Supplemental Material for "Dynamic Incentives in Incompletely Specified Environments"

Gabriel Carroll, University of Toronto
gabriel.carroll@utoronto.ca

August 11, 2024

This Supplemental Material contains additional results and examples. Section S-1 gives details about the connection between XPE and SPE outcomes that was overviewed briefly in Section 5.4 of the main paper. Section S-2 provides versions of the counterexamples from Section 6 that have a stationary structure.

S-1 Details of XPE-SPE connection

We flesh out here the results described in Section 5.4. Throughout this section, we assume for simplicity that A is finite, i.e., the set of stage games and the action spaces are all finite.

We first consider the setting where the long-run player has expected utility with respect to some belief about the evolution of the stage games. We start by developing the formalism.

The long-run player's belief is described by a *stage game process*, which consists of a specification of $\pi_{(G^0,\dots,G^t)} \in \Delta(\mathcal{G})$ for each initial history of stage games $(G^0,\dots,G^t)$, determining the distribution over G^{t+1} given the previous realizations. We denote such a process by π.[1] 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

[1] Alternatively, we could define a stage game process directly as an ex-ante 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.

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 exactly as in the main text. Then, the statement that s supports the realizable outcome (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 every stage game process. Below is the example showing that the converse is not true.

Example S-1.1. Consider two possible stage games, G and G', as shown in Figure S-1. As with Figure 3 in the main text, part (a) presents them in the usual matrix form, while part (b) just shows the values of u and $\widehat{u}$ on the action profiles in $A^*(G), A^*(G')$.

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 S-1: 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 arises every period, and the action profiles played are $(bb, cc, bb, aa, aa, aa, aa, \ldots)$. This outcome does not satisfy (5.2)

with $\underline{t} = 0$, $\bar{t} = 2$, so it cannot be supported in XPE. However, we claim that it can always be supported in SPE for any 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 subgame 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 $(1-\delta)[7 + \delta \cdot (-40) + \delta^2 \cdot 16] < 0$.

△

We now proceed to broaden the allowed preferences of the long-run player to accommodate ambiguity aversion. We adapt the dynamic variational preferences of Maccheroni, Marinacci and Rustichini (2006) to our setting. Such preferences 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, bounded below, and 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

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

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. (In particular, while we allow ambiguity over the stage games, the public randomization device remains unambiguous.)

Note that expected utility with respect to a particular stage game process π is a special

case of such preferences, 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 strategies s_1' that coincide with s_1 at every history except h^t.

Preferences (S-1) 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.[2] However, it remains the case that if s is an XPE then it is also an SPE (and not just a one-shot SPE) for any such preferences. 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. This observation is also made by Krasikov and Lamba (2023).

With this broader class of preferences, we do have our desired "converse" result: any outcome that is supportable in SPE for all preferences of the form (S-1) is in fact supportable in XPE. Moreover, this result holds even if SPE is relaxed to one-shot SPE; thus, it is not relying on the dynamic inconsistency as a device to rule out potential SPE's.

Theorem S-1. *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 $E_\bullet = (G_\bullet^0, G_\bullet^1, \ldots)$. 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 $E = (G_\bullet^0, G_\bullet^1, \ldots, G_\bullet^t, \widetilde{G}^0, \widetilde{G}^1, \widetilde{G}^2, \ldots)$. The proof of Lemma 4.5 shows that, in any SPE for this

[2] Maccheroni, Marinacci and Rustichini (2006) do identify a subclass of dynamic variational preferences that are dynamically consistent. However, this subclass is generally incompatible with our maintained assumption that future stage games are independent of future random signals; and if we were to drop this assumption, we would lose the result that every XPE is always an SPE.

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 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, 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 (S-1) is attained at a corner of the set of possible ψ's, which allows us to simplify (S-1) 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 $t \leq \bar{t}$,

$$U(s|c, h^t) = \min_{E \text{ valid for } h^t} \left(U(s|E, h^t) - (1-\delta) \sum_{t'=t}^{\bar{t}} \delta^{t'-t} \widehat{u}(\underline{a}(G^{t'})) \right). \tag{S-2}$$

Denote the minimand in (S-2) as $\check{U}(s|E, h^t)$, and note for future reference the recursion

$$\check{U}(s|E, h^t) = (1-\delta)(u(s(h^t)) - \widehat{u}(\underline{a}(G^t))) + \delta \mathbb{E}^t[\check{U}(s|E, (h^t, s(h^t), G^{t+1}, \omega^{t+1}))] \tag{S-3}$$

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 (S-2), 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

$$U(s|c, h^t) \leq (1-\delta)(u(a^t) - \widehat{u}(\underline{a}(G^t))) + (\delta - \delta^{\bar{t}+1-t})w_k + \delta^{\bar{t}+1-t}(\underline{\widetilde{U}} + w_k). \tag{S-4}$$

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 (S-3) and the induction hypothesis for the continuation payoffs from time $t+1$ onward. Since this holds for each E, we have

$$U(s_1', s_{-1}|c, h^t) \geq (1-\delta)(\widehat{u}(a^t) - \widehat{u}(\underline{a}(G^t))) + \delta^{\bar{t}+1-t}\underline{\widetilde{U}}. \tag{S-5}$$

Since s is a one-shot SPE, $U(s|c, h^t) \geq U(s_1', s_{-1}|c, h^t)$; combining with (S-4) and (S-5)

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 (S-5) 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 (S-4) 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 (S-3) 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.

□

Finally, our converse result holds for full outcomes as well:

Theorem S-2. *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.*

Proof. If z is not supportable in XPE, then by Theorem 5.3, one of its constituent realizable outcomes is not either. By Theorem S-1, 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.

□

S-2 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 S-2.1. We consider two long-run players. There are four possible stage games, shown in Figure S-2. We assume both players use a discount factor of $\delta = 1/10$.

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 S-2: Stationary example of no universal penal code with two long-run players.

We will use the term "reward" for the high-payoff action 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, consider as a target the realizable outcome $(aq, ds, gw, fv, fv, fv, \ldots)$. (For brevity, we suppress the list of stage games involved.) It can be supported as follows. If "Nature deviates" by choosing a stage game different from those in the target outcome (and player 1 has not deviated in the past), 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 (note that there are multiple 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; this shows the nonexistence of a universal penal code for this game. If such an $\underline{s}$ did exist, it 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.

$\triangle$

Example S-2.2. We assume only one long-run player but no public randomization. We assume $\mathcal{G}$ consists of five stage games as shown in Figure S-3. The discount factor is again $\delta = 1/10$. 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 3(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 S-3: Stationary example of no universal penal code without public randomization.

As in Example S-2.1, we will refer to v, w, x, y, z as "reward" actions and a, e, f, h, j as "punishment" actions.

There exists an XPE that supports the realizable outcome $(c, i, j, j, j, \ldots)$. In particular, specify that if Nature deviates, then reward actions are played from then onward; if player 1 deviates from c in the first period, then punishment actions are played subsequently. All other deviations can be ignored since there are no short-run gains. Refer to this XPE as s_c.

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 punishment actions in all subsequent periods. Refer to this XPE as s_d.

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 play s_d henceforward, and otherwise we simply play punishment actions in every period. It is straightforward to check that this deters the deviation to b in every possible environment (once again, there are several cases to check depending when the environment first differs from the target 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 play s_c henceforward, and otherwise play punishment actions 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, and thus no universal penal code. 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.

$\triangle$

References

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

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

Krasikov, Ilia and Rohit Lamba (2023) "Uncertain Repeated Games," Unpublished manuscript, Pennsylvania State University.

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