$ instead of $\overline{s}$)

$$\geq (1-\delta) \min_{a \in A^*(\overline{G}_k, w_{k-1})} \widehat{u}(a) + \delta \inf_{s' \text{ is SPE for } E^-} U(s'|E^-).$$

Subtracting, and using the fact that two different SPE payoffs in environment E^- differ by at most w_{k-1} by induction, gives us exactly

$$U(\overline{s}|E) - U(\underline{s}|E) \leq B(w_{k-1}, \overline{G}_k).$$

Since this is less than w_k, the desired statement follows.

□

This result partially justifies an understanding of w^* as the largest reward-punishment gap that can be sustained in XPE. We say "partially" because it shows that a higher gap cannot be sustained, but it does not show that w^* is attainable; this will follow from Section 4.3.

As a consequence of the preceding analysis, we can return to make good on the promise at the beginning of this section, to rule out some actions where deviation is too tempting:

Lemma 4.5. *In any XPE, at any history h^t ending in a current stage game G^t, the action*

profile played must be in $A^*(G^t, w^*)$.

Proof. It suffices to prove this for date-0 histories. Consider any initial stage game G and any $\epsilon > 0$. Consider any environment E that begins with G followed by the finite sequence of stage games given by Lemma 4.4. For any SPE s for this environment, any action profile a^0 played at date 0 must satisfy

$$(1-\delta)(\widehat{u}(a^0) - u(a^0)) \leq \delta(w^* + \epsilon),$$

by the same logic used in the proof of Lemma 4.4 (and the fact that continuation payoffs of two different SPE's from period 1 onward differ by at most $w^* + \epsilon$).

Therefore, if s is an XPE, then at any date-0 history with any stage game G^0, the action profile to be played must satisfy $\widehat{u}(a^0) - u(a^0) \leq \frac{\delta}{1-\delta}(w^* + \epsilon)$. Since $\epsilon > 0$ is arbitrary, the right side can be replaced by $\frac{\delta}{1-\delta}w^*$, giving the desired result.

□

As a side observation, there may be nontrivial interactions between the different stage games in $\mathcal{G}$ in determining the value of w^*. That is: Suppose that for each $G \in \mathcal{G}$, we define $w^*(G)$ as the highest fixed point of $w \mapsto B(w; G)$. Then, w^* may be bounded strictly below all of the $w^*(G)$. This also implies that the adversarial environments constructed in Lemma 4.4 may need to have the stage game vary from one period to the next.

In fact, we effectively saw this in the opening Example 2.1, with the two stage games G' and G''. Suppose that $\delta = 2/5$, which is in the parameter region where the cooperative outcome was sustainable in either game individually, but not with both games available. It is straightforward to check that the $B(\cdot;\cdot)$ functions are given by

$$B(w; G') = \begin{cases} \frac{2}{5}w & \text{if } w < 3, \\ \frac{12}{5} + \frac{2}{5}w & \text{if } w \geq 3 \end{cases}$$

and

$$B(w; G'') = \begin{cases} \frac{2}{5}w & \text{if } w < \frac{3}{2}, \\ \frac{3}{5} + \frac{2}{5}w & \text{if } \frac{3}{2} \leq w < \frac{9}{2}, \\ 3 + \frac{2}{5}w & \text{if } w \geq \frac{9}{2}. \end{cases}$$

These are plotted as the solid and dashed lines, respectively, in Figure 2 (the two functions coincide for $w < 3/2$ but are shown spaced apart for visibility). The highest fixed points of the two functions are $w^*(G') = 4$ and $w^*(G'') = 5$, as shown in the figure by the intersections of the solid and dashed lines (respectively) with the 45° diagonal. However,

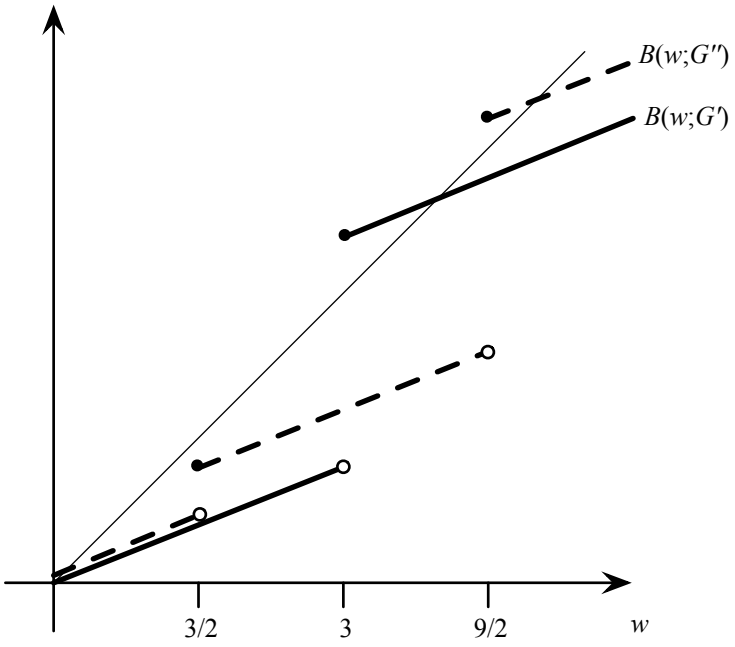

Figure 2: An example where the sustainable reward-punishment gap w^* is lower than the gap sustained by any individual stage game.

w^* is the largest value for which the *lower* of the two functions meets the diagonal, and this happens only at $w^* = 0$.

4.2 Quasi-minmax payoffs

Lemma 4.5 leads to bounds on the payoffs that can arise in any XPE. In particular, for each stage game G, let us pick "most effective reward" and "most effective punishment" action profiles

$$\overline{a}(G) \in \operatorname*{argmax}_{a \in A^*(G,w^*)} u(a); \qquad \underline{a}(G) \in \operatorname*{argmin}_{a \in A^*(G,w^*)} \widehat{u}(a).$$

(As before, these exist, by compactness, and by the fact that $B(w^*) \neq -\infty$ implying $A^*(G, w^*)$ is nonempty.)

The latter give a lower bound on payoffs in XPE. For any environment $E = (G^0, G^1, \ldots)$, define

$$\underline{U}(E) = (1-\delta) \sum_{t=0}^{\infty} \delta^t \widehat{u}(\underline{a}(G^t)).$$

Lemma 4.6. *If s is an XPE, then for any environment E,*

$$U(s|E) \geq \underline{U}(E).$$

Proof. Fix the environment E. Suppose that players $2, \ldots, n$ follow the strategy s, whereas 1 simply plays the myopically optimal deviation at each history. By Lemma 4.5, at each period t, regardless of the past history, s specifies playing an action in $A^*(G^t, w^*)$. Therefore, by myopically deviating, player 1 gets a payoff of at least $\widehat{u}(\underline{a}(G^t))$ in this period. Summing across all periods shows that 1's payoff from the repeated deviation is at least $\underline{U}(E)$. Hence, the payoff from conforming to s is at least this much. □

We can think of $\widehat{u}(\underline{a}(G))$ as a "quasi-minmax" payoff for player 1 when the stage game is G, providing a straightforward lower bound on 1's equilibrium payoffs. Although it involves a minimum over action profiles of 1's myopic best-reply payoff, it differs from the usual minmax in two ways. First, the min is taken only over a restricted set of action profiles, those in $A^*(G, w^*)$. This is natural; because we are not in a folk theorem setting but are considering a fixed δ, some action profiles are ruled out as unsustainable. And second, the action profile that actually produces the stage payoff of $\widehat{u}(\underline{a}(G))$ typically cannot be played in equilibrium, as it does not satisfy the incentive constraints of the short-run players. This is again familiar from the literature on repeated games with short-run players, such as Fudenberg, Kreps and Maskin (1990) (though they nonetheless use the term "minmax value" for the analogous quantity).

Now that we have a lower bound on XPE payoffs, our next step is to develop an XPE strategy profile whose payoffs exceed this bound by a controlled amount.

4.3 Automaton strategies

We present the strategy profile in the form of an automaton, as in (Mailath and Samuelson, 2006, Section 2.3).[3] The automaton enters each period t in some state. After the stage game G^t and public random signal ω^t are realized, the automaton specifies the action profile a^t to be played, and then the automaton transitions to a new state for period $t+1$ depending on the actions observed. In fact, since dynamic incentives are irrelevant for

[3]Section 5.7 of Mailath and Samuelson (2006) develops automata for dynamic games. The formalism used there is in some sense closer to our setting, since actions and transitions depend on the current game state, which is analogous to our stage game G^t. However, in their setup, automaton state transitions happen after the game state in period t is realized and before actions at time t are chosen, whereas for our purposes it is more convenient to have state transitions between periods.

players $2, \ldots, n$, we can focus on state transitions that depend only on 1's action.

More specifically, we consider an automaton whose state space is the interval $W = [0, w^*]$. The state $w \in W$ is to be interpreted as a promise that the payoff will exceed the lower bound $\underline{U}(E)$ by exactly w. The main elements to be specified are the action (output) function $f : W \times \mathcal{G} \times [0,1] \to A$ (which, of course, must output an action profile in $A(G)$ when the input involves stage game G) and the state transition function $\tau : \cup_{G \in \mathcal{G}} (W \times \{G\} \times [0,1] \times A(G)) \to W$. These objects, together with a choice of initial state $w \in W$, determine a strategy profile in the natural way.

For any $G \in \mathcal{G}$, define

$$\lambda(G) = \frac{1}{(1-\delta)(u(\overline{a}(G)) - \widehat{u}(\underline{a}(G))) + \delta w^*}.$$

The denominator equals $B(w^*; G) \geq B(w^*) = w^*$, so for any $w \in W$, we have $\lambda(G)w \in [0,1]$. (The denominator of $\lambda(G)$ may be zero, but only if $w^* = 0$, in which case $w = 0$. In this case, interpret $\lambda(G)w$ as 0 throughout the following.)

Now, for any w, G, ω, a:

- If $\omega \leq \lambda(G)w$: Put $f(w, G, \omega) = \overline{a}(G)$, and

$$\tau(w, G, \omega, a) = \begin{cases} w^* & \text{if } a_1 = \overline{a}_1(G), \\ w^* - \frac{1-\delta}{\delta}(\widehat{u}(\overline{a}(G)) - u(\overline{a}(G))) & \text{otherwise;} \end{cases}$$

- If $\omega > \lambda(G)w$: Put $f(w, G, \omega) = \underline{a}(G)$, and

$$\tau(w, G, \omega, a) = \begin{cases} \frac{1-\delta}{\delta}(\widehat{u}(\underline{a}(G)) - u(\underline{a}(G))) & \text{if } a_1 = \underline{a}_1(G), \\ 0 & \text{otherwise.} \end{cases}$$

In words, we use public randomization to play $\overline{a}(G)$ with probability $\lambda(G)w \in [0,1]$ and play $\underline{a}(G)$ with complementary probability, and then transition to a new state depending on which of the two action profiles was to be played and on whether player 1 deviated. We need to check that all the possible values specified for τ are indeed valid states (i.e. they lie in the interval $[0, w^*]$); this follows from the fact that $\overline{a}(G)$ and $\underline{a}(G)$ are in $A^*(G, w^*)$.

Starting in any state $w \in [0, w^*]$ and proceeding according to the automaton defines a strategy profile. Denote this strategy profile by $s[w]$. The following is a key step in our analysis.

Proposition 4.7. *Pick any $w \in [0, w^*]$, and let E be any environment. Then:*

1. *For each w, $U(s[w]|E) = \underline{U}(E) + w$.*

2. *If the short-run players are following $s[w]$, then at any history h^t, player 1 is indifferent between following $s[w]$ and playing the myopically optimal (one-shot) deviation.*

3. *Strategy profile $s[w]$ is an XPE.*

Proof. 1: Suppose $G = G^0$ is the first stage game encountered in E. By directly considering the possible cases depending on public randomization, and splitting each case into the initial stage and the continuation payoff, we have

$$\begin{aligned} U(s[w]|E) &= \lambda(G)w \times \left((1-\delta)u(\overline{a}(G)) + \delta U(s[w^*]|E^-)\right) + \\ &\quad (1-\lambda(G)w) \times \left((1-\delta)u(\underline{a}(G)) + \delta U\left(s\left[\frac{1-\delta}{\delta}(\widehat{u}(\underline{a}(G)) - u(\underline{a}(G)))\right]\middle| E^-\right)\right). \end{aligned} \tag{4.2}$$

In contrast, write $\widetilde{U}(w|E) = \underline{U}(E) + w$. We will show that $\widetilde{U}$ satisfies the same recurrence:

$$\begin{aligned} \widetilde{U}(w|E) &= \lambda(G)w \times \left((1-\delta)u(\overline{a}(G)) + \delta\widetilde{U}(w^*|E^-)\right) + \\ &\quad (1-\lambda(G)w) \times \left((1-\delta)u(\underline{a}(G)) + \delta\widetilde{U}\left(\frac{1-\delta}{\delta}(\widehat{u}(\underline{a}(G)) - u(\underline{a}(G)))\middle| E^-\right)\right). \end{aligned} \tag{4.3}$$

To see this, expand both the $\widetilde{U}$ terms on the right-hand side of (4.3) and obtain (after slightly simplifying the second line)

$$\begin{aligned} &\lambda(G)w \times \left((1-\delta)u(\overline{a}(G)) + \delta\underline{U}(E^-) + \delta w^*\right) + \\ &\quad (1-\lambda(G)w) \times \left((1-\delta)\widehat{u}(\underline{a}(G)) + \delta\underline{U}(E^-)\right). \end{aligned}$$

Now by combining the terms with the $\lambda(G)w$ coefficient, this rearranges to

$$((1-\delta)\widehat{u}(\underline{a}(G)) + \delta\underline{U}(E^-)) + \lambda(G)w \times ((1-\delta)(u(\overline{a}(G)) - \widehat{u}(\underline{a}(G))) + \delta w^*).$$

But the first parenthesized term is simply $\underline{U}(E)$ from the definition, and the second term is $\lambda(G)w/\lambda(G) = w$, so the whole expression reduces to $\underline{U}(E) + w = \widetilde{U}(w|E)$ as claimed.

Now a standard contraction argument shows that the solution to the recurrence is unique: Write $\Delta(w|E) = U(s[w]|E) - \widetilde{U}(w|E)$. Subtracting (4.3) from (4.2) gives

$$\Delta(w|E) = \lambda(G)w \times \delta\Delta(w^*|E^-) + (1-\lambda(G)w) \times \delta\Delta\left(\frac{1-\delta}{\delta}(\widehat{u}(\underline{a}(G)) - u(\underline{a}(G)))\middle| E^-\right).$$

Put $C = \sup_{w,E} |\Delta(w|E)|$, and note the supremum is finite since both U and $\widetilde{U}$ are bounded. Using C to bound each of the $\Delta(\cdots)$ terms in the previous equation gives

$$|\Delta(w|E)| \leq \lambda(G)w \times \delta C + (1 - \lambda(G)w) \times \delta C = \delta C.$$

Thus, for all w and E, we have $|\Delta(w|E)| \leq \delta C$. In other words, $C \leq \delta C$, which forces $C = 0$. Therefore, $U(s[w]|E) = \widetilde{U}(w|E)$ for all w and E, which completes the proof of part 1.

2: It suffices to prove the statement at period-0 histories. So suppose the date-0 history is $h^0 = (G^0, \omega^0)$. Assume that the automaton specifies an action profile a^0 for which 1's action is not already a myopic best reply (otherwise there is nothing to prove). There are two cases:

- If $\omega^0 \leq \lambda(G^0)w$, then the action profile to be played is $\overline{a}(G^0)$. If player 1 conforms, the state next period is w^*, so the continuation payoff will be $\underline{U}(E^-) + w^*$ by part 1, and therefore the total payoff is

  $$(1-\delta)u(\overline{a}(G^0)) + \delta\left(\underline{U}(E^-) + w^*\right).$$

  If player 1 deviates (optimally) then the stage payoff is $\widehat{u}(\overline{a}(G^0))$ and the state next period is $w^* - \frac{1-\delta}{\delta}\left(\widehat{u}(\overline{a}(G^0)) - u(\overline{a}(G^0))\right)$, so by a similar calculation, the total payoff is

  $$(1-\delta)\widehat{u}(\overline{a}(G^0)) + \delta\left(\underline{U}(E^-) + w^* - \frac{1-\delta}{\delta}\left(\widehat{u}(\overline{a}(G^0)) - u(\overline{a}(G^0))\right)\right)$$
  $$= (1-\delta)u(\overline{a}(G^0)) + \delta\left(\underline{U}(E^-) + w^*\right).$$

- If $\omega^0 > \lambda(G^0)w$, then the action profile to be played is $\underline{a}(G^0)$. Similar calculations show that the total payoff if player 1 conforms is

  $$(1-\delta)u(\underline{a}(G^0)) + \delta\left(\underline{U}(E^-) + \frac{1-\delta}{\delta}\left(\widehat{u}(\underline{a}(G^0)) - u(\underline{a}(G^0))\right)\right)$$
  $$= (1-\delta)\widehat{u}(\underline{a}(G^0)) + \delta\underline{U}(E^-)$$

  and if player 1 deviates is

  $$(1-\delta)\widehat{u}(\underline{a}(G^0)) + \delta\underline{U}(E^-).$$

So in each case, the payoffs from conforming and deviating are equal.

3: We have just shown that in every environment and at every history, player 1 is indifferent to the myopically optimal one-shot deviation. Playing a non-optimal deviation cannot do better, since it leads to the same next-period state (and so the same continuation payoff) as the optimal deviation while giving a lower stage payoff. (Note that if the action profile a specified is such that 1's action is already a best reply, then $\widehat{u}(a) = u(a)$, so by inspection of the formulas, the next-period state after a deviation is the same as after conforming, and the same argument applies.) So, player 1 cannot benefit from a one-shot deviation of any sort, and 1's incentive constraint is satisfied.

The other players' incentives are also satisfied, since whenever a stage game G is to be played, the automaton specifies an action profile in $A^*(G, w^*) \subseteq A^*(G)$. So we have an XPE.

□

With this result, we are now justified in thinking of w^* as the largest sustainable reward-punishment gap (as mentioned in Section 4.1), since we do indeed have two XPE's—namely, $s[w^*]$ and $s[0]$—whose payoffs differ by w^* in every environment.

5 Main results

With this machinery in hand, we are ready to take up our main question of interest: what outcomes might arise in equilibrium?

5.1 Defining outcomes

A first question is how outcomes should be defined, in this setting without a prior over environments. One option is to take the perspective that there exists a "true" (but initially unknown) environment; the outcome should then consist of this realized environment, together with the action profiles played in each period. Of course, the latter may be random, depending on the public signals. Accordingly, we define a *realizable outcome* as a pair (E, z), where $E = (G^0, G^1, \ldots)$ is an environment, and $z : \cup_{t=0}^{\infty}[0,1]^{t+1} \to A$, specifying an action profile $z(\omega^{0,\ldots,t}) \in A(G^t)$ for each date and history of public signals.[4]

[4]Alternatively, we could simply define a realizable outcome as consisting of an environment and a joint distribution over $(a^0, a^1, \ldots)$. This description would contain less information, since it does not specify which periods' public random signals are involved in determining a^t for any given t. (For example, suppose the a^t are independent. We could have each a^t determined by ω^t, thus unpredictable until time t; or could have all of $(a^0, a^1, \ldots)$ determined by the date-0 signal ω^0; or anywhere in between.)

In the special case where the time-t action is independent of the random signals for each t, we call the outcome *deterministic*; such an outcome can simply be described by a single stage game G^t and action profile $a^t \in A(G^t)$ in each period. The reader may find it useful to focus on deterministic outcomes for concreteness, but we will state results for the general case since they are not much more involved.

An alternative perspective is to view an outcome as a full description of the actions that may be played "on-path," for whatever environment may be realized. Accordingly, define a *full outcome* to be a function $z : \cup_{t=0}^{\infty} (\mathcal{G} \times [0,1])^{t+1} \to A$, specifying an action profile $z(G^0, \omega^0, \ldots, G^t, \omega^t) \in A(G^t)$ for each possible initial sequence of stage games and public signals. We again say that such an outcome is *deterministic* if actions are always independent of the signals. A realizable outcome (E, z'), where $E = (G^0, G^1, \ldots)$, *belongs to* the full outcome z if $z(G^0, \omega^0, \ldots, G^t, \omega^t) = z'(\omega^{0,\ldots,t})$ for all t and $\omega^{0,\ldots,t}$.

We will mostly focus on realizable outcomes for expository simplicity. Section 5.5 will state the corresponding results for full outcomes.

Let us say that a strategy profile s *supports* the realizable outcome given by (E, z) if, for all t and $\omega^{0,\ldots,t}$, if we define $a^{t'} = z(\omega^{0,\ldots,t'})$ for each $t' \leq t$ then s satisfies

$$s(G^0, \omega^0, a^0; G^1, \omega^1, a^1; \ldots, G^t, \omega^t) = a^t.$$

Thus, in the environment E, actions on-path are chosen as specified by z. We similarly say that s *supports* a full outcome z if it supports every realizable outcome that belongs to z.

Note that even if the outcome itself is deterministic, randomization off-path may be needed to support it in equilibrium.

5.2 Supportable outcomes

It is not hard to see that the following are necessary conditions for a realizable outcome $(G^0, G^1, \ldots; z)$ to be supported by an XPE s:

$$z(\omega^{0,\ldots,t}) \in A^*(G^t, w^*) \qquad \text{for all } t \text{ and } \omega^{0,\ldots,t}; \tag{5.1}$$

$$\left(\widehat{u}(a^{\underline{t}}) - u(a^{\underline{t}})\right) + \sum_{t=\underline{t}+1}^{\bar{t}} \delta^{t-\underline{t}} \left(\widehat{u}(\underline{a}(G^t)) - \mathbb{E}^{\underline{t}}[u(a^t)]\right) \leq \frac{\delta^{\bar{t}+1-\underline{t}}}{1-\delta} w^* \tag{5.2}$$

$$\text{for all } \underline{t} < \bar{t} \text{ and } \omega^{0,\ldots,\underline{t}},$$

where $a^{\underline{t}} = z(\omega^{0,\cdots,\underline{t}})$, and likewise a^t for $t > \underline{t}$.

(Note that an equivalent formulation is simply that $a^t \in A^*(G^t)$ for all t and (5.2) holds for all $\underline{t} \leq \bar{t}$, where the sum is empty if $\underline{t} = \bar{t}$.)

Indeed, we have already seen that (5.1) is necessary. For (5.2), consider any $\epsilon > 0$, and consider the environment E' that consists of $(G^0, \ldots, G^{\bar{t}})$, followed by the sequence of stage games identified in Lemma 4.4 for this ϵ (and any other stage games thereafter). Consider player 1's decision at time $\underline{t}$, with history $h^{\underline{t}}$. Conforming to s gives a payoff

$$U(s|E', h^{\underline{t}}) = (1-\delta)\left(\sum_{t=\underline{t}}^{\bar{t}} \delta^{t-\underline{t}} \mathbb{E}^{\underline{t}}[u(a^t)]\right) + \delta^{\bar{t}+1-\underline{t}} \mathbb{E}^{\underline{t}}[U(s|E', h^{\bar{t}}, a^{\bar{t}})]. \tag{5.3}$$

(Here, $h^{\bar{t}}$ represents the history arising at period $\bar{t}$.)

An alternative strategy s_1' would play a myopic best reply to the short-run players' anticipated actions at each period $t = \underline{t}, \ldots, \bar{t}$, and then follow s_1 from date $\bar{t}+1$ onward. This would give a stage payoff of $\widehat{u}(a^{\underline{t}})$ in period $\underline{t}$, and would guarantee at least $\widehat{u}(\underline{a}(G^t))$ in each period $t = \underline{t}+1, \ldots, \bar{t}$. So player 1's deviation payoff satisfies

$$U(s_1', s_{-1}|E', h^{\underline{t}}) \geq (1-\delta)\left(\widehat{u}(a^{\underline{t}}) + \sum_{t=\underline{t}+1}^{\bar{t}} \delta^{t-\underline{t}} \widehat{u}(\underline{a}(G^t))\right) + \delta^{\bar{t}+1-\underline{t}} \mathbb{E}^{\underline{t}}[U(s|E', \widetilde{h}^{\bar{t}}, \widetilde{a}^{\bar{t}})] \tag{5.4}$$

(where $\widetilde{h}^{\bar{t}}$ and $\widetilde{a}^{\bar{t}}$ denote the history and period-$\bar{t}$ actions produced by 1's deviations). Since the deviation should not be profitable, subtracting (5.3) from (5.4) and dividing by $1-\delta$ gives

$$\begin{aligned}\left(\widehat{u}(a^{\underline{t}}) - u(a^{\underline{t}})\right) + \sum_{t=\underline{t}+1}^{\bar{t}} \delta^{t-\underline{t}} \left(\widehat{u}(\underline{a}(G^t)) - \mathbb{E}^{\underline{t}}[u(a^t)]\right) + \\ \frac{\delta^{\bar{t}+1-\underline{t}}}{1-\delta}\left(\mathbb{E}^{\underline{t}}[U(s|E', \widetilde{h}^{\bar{t}}, \widetilde{a}^{\bar{t}})] - \mathbb{E}^{\underline{t}}[U(s|E', h^{\bar{t}}, a^{\bar{t}})]\right) \leq 0.\end{aligned}$$

However, the two $U(s|\cdots)$ terms both represent SPE payoffs in the environment starting at date $\bar{t}+1$, and so by Lemma 4.4, they differ by less than $w^* + \epsilon$. Applying this bound, rearranging, and taking $\epsilon \to 0$ gives (5.2).

Condition (5.2) essentially says that the payoff gains from repeated myopic deviation across any interval of periods must be bounded by w^* (suitably discounted). Notice that the terms $\widehat{u}(\underline{a}(G^t)) - \mathbb{E}^{\underline{t}}[u(a^t)]$ may be positive or negative, so it is unknown a priori for which pairs $(\underline{t}, \bar{t})$ the constraint will be tightest.

Our first main result is that conditions (5.1)–(5.2) actually give a complete characterization of the realizable outcomes that can be supported in XPE. At first, this may be surprising in light of the derivation of (5.2) above, which connects it specifically to repeated deviations that continue until the realized environment differs from that in the target outcome. Why does it also suffice to rule out other kinds of deviations? An intuition comes from the indifference result of Proposition 4.7: If deviations are optimally punished using the automaton strategies $s[0]$, then the payoff from deviating repeatedly is the same as from deviating once, so the condition suffices to rule out one-shot deviations (and therefore all others).

Theorem 5.1. *A realizable outcome (E, z) is supported by some XPE s if and only if it satisfies the necessary and sufficient conditions (5.1)–(5.2).*

Proof. Necessity was just argued, so we prove sufficiency. Construct a strategy profile s as follows:

- At any history h^t such that all stage games $(G^0, \ldots, G^t)$ so far have been consistent with E and all actions so far $(a^0, \ldots, a^{t-1})$ have been as prescribed by z, play as specified by z.
- For any history h^t where the stage games and action profiles through time $t-1$ were all as specified by (E, z), but the period-t stage game is different, play according to $s[w^*]$ from h^t onward.
- For any history h^t where all past stage games through time $t-1$ and all action profiles through time $t-2$ were as specified by (E, z), but the action profile observed at $t-1$ was different from that indicated by z, play according to $s[0]$ from period t onward.

Notice that at every history, either all stage games and action profiles so far agreed with (E, z), or there was a unique earliest stage game or action profile that did not agree with (E, z), so this description does specify a well-defined strategy profile. By construction it supports (E, z); we need to check that it is an XPE.

At any history where any stage game or past action has differed from (E, z), there is no incentive to deviate; this follows because we already know that $s[w^*]$ and $s[0]$ are XPE's. Moreover, the incentives of the short-run players are automatically satisfied since an action profile in $A^*(G)$ is indicated at every history. So we only need to check the incentives of player 1 to deviate at histories h^t that have so far agreed with (E, z).

Consider such a history h^t, and any environment $\widetilde{E}$ consistent with it. Suppose $\widetilde{E} \neq E$. Let $\bar{t}+1$ be the earliest period in which $\widetilde{E}$ and E differ. So, writing $E = (G^0, G^1, \ldots)$ as usual, then $\widetilde{E}$ begins $(G^0, G^1, \ldots, G^{\bar{t}}, \widetilde{G}^{\bar{t}+1}, \ldots)$. Evidently $\bar{t} \geq t$.

If $\bar{t} = t$, then by conforming when asked to play a^t, player 1 achieves a payoff (from the period-t vantage point) of $(1-\delta)u(a^t) + \delta(\underline{U}(\widetilde{E}^{-(\bar{t}+1)}) + w^*)$, since play transitions to $s[w^*]$ next period. By deviating, player 1's payoff is $(1-\delta)\widehat{u}(a^t) + \delta\underline{U}(\widetilde{E}^{-(\bar{t}+1)})$ (or less, if a non-optimal deviation is chosen). So the overall gain from deviating is $(1-\delta)(\widehat{u}(a^t) - u(a^t)) - \delta w^*$, which is ≤ 0 by condition (5.1).

If $\bar{t} > t$, then by conforming, player 1 achieves a payoff of

$$(1-\delta)\left(\sum_{t'=t}^{\bar{t}} \delta^{t'-t}\mathbb{E}^t[u(a^{t'})]\right) + \delta^{\bar{t}+1-t}\left(\underline{U}(\widetilde{E}^{-(\bar{t}+1)}) + w^*\right).$$

By deviating, player 1's payoff is

$$(1-\delta)\widehat{u}(a^t) + \delta\underline{U}(\widetilde{E}^{-(t+1)}).$$

By expanding using the definition of $\underline{U}$, we get $\underline{U}(\widetilde{E}^{-(t+1)}) = (1-\delta)\left(\sum_{t'=t+1}^{\bar{t}} \delta^{t'-(t+1)}\widehat{u}(\underline{a}(G^{t'}))\right) + \delta^{\bar{t}-t}\underline{U}(\widetilde{E}^{-(\bar{t}+1)})$, and so the deviation payoff is

$$(1-\delta)\left(\widehat{u}(a^t) + \sum_{t'=t+1}^{\bar{t}} \delta^{t'-t}\widehat{u}(\underline{a}(G^{t'}))\right) + \delta^{\bar{t}+1-t}\underline{U}(\widetilde{E}^{-(\bar{t}+1)}).$$

Now condition (5.2) (with t in place of $\underline{t}$) implies that the deviation is unprofitable.

One loose end remains: what about play in the exact environment E specified by the target outcome? In this case, for each $t' > t$, let $\widetilde{E}^{t'}$ be an alternative environment that agrees with E until period t' and disagrees with it starting at $t'+1$. History h^t is then consistent with $\widetilde{E}^{t'}$. Taking limits as $t' \to \infty$, we have $U(s|\widetilde{E}^{t'}, h^t) \to U(s|E, h^t)$ and, for any proposed deviating strategy s_1', $U(s_1', s_{-1}|\widetilde{E}^{t'}, h^t) \to U(s_1', s_{-1}|E, h^t)$. So the fact that the deviation is not profitable in any $\widetilde{E}^{t'}$ (which we have already shown) implies, by taking limits, that it is not profitable in E either.

□

A few remarks are in order.

First, as a special case when $\mathcal{G} = \{G\}$ is a singleton, we can cover the case of a standard repeated game with a single long-run player; our analysis so far identifies the

long-run player's worst SPE payoff (which reduces simply to $\widehat{u}(\underline{a}(G))$) and characterizes the supportable outcomes. This does not seem to be noted in existing literature.

Second, we can compare the conditions for an XPE outcome against those for an SPE outcome in standard repeated games. By taking the limit as $\bar{t} \to \infty$ in (5.2), we get

$$\left(\widehat{u}(a^{\underline{t}}) - u(a^{\underline{t}})\right) + \sum_{t=\underline{t}+1}^{\infty} \delta^{t-\underline{t}} \left(\widehat{u}(\underline{a}(G^t)) - \mathbb{E}^{\underline{t}}[u(a^t)]\right) \leq 0 \qquad \text{for all } \underline{t}, \omega^{0,\ldots,\underline{t}}. \tag{5.5}$$

This condition says that the payoff from following the proposed outcome, beginning in period $\underline{t}$, is at least as high as that from a one-period deviation followed by the ensuing punishment. In repeated games, the corresponding condition is in fact sufficient for supportability in SPE (see Abreu, 1988, Proposition 4). Here, we need a condition indexed both by $\underline{t}$ and $\bar{t}$ because of the possibility of different stage games arising in future periods. That is, the proposed realizable outcome may satisfy (5.5) if there is a large temptation to deviate at period $\underline{t}$ but large rewards promised at some future period. Such an outcome may not be supportable because, when the future comes along, the stage game may be one in which large rewards are impossible, and thus the deviation at $\underline{t}$ cannot be discouraged.

Third, at least in the *deterministic* case, we can slightly rewrite the conditions in a way that offers an alternative interpretation. Given a deterministic realizable outcome (E, z), recursively define $d^{-1}(z) = 0$ and

$$d^t(z) = \max\left\{\frac{1}{\delta} d^{t-1}(z) + \left(\widehat{u}(\underline{a}(G^t)) - u(a^t)\right), \widehat{u}(a^t) - u(a^t)\right\}$$

for $t = 0, 1, \ldots$. Then we have (proof in Appendix A):

Proposition 5.2. *A deterministic realizable outcome (E, z), with $E = (G^0, G^1, \ldots)$ and $z = (a^0, a^1, \ldots)$, is supported by some XPE if and only if it satisfies $a^t \in A^*(G^t)$ for all t, and $d^t(z) \leq \frac{\delta}{1-\delta} w^*$ for all t.*

We can think of d^t as the "debt" owed to player 1 after period t for refraining from deviation in the past. The proposition then says that an outcome can be supported in XPE just so long as the debt owed never exceeds the amount that can be promised. In each period t, the debt repayment promised in the future needs to be large enough to cover the previous debt, with "interest," adjusted by whatever portion is being delivered in the present period (this is the first term of the max); it also needs to be large enough to outweigh the gains from a one-time deviation at t.

Fourth, we have so far been viewing the set of possible stage games as fixed and asking what outcomes are supportable. But we could equally well flip things around and ask: given a proposed outcome, what sets of stage games allow it to be supported? This question might be of interest, for example, to a long-run player who is confident about the environment and has a desired outcome in mind, but who worries that the short-run players are more uncertain about the environment, and who wants to know what the short-run players need to know in order for them to be assured that the long-run player is willing to follow the plan.

More formally, let $\overline{\mathcal{G}}$ be some "universe" of potential stage games, and $u_i : \overline{A} \to \mathbb{R}$ the corresponding payoff functions, satisfying the assumptions of Section 3 (where $\overline{A}$ is the disjoint union of the sets $A(G)$ for $G \in \overline{\mathcal{G}}$). Let (E, z) be a realizable outcome and $\underline{\mathcal{G}} \subseteq \overline{\mathcal{G}}$ such that each stage game G^t of E lies in $\underline{\mathcal{G}}$. We consider various sets $\mathcal{G}$ with $\underline{\mathcal{G}} \subseteq \mathcal{G} \subseteq \overline{\mathcal{G}}$; any choice of such a $\mathcal{G}$ specifies the possible stage games in such a way that the environment E can occur. Under what conditions on $\mathcal{G}$ will it be the case that (E, z) is supportable in XPE over $\mathcal{G}$? For the deterministic case, Proposition 5.2 gives an answer: this happens if and only if the value of w^* for $\mathcal{G}$ is greater than or equal to $\frac{1-\delta}{\delta} \sup_t d^t(z)$; equivalently, if and only if there exists some $w \geq \frac{1-\delta}{\delta} \sup_t d^t(z)$ such that $B(w; G) \geq w$ for each $G \in \mathcal{G}$. (And likewise, for the more general case, this condition must hold for some w large enough to satisfy (5.1)–(5.2) along all signal histories.) As a side note, observe that this condition is not closed under taking unions: that is, it may be that a set of stage games $\mathcal{G}$ supports z as an XPE outcome, and another set $\mathcal{G}'$ does also, but their union $\mathcal{G} \cup \mathcal{G}'$ does not, because the value of w that works for $\mathcal{G}$ is different than the one that works for $\mathcal{G}'$. Indeed, we saw this in Example 2.1.

5.3 Universal penal codes

Another central result from standard repeated games that does carry over to our setting is the existence of "worst punishments" that can be used to support any equilibrium outcome path. Explicitly, let us say that a strategy profile $\underline{s}$ is a *universal penal code* if it has the following property: For every realizable outcome (E, z) that is supportable in XPE, there is in particular an XPE supporting z where, following any initial deviation by player 1 (i.e. a history $h^t = (G^0, \omega^0, a^0; \ldots; G^t, \omega^t)$ consistent with E, where all actions so far are as specified by z, followed by an action profile (a'_1, a^t_{-1}) where 1's action differs from that given by z), continuation play is given by $\underline{s}$. We then have the following result:

Theorem 5.3. *There exists a universal penal code.*

Proof. It follows from the proof of Theorem 5.1 that $s[0]$ is a universal penal code, since, for any realizable outcome meeting conditions (5.1)–(5.2), that proof constructs an XPE supporting it with player 1's deviations punished by $s[0]$.

□

(However, unlike the repeated-game setting, here the statement that z should be played with deviations punished by $\underline{s}$ does not give a full description of the strategy profile, since it does not specify what happens once a stage game differing from E gets realized.)

As a brief note on literature, Abreu (1988) is usually credited for the notion of penal codes. The relevant definition there is that of an *optimal penal code*, which is a specification of an SPE for each player that delivers to that player the lowest payoff among all SPE's. Although it is also true here that our $s[0]$ is an optimal penal code, in the strong sense of delivering the lowest XPE payoff in *any* environment, this definition does not explicitly relate to its use as a punishment, which is why we have instead emphasized the definition of universal penal codes here. In general, the notions of an optimal penal code and a universal penal code need not coincide.

5.4 Comparing XPE to SPE outcomes

As mentioned in the introduction, a difficulty with giving a positive interpretation to XPE is that it is not rooted in individual maximization. One might instead argue that agents should play an SPE of the dynamic game induced by whatever process (perhaps random) they believe governs the stage games. (And even if the agents are unsure about this process, and there is asymmetric information about it, they should presumably play an equilibrium of the the resulting incomplete information game.)

Of course, any XPE is automatically an equilibrium of any such fully-specified game as well, and so an XPE-supported outcome is one that the analyst can confidently describe as being attainable in whatever world the players actually live in. However, a natural converse statement is not true. As shown by Example 5.1 below, there can be realizable outcomes that can be supported in SPE no matter what process governs the stage games, but that require different punishments for deviation depending on the process, and so cannot be supported in XPE.[5]

[5]A parallel is the question of foundations for dominant-strategy implementation in mechanism design (Bergemann and Morris, 2005; Chung and Ely, 2007). Dominant-strategy mechanisms, when they exist, allow for a desired outcome to be achieved regardless of agents' beliefs or higher-order beliefs about each other. This robustness has led to a large literature focusing on such mechanisms. But even when they do not exist, it may be still be possible to implement the desired outcome with a mechanism where agents'

Before giving this example, we sketch the concepts in a little more detail. For simplicity, here and for the rest of this subsection, we assume A is finite, i.e. the set of stage games and the action spaces are all finite.

A *stage game process* consists of a specification of $\pi_{(G^0,\ldots,G^t)} \in \Delta(\mathcal{G})$ for each initial history of stage games $(G^0, \ldots, G^t)$, describing the distribution over G^{t+1} given the previous realizations. We denote such a process by π.[6] Histories and strategies are defined exactly as in the main model. At any history of stage games, the transition probabilities given by π recursively determine a conditional distribution over $(G^{t+1}, G^{t+2}, \ldots)$, which allows us to define the expected payoffs from a strategy profile at any history (with the understanding that the public random signals $(\omega^0, \omega^1, \ldots)$ are drawn independently of the stage game transitions). Strategy profile s is an SPE for π if, at each history, no player can improve his expected payoff by deviating.

The definitions of realizable (and full) outcomes, and strategies supporting such outcomes, are unchanged. Then, the statement that s supports (E, z) can be interpreted as saying that z describes the actions played conditional on E realizing.

Evidently, any strategy profile that is an XPE is an SPE for any stage game process: since deviating can never increase the payoff in any environment, it cannot increase the payoff in expectation either. A fortiori, any XPE-supportable outcome is SPE-supportable for any stage game process. Below is the example showing that the converse is not true. A rough intuition is that the argument for necessity of (5.2), applied in an SPE setting, would require two things: first, that at time $\underline{t}$, the subsequent stage games $G^{\underline{t}+1}, \ldots, G^{\bar{t}}$ in the target outcome are expected to arise with high probability (otherwise (5.2) is not relevant to incentives from deviation); and second, at each intervening time t, the environment starting at time $t+1$ is likely to be the adversarial one identified in Lemma 4.4 (because otherwise actions outside of $A^*(G^t, w^*)$ can be played, so $\underline{a}(G^t)$ is not the worst available punishment). A single stage game process cannot simultaneously satisfy both conditions.

Example 5.1. Consider two possible stage games, G and G', as shown in Figure 3. Part (a) of the figure illustrates them in the standard matrix form, whereas part (b) rearranges them to a form more suitable for us, by showing the action profiles in $A^*(G)$ and $A^*(G')$, and the player-1 payoffs u and $\widehat{u}$ for each.

strategies depend on their beliefs.

[6] Alternatively, we could define a stage game process directly as a distribution over environments E, but then we would need to add a full-support assumption to avoid the difficulty of defining expectations about the future stage games at probability-zero histories.

G :

	a	b	c	d
a	$24,1$	$0,0$	$0,0$	$-22,0$
b	$40,0$	$8,1$	$0,0$	$-40,0$
c	$0,0$	$15,0$	$0,1$	$-40,0$
d	$0,0$	$0,0$	$0,0$	$-40,1$

G' :

	e	f
e	$16,1$	$0,0$
f	$16,0$	$0,1$

(a)

G :

	aa	bb	cc	dd
u	24	8	0	-40
$\widehat{u}$	40	15	0	-22

G' :

	ee	ff
u	16	0
$\widehat{u}$	16	0

(b)

Figure 3: Example with a realizable outcome that is supportable in SPE for any stage game process, but not supportable in XPE.

We take the discount factor $\delta = 1/2$. This leads to $w^* = 16$, $A^*(G, w^*) = \{aa, bb, cc\}$ and $A^*(G', w^*) = A^*(G') = \{ee, ff\}$. In particular, $\widehat{u}(\underline{a}(G)) = \widehat{u}(\underline{a}(G')) = 0$.

Consider the deterministic realizable outcome in which G is played every period, and the action profiles are $(bb, cc, bb, aa, aa, aa, aa, \ldots)$. This outcome does not satisfy (5.2) with $\underline{t} = 0$, $\overline{t} = 2$, so it cannot be supported in XPE. However, we claim that it can always be supported in SPE for any full-support stage game process π. To see this, write q for the probability of $G^2 = G$ given that $(G^0, G^1) = (G, G)$, and consider two cases:

Case 1: $q \leq 1/2$.

Consider the following strategy profile. For the "on-path" actions (as long as there have been no deviations), as long as G has arisen in every period, play according to the target outcome; once G' realizes, play ee, and then play either aa or ee in every subsequent period. If there is ever a deviation by player 1, play the punishment actions cc or ff in every subsequent period. (Further deviations can be ignored.)

Let us check that there is never an incentive to deviate. During the punishment phase, there is no gain from deviating. During the on-path phase, if G' has ever arisen, or if only G has ever arisen and the current period is $t \geq 2$, then the deviation brings a short-run gain of at most 16 but a loss of at least 16 in each subsequent period, so is not optimal. If only G has ever arisen and $t = 1$, then there is no myopic gain, only a subsequent loss.

This leaves only the case $t = 0$, when playing G in the initial period. The myopic gain is 7. We consider two possibilities:

- Conditional on $G^1 = G'$, the deviation in period 0 leads to a loss of at least 16 in each subsequent period, so a loss overall.

- Conditional on $G^1 = G$, the punishment entails no loss in period 1, but it entails a loss in period 2 of either 8 or 16 depending on whether $G^2 = G$ or G', and then a loss of at least 16 in every subsequent period. Hence, the total net gain, in period-0 payoff terms, is at most

$$(1-\delta)[7-\delta^2\cdot(q\cdot 8+(1-q)\cdot 16)-(\delta^3+\delta^4+\cdots)\cdot 16] = \left(\frac{1}{2}\right)[7-(4-2q)-4] \leq 0.$$

Case 2: $q \geq 1/2$.

In this case, we first consider the following strategy profile, call it s_d, for the game from period 1 onwards: If $G^1 = G$, we play dd in period 1, and then on-path play aa or ee in all subsequent periods. If there is ever a deviation, punish using cc or ff in all subsequent periods (and ignore further deviations). If $G^1 = G'$, play ff in period 1, and then cc or ff in all subsequent periods (and ignore deviations).

We claim that s_d is an SPE of the subgame starting in period 1 conditional on $G^0 = G$. There is no incentive to deviate whenever cc, ee, or ff is specified. When aa is indicated, deviating brings a short-run gain of at most 16 and a loss at least 16 in each subsequent period. This leaves us only to check the incentive to deviate from d in period 1 when $G^1 = G$. This deviation brings an immediate gain of 18, and a loss of at least 16 in each subsequent period, including a loss of 24 in period 2 if $G^2 = G$ (which happens with probability q), hence an overall net gain at most

$$(1-\delta)[18-\delta\cdot(q\cdot 24+(1-q)\cdot 16)-(\delta^2+\delta^3+\cdots)\cdot 16] = \left(\frac{1}{2}\right)[18-(8+4q)-8] \leq 0.$$

With this in mind, we consider the following strategy profile for the overall game. On-path actions are as in Case 1. A deviation in period 0, if $G^0 = G$, is punished by switching to s_d in subsequent periods. Any other deviation is punished by playing cc or ff in all subsequent periods (and further deviations are ignored).

As in Case 1, it is easy to check there is no incentive to deviate at all histories except at the initial period when playing G. For this last, the short-run gain from deviating is 7. Conditional on $G^1 = G'$, the loss in every period from 1 onward is at least 16, so the deviation is not beneficial. And conditional on $G^1 = G$, the loss in period 1 is 40 and there is no further gain except possibly of 16 in period 2 (if $G^2 = G$), so the net effect is

at best $(1-\delta)[7+\delta\cdot(-40)+\delta^2\cdot 16]<0$.

$\triangle$

The lesson of this example, however, rests on the assumption that the long-run player maximizes expected utility with respect to some belief about the future stage games. One might instead imagine—and it is arguably in keeping with the spirit of our overall exercise—that the long-run player evaluates the uncertainty over the future with some non-expected utility; for example, he might be ambiguity averse.

We therefore proceed to consider a large class of (weakly) ambiguity-averse preferences that mesh with the discounting structure of repeated games: namely, *dynamic variational preferences* (Maccheroni, Marinacci and Rustichini, 2006). Such preferences, adapted for our setting, are parameterized by a *dynamic ambiguity index* c, which specifies, for each $t\geq 0$ and each initial history of stage games $(G^0,\ldots,G^t)$, a function $c_{(G^0,\ldots,G^t)}:\Delta(\mathcal{G}^\infty)\to\mathbb{R}\cup\{\infty\}$ that is convex and is not everywhere infinite.

Given a dynamic ambiguity index c, at any history $h^t=(G^0,\omega^0,a^0;\ldots;G^t,\omega^t)$, we define the subgame payoff for a strategy profile s by

$$U(s|c,h^t)=(1-\delta)\inf_{\psi\in\Delta(\mathcal{G}^\infty)}\left(\mathbb{E}_\psi\left[\sum_{t'=t}^{\infty}\delta^{t'-t}u(a^{t'})\right]+c_{(G^0,\ldots,G^t)}(\psi)\right). \tag{5.6}$$

Here, the expectation is with respect to future stage games $(G^{t+1},G^{t+2},\ldots)$ drawn from distribution ψ and signals $(\omega^{t+1},\omega^{t+2},\ldots)$ drawn independently $U[0,1]$, and $a^{t'}$ are the actions played by following s starting at h^t, as usual.

Note that expected utility with respect to a particular stage game process π is a special case, where we simply take $c_{(G^0,\ldots,G^t)}(\psi)$ to be 0 if ψ coincides with the distribution over future stage games generated by π after $(G^0,\ldots,G^t)$, and ∞ for any other ψ. Other commonly-studied special cases include maxmin utility with multiple priors (Epstein and Schneider, 2003) and multiplier preferences (Hansen and Sargent, 2001).

We then say that s is an SPE for c if, for every history h^t and any possible deviation s_1', $U(s|c,h^t)\geq U(s_1',s_{-1}|c,h^t)$, and the short-run players' incentives are always satisfied. Let us also define a *one-shot SPE* for c by the same conditions except that we require $U(s|c,h^t)\geq U(s_1',s_{-1}|c,h^t)$ only for each s_1' that differs from s_1 only at h^t.

Preferences (5.6) are not dynamically consistent in general, and therefore the one-shot deviation principle need not apply: a one-shot SPE may not be an SPE.[7] However, it

[7]Maccheroni, Marinacci and Rustichini (2006) also identify a subclass of dynamic variational preferences that are dynamically consistent. However, adapting this feature to our setting would require

remains the case that if s is an XPE then it is also an SPE for any such preferences (even allowing repeated deviations). This follows since a deviation from s_1 to any alternate strategy s_1' can never increase the expression inside the infimum for any particular ψ, and therefore cannot increase the value of the infimum.

With this broader class of preferences, we restore the desired "converse" result relating XPE-supportable outcomes to SPE-supportable ones, and moreover it holds even if SPE is relaxed to one-shot SPE:

Theorem 5.4. *If a realizable outcome $(E_\bullet, z)$ is not supported by any XPE, then there exists a dynamic ambiguity index c such that $(E_\bullet, z)$ is not supported by any one-shot SPE for c.*

Proof. Write $E_\bullet = (G^0_\bullet, G^1_\bullet, \ldots)$. Let $w_0, w_1, \ldots$ be the sequence from Lemma 4.3.

By Theorem 5.1, z must violate either (5.1) or (5.2). The former case is easy to dispose of: In this case, there exist some t and $\omega^{0,\ldots,t}$ such that $a^t = z(\omega^{0,\ldots,t})$ satisfies $\widehat{u}(a^t) - u(a^t) > \frac{\delta}{1-\delta} w_k$ for some k. (Or one of the short-run players' incentives is violated, but then our conclusion is immediate.) Lemma 4.4 gives an environment $\widetilde{E} = (\widetilde{G}^0, \widetilde{G}^1, \ldots)$ in which any two SPE payoffs differ by less than w_k. Consider the environment $E = (G^0_\bullet, G^1_\bullet, \ldots, G^t_\bullet, \widetilde{G}^0, \widetilde{G}^1, \widetilde{G}^2, \ldots)$. The proof of Lemma 4.4 shows that, in any SPE for this environment, a^t can never be played at time t. So our conclusion follows, with c actually given by expected utility for the (degenerate) stage game process that always follows E.

This leaves us with the case where (5.2) is violated for some $\underline{t} < \bar{t}$ and $\omega^{0,\ldots,\underline{t}}$. Again, (5.2) will remain violated if its right side is replaced by $\frac{\delta^{\bar{t}+1-\underline{t}}}{1-\delta} w_k$ for large enough k. Also, our finiteness assumption implies that for all $G \in \mathcal{G}$ we have $A^*(G, w_k) = A^*(G, w^*)$ for k large enough, so assume this holds as well. As above, let $\widetilde{E} = (\widetilde{G}^0, \widetilde{G}^1, \ldots)$ be the environment given by Lemma 4.4 for w_k. Let $\underline{\widetilde{U}}$ be the infimum of payoffs of SPE's for $\widetilde{E}$, so by the lemma, any SPE for $\widetilde{E}$ has payoff at most $\underline{\widetilde{U}} + w_k$. Also, write $\overline{G}$ for the stage game that was $\overline{G}_{k+1}$ in the proof of Lemma 4.4, so that $B(w_k; \overline{G}) < w_{k+1} < w_k$.

We construct the ambiguity index c as follows:

- For each $t > \bar{t}$ and any $(G^0, \ldots, G^t)$, let $c_{(G^0,\ldots,G^t)}$ be the function that assigns value 0 to ψ if ψ places probability 1 on the future stage games $(G^{t+1}, G^{t+2}, G^{t+3}, \ldots)$ being equal to $(\widetilde{G}^{t-\bar{t}}, \widetilde{G}^{t-\bar{t}+1}, \widetilde{G}^{t-\bar{t}+2}, \ldots)$, and assigns ∞ to any other ψ.

- For each $t \leq \bar{t}$ and any $(G^0, \ldots, G^t)$, let $c_{(G^0,\ldots,G^t)}$ be the function that assigns ∞ to ψ if ψ places positive probability on $G^{t'} \neq \widetilde{G}^{t'-\bar{t}-1}$ for some $t' > \bar{t}$, and otherwise

allowing a broader space of ψ's in which future stage games may be correlated with future random signals; we would then lose the result that every XPE is always an SPE.

assigns ψ a value equal to $-\mathbb{E}_\psi\left[\sum_{t'=t}^{\bar{t}} \delta^{t'-t}\widehat{u}(\underline{a}(G^{t'}))\right]$. (Note that this sum includes a term for G^t for which is already determined by the history, as well as terms for future stage games drawn from ψ.)

This function is indeed convex, since it is finite-valued only for a convex set of ψ's and is affine on this set.

The affineness for $t \leq \bar{t}$ means that the infimum in (5.6) is attained at a corner of the set of possible ψ's, which allows us to simplify (5.6) as follows. Given history h^t, say that an environment $E = (G^0, G^1, \ldots)$ is *valid* for h^t if the stage games of E from time 0 to t agree with those of h^t and the stage games from $\bar{t}+1$ onward are $(\widetilde{G}^0, \widetilde{G}^1, \ldots)$. (The intervening stage games may be arbitrary.) Then, for $t \leq \bar{t}$,

$$U(s|c, h^t) = \min_{E \text{ valid for } h^t} \left(U(s|E, h^t) - (1-\delta)\sum_{t'=t}^{\bar{t}} \delta^{t'-t}\widehat{u}(\underline{a}(G^{t'})) \right). \tag{5.7}$$

Denote the minimand in (5.7) as $\check{U}(s|E, h^t)$, and note for future reference the recursion

$$\check{U}(s|E, h^t) = (1-\delta)(u(s(h^t)) - \widehat{u}(\underline{a}(G^t))) + \delta\mathbb{E}^t[\check{U}(s|E, (h^t, s(h^t), G^{t+1}, \omega^{t+1}))] \tag{5.8}$$

when $t < \bar{t}$.

Let s be any one-shot SPE. At any history at time $\bar{t}$, the continuation game starting in the next period is expected to deterministically follow the environment $\widetilde{E}$, and so continuation play will be an SPE for this environment. Now we make the following claim: for any $t \leq \bar{t}$, at any history h^t, ending in any stage game G^t, we have $s(h^t) \in A^*(G^t, w_k)$, and $U(s|c, h^t) \in [\delta^{\bar{t}+1-t}\underline{\widetilde{U}}, \delta^{\bar{t}+1-t}\underline{\widetilde{U}} + B(w_k; G^t)]$.

We show this claim by downward induction on t. Suppose the claim holds for all times from $t+1$ to $\bar{t}$ (this hypothesis is vacuous in the base case $t = \bar{t}$). We prove that it holds for t. Consider any time-t history h^t, ending in some stage game G^t. Consider the specific valid environment in which $\overline{G}$ realizes at every date $t+1, \ldots, \bar{t}$ (again, if $t = \bar{t}$ there are

no such dates). Applying this particular environment in (5.7), we have

$$\begin{aligned} U(s|c,h^t) &\leq (1-\delta)\left(\sum_{t'=t}^{\infty}\delta^{t'-t}\mathbb{E}^t[u(a^{t'})] - \sum_{t'=t}^{\bar{t}}\delta^{t'-t}\widehat{u}(\underline{a}(G^{t'}))\right) \\ &= (1-\delta)\left((u(a^t)-\widehat{u}(\underline{a}(G^t))) + \sum_{t'=t+1}^{\bar{t}}\delta^{t'-t}\left(\mathbb{E}^t[u(a^{t'})]-\widehat{u}(\underline{a}(\overline{G}))\right)\right) \\ &\quad +\delta^{\bar{t}+1-t}\mathbb{E}^t[U(s|c,h^{\bar{t}+1})]. \end{aligned}$$

Since each $a^{t'}$ for $t < t' \leq \bar{t}$ always lies in $A^*(\overline{G}, w_k)$ by the induction hypothesis, each term $\left(\mathbb{E}^t[u(a^{t'})] - \widehat{u}(\underline{a}(\overline{G}))\right)$ is at most $(B(w_k;\overline{G}) - \delta w_k)/(1-\delta) < w_k$; and the final term is at most $\delta^{\bar{t}+1-t}(\underline{\widetilde{U}} + w_k)$ because continuation play starting at time $\bar{t}+1$ must be an SPE for $\widetilde{E}$. Combining gives

$$U(s|c,h^t) \leq (1-\delta)(u(a^t) - \widehat{u}(\underline{a}(G^t))) + (\delta - \delta^{\bar{t}+1-t})w_k + \delta^{\bar{t}+1-t}(\underline{\widetilde{U}} + w_k). \tag{5.9}$$

Meanwhile, consider the strategy s_1' that myopically deviates at h^t and follows s_1 everywhere else. Still writing $a^t = s(h^t)$, we have, for any valid environment E, that

$$\check{U}(s_1', s_{-1}|E,h^t) \geq (1-\delta)(\widehat{u}(a^t) - \widehat{u}(\underline{a}(G^t))) + \delta\cdot\delta^{\bar{t}+1-(t+1)}\underline{\widetilde{U}},$$

where if $t = \bar{t}$ the last term follows from the lower bound for SPE payoffs in environment $\widetilde{E}$, and otherwise it comes from (5.8) and the induction hypothesis for the continuation payoffs from time $t+1$ onward. Since this holds for each E, we have

$$U(s_1', s_{-1}|c,h^t) \geq (1-\delta)(\widehat{u}(a^t) - \widehat{u}(\underline{a}(G^t))) + \delta^{\bar{t}+1-t}\underline{\widetilde{U}}. \tag{5.10}$$

Since s is a one-shot SPE, $U(s|c,h^t) \geq U(s_1', s_{-1}|c,h^t)$; combining with (5.9) and (5.10) and rearranging gives

$$(1-\delta)(\widehat{u}(a^t) - u(a^t)) \leq \delta w_k.$$

Consequently, $a^t \in A^*(G^t, w_k)$, giving the first part of the claim for t. As for the second part, now that $a^t \in A^*(G^t, w_k) = A^*(G^t, w^*)$, we know $\widehat{u}(a^t) - \widehat{u}(\underline{a}(G^t)) \geq 0$ so that (5.10) gives the lower bound, and likewise $(1-\delta)(u(a^t) - \widehat{u}(\underline{a}(G^t))) + \delta w_k \leq B(w_k; G^t)$ so that (5.9) gives the upper bound.

This completes the proof of the claim.

Now consider the particular history $h^{\underline{t}} = (G^0_\bullet, \omega^0, a^0; \ldots; G^{\underline{t}}_\bullet, \omega^{\underline{t}})$, where the stage

games so far are as in the target environment $E_\bullet$, the random signals are those for which (5.2) is violated, and the actions so far are as specified by z. Suppose the one-shot SPE s supports $(E_\bullet, z)$. Consider the valid environment $E = (G^0_\bullet, \ldots, G^{\bar{t}}_\bullet, \widetilde{G}^0, \widetilde{G}^1, \ldots)$. We have

$$U(s|c, h^{\underline{t}}) \leq \check{U}(s|E, h^{\underline{t}}) = (1-\delta)\left(\sum_{t=\underline{t}}^{\bar{t}} \delta^{t-\underline{t}} \left(\mathbb{E}^{\underline{t}}[u(a^t)] - \widehat{u}(\underline{a}(G^t_\bullet))\right) + \sum_{t=\bar{t}+1}^{\infty} \delta^{t-\underline{t}} \mathbb{E}^{\underline{t}}[u(a^t)]\right)$$

(where the future actions a^t are as generated by s)

$$< (1-\delta)(\widehat{u}(a^{\underline{t}}) - \widehat{u}(\underline{a}(G^{\underline{t}}_\bullet))) - \delta^{\bar{t}+1-\underline{t}} w_k + (1-\delta) \sum_{t=\bar{t}+1}^{\infty} \delta^{t-\underline{t}} \mathbb{E}^{\underline{t}}[u(a^t)]$$

by applying the assumed violation of (5.2) (with the right side replaced by $\frac{\delta^{\bar{t}+1-\underline{t}}}{1-\delta} w_k$)

$$\leq (1-\delta)(\widehat{u}(a^{\underline{t}}) - \widehat{u}(\underline{a}(G^{\underline{t}}_\bullet))) - \delta^{\bar{t}+1-\underline{t}} w_k + \delta^{\bar{t}+1-\underline{t}} (\underline{\widetilde{U}} + w_k)$$

since play from period $\bar{t}+1$ onward is an SPE of $\widetilde{E}$ and so has payoff at most $\underline{\widetilde{U}} + w_k$.

On the other hand, consider the strategy s'_1 given by a one-shot optimal deviation at $h^{\underline{t}}$. For any valid environment E, applying (5.8) and the claim for the continuation payoff from date $\underline{t}+1$, we have

$$\check{U}(s'_1, s_{-1}|E, h^{\underline{t}}) \geq (1-\delta)(\widehat{u}(a^{\underline{t}}) - \widehat{u}(\underline{a}(G^{\underline{t}}_\bullet))) + \delta \cdot \delta^{\bar{t}+1-(\underline{t}+1)} \underline{\widetilde{U}}.$$

Since this holds for any E, we have $U(s'_1, s_{-1}|c, h^{\underline{t}}) \geq (1-\delta)(\widehat{u}(a^{\underline{t}}) - \widehat{u}(\underline{a}(G^{\underline{t}}_\bullet))) + \delta^{\bar{t}+1-\underline{t}} \underline{\widetilde{U}} > U(s|c, h^{\underline{t}})$. So the deviation at date $\underline{t}$ is strictly profitable, a contradiction.

□

The preferences constructed in the proof of Theorem 5.4 have a simple interpretation: After renormalizing the stage game payoffs so that $\widehat{u}(\underline{a}(G)) = 0$ for each G, the long-run player acts as though any sequence of stage games up until time $\bar{t}$ as possible, but the subsequent stage games definitely follow $\widetilde{E}$, and payoffs are evaluated by the worst case over the uncertain early stages.

This result provides a "positive foundation" for focusing on XPE-supportable outcomes: Even if the analyst thinks that (one-shot) SPE is the appropriate prediction of behavior, the set of XPE-supportable outcomes nonetheless plays a distinguished role, in that these are exactly the outcomes that are guaranteed to be supportable no matter what the long-run player's attitude toward the future stage games may be (if he may be

ambiguity averse).

5.5 Full outcomes

All of the preceding results of this section have analogues for full outcomes rather than realizable outcomes. We briefly develop the statements, leaving proofs to Appendix A.

Evidently, a full outcome z can only be supported in XPE if each of the realizable outcomes belonging to it can, or equivalently, if each such realizable outcome satisfies (5.1)–(5.2). The converse is also true, and thus:

Theorem 5.5. *A full outcome z is supported by some XPE if and only if each of the realizable outcomes belonging to it can be supported by some XPE.*

Actually, more can be said. Recall that (5.2) implied (5.5), by taking the limit as $\bar{t} \to \infty$. It turns out that for full outcomes, we can replace (5.2) with this weaker condition:

Theorem 5.6. *A full outcome z is supported by some XPE if and only if each of the realizable outcomes belonging to it satisfies (5.1) and (5.5).*

We no longer need to worry about what happens when the realized environment departs from the specified outcome, because a full outcome by definition considers all possible environments.

Say that a strategy profile $\underline{s}$ is a *universal penal code for full outcomes* if it has the following property: For every full outcome z that is supportable in XPE, there is in particular an XPE supporting z where, following any initial deviation by player 1 (i.e. a period-t history with $a^{t'} = z(G^0, \omega^0, \ldots, G^{t'}, \omega^{t'})$ for each $t' = 0, \ldots, t-1$, followed by an action profile (a'_1, a^t_{-1}) with $a'_1 \neq z_1(G^0, \omega^0, \ldots, G^t, \omega^t)$), continuation play is given by $\underline{s}$. Then:

Theorem 5.7. *There exists a universal penal code for full outcomes.*

(In contrast to realizable outcomes, here, specifying a full outcome, together with the punishments after 1's deviations, fully describes the strategy profile—aside from the detail that a strategy profile should say what happens after deviations by short-run players, but these can simply be ignored.)

Finally:

Theorem 5.8. *Assume that A is finite. If a full outcome z is not supported by any XPE, then there exists a dynamic ambiguity index c such that z is not supported by any one-shot SPE for c.*

6 Discussion

In some sense, the analysis so far seems quite straightforward: there is some maximum reward gap w^* that can be credibly promised to the long-run player; a contemplated outcome is supportable just so long as doing so never requires accumulating a debt for forgoing temptations (possibly in expected terms) that exceeds w^*. However, as discussed in the introduction, our framework involves particular features, in particular the restriction that there is only a single long-run player for whom dynamic incentives need to be provided, and also the availability of public randomization. We will show here that the results change if either of these conditions is removed. In particular, a universal penal code may no longer exist. (Relatedly, an *optimal* penal code, i.e. an XPE giving the lowest payoff in any environment, can also fail to exist.) Given the central role of the universal penal code in the analysis, this suggests that the theory of XPE's in these broader settings would have to look very different.

In standard repeated games, the usual recursive analysis by way of continuation values, and its corresponding role for optimal penal codes, applies equally well with or without public randomization, and with any combination of long-run and short-run players. This contrast suggests that the mapping between the reward-punishment gaps emphasized in our analysis and continuation values as in the usual approach is not just a mechanical renormalization.

For ease of exposition, the examples in this section are presented in a slightly different framework than the main model: we assume a nonstationary framework (i.e. for each period t, there is a different set of possible stage games $\mathcal{G}^t$); we also assume a finite horizon, and no discounting. None of these changes matters conceptually. For completeness, Appendix B shows how to build on the examples below to express the same ideas while retaining stationarity and discounting.

For brevity, we skip over some of the formalities for these examples.

6.1 No public randomization

We first consider a framework without public randomization available. Our example features three periods, $t = 0, 1, 2$. The best way to give the long-run player a low payoff starting at date 1 depends which stage game will be realized at date 2, because the latter determines how expensive it is to deter stage-1 deviations. This means that there is no XPE that gives the lowest payoff starting from date 1 in every environment. This in turn leads to the lack of a universal penal code.

Example 6.1. Consider the sets of stage games shown in Figure 4. There is one possible stage game in each period $t = 0, 1$, and two possible stage games in period 2. As with Figure 3, part (a) presents the stage games in standard matrix form, while part (b) shows the sets $A^*(G)$ in each stage game and the values of $u, \widehat{u}$ for each.

G^0 :

	a	b
a	$0,1$	$4,0$
b	$4,0$	$0,1$

$\rightarrow$ G^1 :

	c	d	e
c	$0,1$	$0,0$	$11,0$
d	$5,0$	$6,1$	$0,0$
e	$0,0$	$7,0$	$11,1$

$\nearrow$ $\searrow$

G^2 :

	f	g	h
f	$0,1$	$1,0$	$10,0$
g	$0,0$	$1,1$	$10,0$
h	$0,0$	$1,0$	$10,1$

$G^{2'}$:

	i	j
i	$0,1$	$5,0$
j	$0,0$	$5,1$

(a)

G^0 :

	aa	bb
u	0	0
$\widehat{u}$	4	4

$\rightarrow$ G^1 :

	cc	dd	ee
u	0	6	11
$\widehat{u}$	5	7	11

$\nearrow$ $\searrow$

G^2 :

	ff	gg	hh
u	0	1	10
$\widehat{u}$	0	1	10

$G^{2'}$:

	ii	jj
u	0	5
$\widehat{u}$	0	5

(b)

Figure 4: Example without public randomization. No universal penal code exists.

Let s_c denote the XPE profile for the subgame beginning in period 1 that plays actions cc, hh, jj along the path of play and, if player 1 deviates in period 1, plays ff or ii in period 2 accordingly. (As usual, deviations by 2 can be ignored.) Player 1's total payoff across the two stages is 10 or 5, depending whether G^2 or $G^{2'}$ is realized.

Let s_d denote the XPE profile beginning in period 1 that plays actions dd, gg, jj on-path, with ff or ii in period 2 if player 1 deviates in period 1. Player 1's total payoff for the two periods is 7 or 11, respectively.

In the three-period game, the realizable outcome (deterministic, of course) with actions aa, ee, ff can be supported in XPE. Namely, specify aa, ee, ff, jj as on-path actions. If player 1 deviates in period 0, then switch to s_d as punishment. (Deviation in period 1

only can be ignored, since it brings no within-period gain.) For both realizations of the period-2 game, the punishment is sufficient to deter the period-0 deviation.

The realizable outcome with actions aa, ee, ii can also be supported in XPE: specify aa, ee, hh, ii, and deter a period-0 deviation by using s_c as punishment.

However, there is no punishment that can support both of these outcomes at once. Indeed, to support aa, ee, ff, the punishment after a period-0 deviation to b has to deliver total payoff ≤ 7 across the two remaining periods in environment (G^0, G^1, G^2), which requires beginning with dd (since ee is clearly too generous, and cc would have to be followed by hh, otherwise 1 would deviate in period 1, but hh is also too generous). To support aa, ee, ii, the punishment after b has to deliver total payoff ≤ 7 across G^1 and $G^{2\prime}$, but this cannot be done using dd (because dd must be followed by jj to deter a period-1 deviation, but this is again too generous). So, no one punishment can support both outcomes.

This shows that Theorem 5.3 fails without public randomization (and the same is true for Theorem 5.7). Note that it also shows that Theorem 5.5 fails, since the full outcome aa, ee, ff, ii cannot be supported in XPE even though its consistuent realizable outcomes can. Finally, a minor variant of this example suffices to give a case with a single full outcome that *can* be supported, but only by using different punishments for different date-1 histories. Namely, create a second stage-0 game $G^{0\prime}$ that is a copy of G^0, and now consider the full outcome that specifies aa, ee, ff, jj if G^0 is realized, and aa, ee, hh, ii if $G^{0\prime}$ is realized. This contrasts with stochastic games, where universal penal codes for full outcomes do exist (Kitti, 2016).

△

6.2 Multiple long-run players

Let us now restore public randomization, but suppose that there are two long-run players, who both act to maximize the sum of payoffs across periods. We again give a three-period example where there is no universal penal code. Although randomization is allowed, deterministic outcomes will suffice for our example.

Similar to the previous example, the most effective way to give the long-run player a low payoff starting at date 1 depends what stage game will be realized at date 2. Here, the reason is that there is an action profile at date 1 that gives a low payoff to player 1 but also a high temptation to deviate for player 2. It may or may not be possible to reward player 2 in the next period for forgoing this temptation without also giving a high

payoff to player 1, depending which period-2 stage game is realized.

Example 6.2. Again, one possible stage game in each period $t = 0, 1$, and two in period 2. The relevant stage games are illustrated in Figure 5.

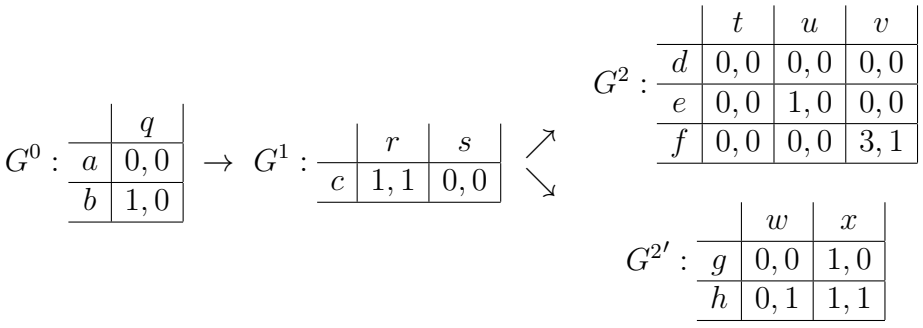

Figure 5: Example with two long-run players. No universal penal code exists.

Let s_r denote the XPE for the subgame starting in period 1 that consists of playing cr followed by dt or gw. (Deviation by player 2 in period 1 can be ignored, since it brings no gain.) This delivers to the two players total payoffs of $(1, 1)$ across the two periods, both when G^2 is realized and when $G^{2'}$ is realized.

Let s_s (with apologies for the notation) be the XPE starting in period 1 that plays cs followed by fv or hw on-path, and that punishes a deviation by player 2 at period 1 by following up with dt or gw. This delivers total payoffs across the two periods of $(3, 1)$ if G^2 is realized and $(0, 1)$ if $G^{2'}$ is realized.

In the overall game, the (deterministic) realizable outcome with actions aq, cr, eu can be supported, with hx being chosen in period 2 if $G^{2'}$ is realized. To see this, we just need to be able to deter deviation to b in period 0 (since the specified play constitutes a stage Nash in each subsequent period), and this can be done using s_r as a punishment. The realizable outcome with actions aq, cr, gw can also be supported, with fv being chosen on-path in period 2 if G^2 is realized. To do this, again we only need to deter the deviation to b, and this can be done by punishing with s_s.

However, no punishment can support both of these outcomes in XPE. Such a punishment would need to deliver an (expected) payoff to player 1 of ≤ 1 across periods 1–2 if G^2 is realized and a payoff ≤ 0 if $G^{2'}$ is realized. The latter, in particular, means that cs must be played with probability 1 in period 1. But then the only way to deter 2 from deviating is to reward her with fv (again with probability 1) if G^2 is realized. This means that player 1's total payoff across the two periods is 3 in this environment, contrary to

the requirement. Note also that this argument has accounted for the availability of public randomization.

This shows that the analogue of Theorem 5.3 with two long-run players does not hold. The example also can be modified to show that various other results from the main analysis do not extend, just as was done in Example 6.1.

△

6.3 Why the difference?

Why does the recursive approach fail to extend to these settings? It may be helpful to try to imagine a common generalization of the approach taken here and APS. Both seek to characterize the set of appropriately-normalized reward vectors that can be achieved in equilibrium, so as to determine which deviations can be deterred. In APS the normalized reward is just the total payoff, whereas here it is the reward-punishment gap. $B(W)$ is the set of normalized rewards that can be attained, given that the continuation rewards are taken from the set W. But what does "attained" mean?

The incentive constraints require inequalities: as long as we can ensure a normalized reward of at least v for each stage game G that might materialize, we can deter a deviation gain of corresponding size. Doing so may involve using different rewards for different realizations of G. The promise-keeping constraints, however, require equalities: a lower bound on the continuation reward is not enough, because the total reward must be kept low in order to be usable as a punishment at the next stage of the recursion. Thus, for the recursive technique to apply, the same $B(\cdot)$ operator needs to be able to calculate both

$$\{v \mid \text{ can ensure normalized reward } \geq v \text{ for all } G\}$$

and

$$\{v \mid \text{ can ensure normalized reward } = v \text{ for all } G\}.$$

In the "one-dimensional" case studied here with public randomization, the two sets are just intervals, and they coincide. In general, this will not be true. (In APS, where there is only one game G, the equality set uniquely determines the inequality set, so it suffices to recurse on the equality set only.)

7 Summary

Repeated games are a standard modeling tool for studying dynamic incentives in long-run interactions. But the standard model adopts a stylized framework in which the same game is played in every period, and the tools for studying this model apply in payoff space. This paper has explored a generalization to a setting where multiple stage games may arise in each period, and the stage game may vary unpredictably from one period to the next, to try to understand how much of the standard toolkit carries through without well-defined payoffs.

We adopted ex-post perfect equilibrium as a solution concept. Under two significant restrictions—a single long-run player interacting with a series of short-run players, and availability of public randomization—the recursive technique from APS adapts to identify the maximum gap between future reward and punishment that can credibly be promised to the long-run player. This leads to a characterization of the outcomes attainable in equilibrium, as ones for which there can never be an expected gain above this maximum gap from deviating repeatedly across an interval of successive periods—or equivalently (for deterministic outcomes), the outcomes for which the debt owed to compensate past obedience never exceeds the maximum gap. Any such outcome is supportable by using a single worst punishment following any deviation. And, using the characterization of supportable outcomes, we are able to connect the concept of ex-post perfect equilibrium to the more standard one of subgame-perfect equilibrium, by showing that the ex-post perfect equilibrium outcomes are exactly the ones that can be supported for any preference specification for the long-run player, although this equivalence requires allowing ambiguity aversion over the future stage games. We also saw that when we try to extend to multiple long-run players, or we drop public randomization, the analysis does not replicate (and, in particular, universal penal codes no longer exist).

A few words of broader perspective: The standard model of repeated games has proven extremely valuable for understanding the possibilities and limitations of providing incentives through repeated interactions. Central to the analysis of this model are certain ideas, such as optimal penal codes and their use to characterize outcomes in payoff space. It seems fitting to ask to what extent these ideas follow from the economic concept of dynamic incentives, as opposed to being convenient features of a particular mathematical model. For this distinction to be meaningful, it is necessary to argue that other models are possible. This paper has undertaken a step in this direction, by offering a minimal departure from the canonical model and showing that, indeed, the ideas do not fully ex-

tend (as the negative examples show), but there is considerable overlap. Perhaps some more general modeling framework will better elucidate the commonalities.

A Omitted proofs

Proof of Lemma 4.2. Consider any decreasing sequence $w_k \to w$. Because $B(w; G)$ is increasing, the limit $\lim_k B(w_k; G)$ is well-defined (with value $-\infty$ if eventually $B(w_k; G) = \infty$), and right-continuity will follow if we can show that $B(w; G) \geq \lim B(w_k; G)$.

We can assume that $B(w_k; G) \neq -\infty$ for all k, since otherwise monotonicity implies $B(w; G) = -\infty = \lim B(w_k; G)$. Granted this, take $a_k, a'_k \in A^*(G, w_k)$ attaining the max and min in the definition of $B(w_k; G)$. By compactness, we can pass to a subsequence for which a_k and a'_k converge to limits a_∞, a'_∞. Continuity of u and $\widehat{u}$ then imply that $a_\infty, a'_\infty \in A^*(G, w)$, and

$$\begin{aligned} B(w; G) &\geq (1-\delta)(u(a_\infty) - u(a'_\infty)) + \delta w \\ &= \lim_k (1-\delta)(u(a_k) - u(a'_k)) + \delta w_k \\ &= \lim_k B(w_k; G). \end{aligned}$$

Thus, $B(w; G)$ is right-continuous. For $B(w)$, again consider a decreasing sequence $w_k \to w$. If $B(w) = -\infty$, then $B(w; G) = -\infty$ for some G, hence the previous argument implies $B(w_k) = B(w_k; G) = -\infty$ for nearby w_k. Otherwise, if the desired right-continuity fails then there exists $\epsilon > 0$ such that $B(w_k) > B(w) + \epsilon$ for all k. Take G such that $B(w; G) < B(w) + \epsilon/2$; then right-continuity of $B(w; G)$ for this specific G implies $B(w_k) \leq B(w_k; G) < B(w) + \epsilon$ for large enough k, a contradiction.

□

Proof of Lemma 4.3. First, note that $w_k > w^*$ by induction: This is clearly true for w_0; then $w_1 > B(w_0) > B(w^*) = w^*$ by strict monotonicity of B, and for $k \geq 2$ we then have $w_k = (B(w_{k-1}) + B(w_{k-2}))/2 > (B(w^*) + B(w^*))/2 = w^*$ by strict monotonicity and induction hypothesis.

In particular, the terms w_k never fall to $-\infty$. Now we prove the ensuing statements:

1: We have $w_1 < w_0$ from the construction, and then $B(w_1) < B(w_0) < w_1$ by strict monotonicity, from which $w_2 = (B(w_0) + B(w_1))/2 < w_1$. Now proceed by induction: if $k > 2$ and $w_{k-1} < w_{k-2} < w_{k-3}$, then $w_k = (B(w_{k-1}) + B(w_{k-2}))/2 < (B(w_{k-2}) + B(w_{k-3}))/2 = w_{k-1}$ by strict monotonicity.

2: For $k = 1$ this is given; for $k \geq 2$ we have $w_k = (B(w_{k-1}) + B(w_{k-2}))/2 > B(w_{k-1})$ using strict monotonicity of B and the fact that $w_{k-2} > w_{k-1}$.

3: Since the sequence is decreasing and bounded below by w^*, it has a limit w_∞. Right-continuity of B implies $w_\infty = \lim_k w_k = \lim_k (B(w_{k-1}) + B(w_{k-2}))/2 = (B(w_\infty) + B(w_\infty))/2 = B(w_\infty)$. But since $w_\infty \geq w^*$, and no value greater than w^* is a fixed point of B, we have equality.

□

Proof of Proposition 5.2. As noted in the text, the conditions (5.1)–(5.2) as stated are equivalent to requiring $a^t \in A^*(G^t)$ for all t and (5.2) for all $\underline{t} \leq \bar{t}$. So it suffices to check that, in the deterministic case, the latter is equivalent to $d^t \leq \frac{\delta}{1-\delta} w^*$ for all t. Rewrite (5.2) as

$$\frac{1}{\delta^{\bar{t}-\underline{t}}} \left(\widehat{u}(a^{\underline{t}}) - u(a^{\underline{t}}) \right) + \sum_{t=\underline{t}+1}^{\bar{t}} \delta^{t-\bar{t}} \left(\widehat{u}(\underline{a}(G^t)) - u(a^t) \right) \leq \frac{\delta}{1-\delta} w^*. \tag{A.1}$$

(We have removed the expectation operator since a^t is no longer random.) Denoting the left-hand side of (A.1) by $d^{\underline{t},\bar{t}}(z)$, requiring (5.2) for all $\underline{t}, \bar{t}$ is then equivalent to $\max_{\underline{t} \in \{0,\ldots,\bar{t}\}} d^{\underline{t},\bar{t}}(z) \leq \frac{\delta}{1-\delta} w^*$ for all $\bar{t}$. But it is easy to see by induction that $d^{\bar{t}}(z) = \max_{\underline{t} \in \{0,\ldots,\bar{t}\}} d^{\underline{t},\bar{t}}(z)$.

□

Proof of Theorem 5.5. Necessity is immediate. For sufficiency, we need to argue that it suffices for each realizable outcome to satisfy (5.1)–(5.2). This follows from sufficiency of the weaker conditions in Theorem 5.6, so we defer to that proof.

□

Proof of Theorem 5.6. Again, we already have necessity, so we focus on sufficiency. Adapting the proof of Theorem 5.1, we construct a strategy profile s as follows: at any history where actions have not yet deviated from z, play as specified by z; when a deviation first occurs at some period $t-1$, play according to $s[0]$ from period t onward. Since z specifies an intended action profile for every possible initial sequence of stage games (and random signals), this description fully specifies a strategy profile.

As in the earlier proof, we just need to check the incentives of player 1 at any history h^t where a deviation has not yet occurred. Fix any environment $E = (G^0, G^1, \ldots)$ such that h^t is consistent with E. Let $a^{t'}$, for each $t' \geq t$, be the actions specified by z in this environment (which may depend on the already-realized signals $\omega^{0,\ldots,t}$, as well as the

random future signals). If player 1 conforms to s, then the payoff starting at h^t from conforming is

$$(1-\delta)\left(\sum_{t'=t}^{\infty}\delta^{t'-t}\mathbb{E}^t[u(a^{t'})]\right). \tag{A.2}$$

If player 1 deviates from s, then subsequent play transitions to $s[0]$ and so the payoff from $t+1$ onward is given by $\underline{U}(E^{-(t+1)})$. Therefore, the payoff from deviating optimally, as measured from h^t, is

$$(1-\delta)\widehat{u}(a^t)+\delta\underline{U}(E^{-(t+1)})=(1-\delta)\left(\widehat{u}(a^t)+\sum_{t'=t+1}^{\infty}\delta^{t'-t}\widehat{u}(\underline{a}(G^{t'}))\right). \tag{A.3}$$

Rearranging (5.5) tells us exactly that (A.2) is greater than or equal to (A.3). Hence, deviating is never profitable, in any environment.

□

Proof of Theorem 5.7. It is immediate from the proof of Theorem 5.6 that $s[0]$ is a universal penal code for full outcomes.

□

Proof of Theorem 5.8. If z is not supportable in XPE, then by Theorem 5.5, one of its constituent realizable outcomes is not either. By Theorem 5.4, there is some dynamic ambiguity index for which this realizable outcome is not supportable in one-shot SPE, and a fortiori the full outcome z is not either.

□

B Stationary versions of counterexamples

We sketch here constructions analogous to Examples 6.1 and 6.2, but retaining the stationary structure of the original model (including infinite horizon and discounting).

Example B.1. For this example, we assume one long-run player and no public randomization, as in Example 6.1. We assume $\mathcal{G}$ consists of five stage games as shown in Figure 6. The discount factor is $\delta = 1/10$. (This makes the numbers simple, but similar examples can be constructed for δ arbitrarily close to 1.) For brevity, we avoid writing out the games in traditional matrix form, and instead just directly name the action profiles assumed to comprise $A^*(G)$ and list the values of u and $\widehat{u}$, as in Figure 4(b).

G_1 :

	a	b	v
u	0	0	10000
$\widehat{u}$	0	4	10000

G_2 :

	c	d	e	w
u	0	60	110	10000
$\widehat{u}$	50	70	110	10000

G_3 :

	f	g	x
u	0	100	10000
$\widehat{u}$	0	100	10000

G_4 :

	h	i	y
u	0	500	10000
$\widehat{u}$	0	500	10000

G_5 :

	j	z
u	0	1000000
$\widehat{u}$	0	1000000

Figure 6: Stationary example of no universal penal code without public randomization.

There exists an XPE that supports the realizable outcome $(c, i, j, j, j, \ldots)$ (here we suppress the list of stage games involved, for brevity). In particular, specify that if "Nature deviates" by ever choosing a stage game different from that specified by the outcome (and player 1 has not deviated in the past), then reward actions (v, w, x, y, z) are played from then onward; if player 1 deviates from c in the first period, then the worst stage-Nash actions (a, e, f, h, j) are played subsequently. All other deviations can be ignored since there are no short-run gains.

There exists an XPE that supports the realizable outcome $(d, g, j, j, j, \ldots)$. If Nature ever deviates, use reward actions as above; if player 1 deviates from d in the first period, then use the worst stage-Nash actions in all subsequent periods.

These, in turn, can be used to support two different realizable outcomes that start with b being played in G_1 in period 0. First, we can support $(b, e, f, j, j, j, \ldots)$ by specifying that reward actions are to be played if Nature deviates, and a deviation from b by player 1 is punished as follows: in period 1, if the stage game drawn is G_2, we commence the $(d, g, j, j, j, \ldots)$ equilibrium, and otherwise we simply play worst stage-Nash in every period. It is straightforward to check that this deters the deviation to b in every possible environment (note that there are several cases to check depending when the environment first differs from the proposed outcome).

Second, we can support $(b, e, h, j, j, j, \ldots)$ by specifying that reward actions are to be played if Nature deviates, and a deviation from b by player 1 is punished as follows: in period 1, if G_2 is drawn, then we commence the $(c, i, j, j, j, \ldots)$ equilibrium, and otherwise we play worst stage-Nash in every period.

Finally, we claim there is no XPE punishment $\underline{s}$ that can support both the $(b, e, f, j, j, j, \ldots)$ and $(b, e, h, j, j, j, \ldots)$ outcomes, thus showing nonexistence of an optimal penal code in this environment. Indeed, to be an effective deterrent, $\underline{s}$ would have to give a total payoff of at most 63 in both environments $(G_2, G_3, G_5, G_5, G_5, \ldots)$ and $(G_2, G_4, G_5, G_5, G_5, \ldots)$.

We show that no XPE $\underline{s}$ can have this property.

Evidently, if G_2 is drawn initially then either c or d must be played. Suppose that c is played. In the continuation environment $(G_3, G_5, G_5, \ldots)$, the total payoff needs to be at most 630. This means that play should begin with f or g, and j must be played for at least the next three periods. But this in turn means that if the continuation environment turns out to be instead $(G_3, G_5, G_5, G_5, G_3, G_3, G_3, \ldots)$, then the total payoff is at most $(1-\delta)(100 + (\delta^4 + \delta^5 + \cdots) \cdot 10000) = 91$, which is not enough reward to prevent the deviation from c in the preceding period. Correspondingly, if $\underline{s}$ begins by playing d in G_2, then the continuation in environment $(G_4, G_5, G_5, G_5, \ldots)$ needs to have payoff at most 90. It therefore needs to begin with h followed by at least three copies of j. This means that if the continuation environment is instead $(G_4, G_5, G_5, G_5, G_3, G_3, \ldots)$ then this continuation has payoff no more than 1, which means it cannot prevent the deviation from d in the initial period.

△

Example B.2. We now restore public randomization but consider two long-run players. We build on Example 6.2, using the ideas of Example B.1 to extend to a stationary environment. The possible stage games are shown in Figure 7. We again assume the discount factor is $\delta = 1/10$ (for both players).

G_1:

	q	r
a	$0, 0$	$0, 0$
b	$1, 0$	$0, 0$
c	$0, 0$	$10000, 10000$

G_2:

	s	t	u
d	$10, 10$	$0, 0$	$0, 0$
e	$0, 0$	$0, 0$	$10000, 10000$

G_3:

	v	w	x
f	$0, 0$	$0, 0$	$0, 0$
g	$0, 0$	$100, 0$	$0, 0$
h	$0, 0$	$0, 0$	$30000, 10000$

G_4:

	y	z
i	$0, 0$	$10000, 0$
j	$0, 10000$	$10000, 10000$

Figure 7: Stationary example of no universal penal code with two long-run players.

We will use the term "reward" for the high-payoff profile in each stage game (cr, eu, hx, jz), which is always stage Nash, and "punishment" for bq, ds, fv, iy, which achieves the lowest payoff for player 1 among stage-Nash profiles.

Let s_s be the XPE that always plays the punishment action profile. Deviations are simply ignored. This is an XPE since it plays a stage Nash in every period and deviations

do not affect future play.

Let s_t be the XPE that does the following: If G_2 is drawn in the initial period, then dt is to be played. If player 2 does not deviate from t, then in the next period, cr, eu, hx, or jy is to be played depending on the stage game (i.e. the reward profile, except that we play jy instead of jz in G_4); and after that, the punishment profile is played in all subsequent periods. If 2 does deviate in the initial period, then the punishment profile is played in all subsequent periods. If the initial stage game is not G_2, then we simply play the punishment profile in every period. All deviations are ignored except deviation by 2 in the initial period as described above. Note that this is an XPE: it specifies stage Nash in every period, except in the initial period if G_2 is drawn, but the punishment the next period is sufficient to deter 2 from deviating to s.

Now, we can use these to support two different (deterministic) realizable outcomes that begin with aq being played in G_1 in period 0. First, the realizable outcome $(aq, ds, gw, fv, fv, fv, \ldots)$ can be supported as follows. If Nature deviates, play reward profiles forever. Deviations by players are ignored unless they bring a short-run gain, as usual. So we need only worry about deviation by player 1 to b in period 0, and we specify that this deviation is punished by switching to s_s. We can check that this punishment deters the deviation in every environment (as in Example B.1, there are cases to check depending when the stage games first diverge from those in the target outcome).

Second, the realizable outcome $(aq, ds, iy, iy, iy, \ldots)$ can be supported by specifying that a deviation by Nature is followed with reward profiles, while a deviation by player 1 in period 0 is punished by following with s_t. Again, this punishment deters the deviation in all environments (with several cases to check).

Finally, we cannot support both $(aq, ds, gw, fv, fv, fv, \ldots)$ and $(aq, ds, iy, iy, iy, iy, \ldots)$ using the same XPE $\underline{s}$ to punish player 1 for a period-0 deviation in both cases. Such an XPE would have to give a payoff to player 1 of at most 9 in the environment $(G_2, G_3, G_3, G_3, \ldots)$ and at most 0 in the environment $(G_2, G_4, G_4, G_4, \ldots)$. The latter implies that in the initial period, in G_2, only action profiles with payoffs $(0, 0)$ can be played with positive probability (accounting for the ability to use public randomization). However, player 2 needs to be guaranteed a total payoff at least 9 in environment $(G_2, G_3, G_3, G_3, \ldots)$, since she can get this much by myopically deviating in the initial period. This means that in this environment, $\underline{s}$ has to give player 1 an expected payoff of at least 27, because 1's payoff is always at least three times 2's payoff (in the initial period this holds because both are getting payoff 0, as argued above, and in subsequent periods it holds because every action profile in G_3 satisfies this relation). This contradicts the requirement that

1's payoff from $\underline{s}$ in this environment should be at most 9.

$\triangle$

References

Abreu, Dilip (1988) "On the theory of infinitely repeated games with discounting," *Econometrica*, 383–396.

Abreu, Dilip, David Pearce, and Ennio Stacchetti (1990) "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica*, 1041–1063.

Bergemann, Dirk and Stephen Morris (2005) "Robust mechanism design," *Econometrica*, **73** (6), 1771–1813.

Chung, Kim-Sau and Jeffrey C Ely (2007) "Foundations of dominant-strategy mechanisms," *The Review of Economic Studies*, **74** (2), 447–476.

Ely, Jeffrey C, Johannes Hörner, and Wojciech Olszewski (2005) "Belief-free equilibria in repeated games," *Econometrica*, **73** (2), 377–415.

Epstein, Larry G and Martin Schneider (2003) "Recursive multiple-priors," *Journal of Economic Theory*, **113** (1), 1–31.

Fudenberg, Drew, David M Kreps, and Eric S Maskin (1990) "Repeated games with long-run and short-run players," *Review of Economic Studies*, **57** (4), 555–573.

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

Hansen, LarsPeter and Thomas J Sargent (2001) "Robust control and model uncertainty," *American Economic Review*, **91** (2), 60–66.

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

Kitti, Mitri (2016) "Subgame perfect equilibria in discounted stochastic games," *Journal of Mathematical Analysis and Applications*, **435** (1), 253–266.

Maccheroni, Fabio, Massimo Marinacci, and Aldo Rustichini (2006) "Dynamic variational preferences," *Journal of Economic Theory*, **128** (1), 4–44.

Mailath, George J and Larry Samuelson (2006) *Repeated games and reputations: long-run relationships*: Oxford University Press.