Replacement and Reputation

Navin Kartik Elliot Lipnowksi Harry Pei

Motivation

Primary instrument to control elected officials: replacement

Replacement serves two roles:

1. Provide incentives
2. Selection

Other settings: organizational leaders, bureaucrats, etc.

Motivation

Primary instrument to control elected officials: replacement

Replacement serves two roles:

1. Provide incentives
2. Selection

Other settings: organizational leaders, bureaucrats, etc.

Our question: Does replacement lead to good outcomes in the long run?

This paper

Simple model of accountability with moral hazard and adverse selection

- Reputational model: long-lived politicians; short-lived voters
- Replacement is voters' only tool

This paper

Simple model of accountability with moral hazard and adverse selection

- Reputational model: long-lived politicians; short-lived voters
- Replacement is voters' only tool

Results

- There is always an eqm with good outcomes in the long run
- Under some conditions, $\exists$ equilibria with good outcomes in every period

This paper

Simple model of accountability with moral hazard and adverse selection

- Reputational model: long-lived politicians; short-lived voters
- Replacement is voters' only tool

Results

- There is always an eqm with good outcomes in the long run
- Under some conditions, $\exists$ equilibria with good outcomes in every period
- But those very conditions also characterize when
  $\exists$ equilibria with bad outcomes even in the long run

This paper

Simple model of accountability with moral hazard and adverse selection

- Reputational model: long-lived politicians; short-lived voters
- Replacement is voters' only tool

Results

- There is always an eqm with good outcomes in the long run
- Under some conditions, $\exists$ equilibria with good outcomes in every period
- But those very conditions also characterize when
  $\exists$ equilibria with bad outcomes even in the long run

Takeaways

- Mechanism for and robustness of good long run outcomes varies
- In "incentive-sufficient" environs, can discipline moral hazard, but norms matter
  $\rightarrow$ despite reputational forces, poor long-run outcomes possible
  $\rightarrow$ novel rationale
- In "incentive-deficient" environs, discipline fails, but selection prevails

Related Literature

Political accountability with selection

- Banks & Sundaram (1993); Fearon (1999); Myerson (2006); Anesi & Buisseret (2022)

Reputation with imperfect monitoring

- Fudenberg & Levine (1992); Cripps, Mailath, Samuelson (2004)

Related Literature

Political accountability with selection

- Banks & Sundaram (1993); Fearon (1999); Myerson (2006); Anesi & Buisseret (2022)

Reputation with imperfect monitoring

- Fudenberg & Levine (1992); Cripps, Mailath, Samuelson (2004)

Reputation with exogenous replacement

- Mailath & Samuelson (2001); Tadelis (2002); Ekmekci, Gossner, Wilson (2012)

Reputation with competition between long-lived players

- Hörner (2002); Atakan & Ekmekci (2015); Deb & Fanning (2025)

Model

Model

Time: $t = 0, 1, \ldots$

Pool of infinitely many identical long-lived politicians

Sequence of short-lived voters, one in each period

Model

Time: $t = 0, 1, \ldots$

Pool of infinitely many identical long-lived politicians

Sequence of short-lived voters, one in each period

In period 0, one politician is exogenously the incumbent

Incumbent chooses effort $a_0 \in \{0, 1\}$ and generates public signal $s_0 \sim f(\cdot|a_0) \in \Delta S$,

- Finite S; $f(\cdot|0) \neq f(\cdot|1)$ and both pmfs have full support

Model

Time: $t = 0, 1, \ldots$

Pool of infinitely many identical long-lived politicians

Sequence of short-lived voters, one in each period

In period 0, one politician is exogenously the incumbent

Incumbent chooses effort $a_0 \in \{0, 1\}$ and generates public signal $s_0 \sim f(\cdot|a_0) \in \Delta S$,

- Finite S; $f(\cdot|0) \neq f(\cdot|1)$ and both pmfs have full support

Period $t \geq 1$:

- Current voter decides whether to replace the incumbent with random new draw from pool
- Incumbent (old or new) chooses $a_t \in \{0, 1\}$ and generates public signal $s_t \sim f(\cdot|a_t)$

Model

Time: $t = 0, 1, \ldots$

Pool of infinitely many identical long-lived politicians

Sequence of short-lived voters, one in each period

In period 0, one politician is exogenously the incumbent

Incumbent chooses effort $a_0 \in \{0, 1\}$ and generates public signal $s_0 \sim f(\cdot|a_0) \in \Delta S$,

- Finite S; $f(\cdot|0) \neq f(\cdot|1)$ and both pmfs have full support

Period $t \geq 1$:

- Current voter decides whether to replace the incumbent with random new draw from pool
- Incumbent (old or new) chooses $a_t \in \{0, 1\}$ and generates public signal $s_t \sim f(\cdot|a_t)$

Replacements are public; once a politician is replaced, he never returns

Payoffs and Types

Voter t's payoff is a_t; just wants incumbent effort in period t

- Paper also allows for replacement cost (not too large)

Payoffs and Types

Voter t's payoff is a_t; just wants incumbent effort in period t

- Paper also allows for replacement cost (not too large)

Each politician is either good or opportunistic

Politicians' types are private info and iid; each is good with prob $\pi_0 \in (0, 1)$

Payoffs and Types

Voter t's payoff is a_t; just wants incumbent effort in period t

- Paper also allows for replacement cost (not too large)

Each politician is either good or opportunistic

Politicians' types are private info and iid; each is good with prob $\pi_0 \in (0, 1)$

- Good type exerts effort (action 1) whenever he is in office

Payoffs and Types

Voter t's payoff is a_t; just wants incumbent effort in period t

- Paper also allows for replacement cost (not too large)

Each politician is either good or opportunistic

Politicians' types are private info and iid; each is good with prob $\pi_0 \in (0,1)$

- Good type exerts effort (action 1) whenever he is in office
- Opportunistic type's stage payoff is

$$u_s = \begin{cases} 0 & \text{if not in office in period } s \\ u(a_s) & \text{if in office in period } s, \end{cases}$$

where $u(0) \equiv 1 > u(1) \equiv 1 - \kappa > 0$. ($\kappa$ is effort cost)

Discount factor δ

Solution Concept

Equilibrium: symmetric weak PBE, i.e., weak PBE in which

- All opportunistic politicians use the same strategy:
  sequence of own signals $\rightarrow$ prob of effort
- All voters use same strategy:
  sequence of incumbent's own signals $\rightarrow$ prob of replacement

Solution Concept

Equilibrium: symmetric weak PBE, i.e., weak PBE in which

- All opportunistic politicians use the same strategy:
  sequence of own signals $\rightarrow$ prob of effort
- All voters use same strategy:
  sequence of incumbent's own signals $\rightarrow$ prob of replacement

Symmetry rules out

- Coordination on previous office-holders' signals / when an incumbent took office
- Different politicians using different strategies

$\rightarrow$ replacement simply starts interaction afresh with new (ex-ante identical) office-holder

Solution Concept

Equilibrium: symmetric weak PBE, i.e., weak PBE in which

- All opportunistic politicians use the same strategy:
  sequence of own signals $\rightarrow$ prob of effort
- All voters use same strategy:
  sequence of incumbent's own signals $\rightarrow$ prob of replacement

Symmetry rules out

- Coordination on previous office-holders' signals / when an incumbent took office
- Different politicians using different strategies

$\rightarrow$ replacement simply starts interaction afresh with new (ex-ante identical) office-holder

Weak PBE, rather than PBE

- Strengthens main direction of result
  $\rightarrow$ all eqa are "good"
- But off-path flexibility plays role in other direction
  $\rightarrow$ construction of a "bad" eqm

Solution Concept

Equilibrium: symmetric weak PBE, i.e., weak PBE in which

- All opportunistic politicians use the same strategy:
  sequence of own signals $\rightarrow$ prob of effort
- All voters use same strategy:
  sequence of incumbent's own signals $\rightarrow$ prob of replacement

Symmetry rules out

- Coordination on previous office-holders' signals / when an incumbent took office
- Different politicians using different strategies

$\rightarrow$ replacement simply starts interaction afresh with new (ex-ante identical) office-holder

Weak PBE, rather than PBE

- Strengthens main direction of result
  $\rightarrow$ all eqa are "good"
- But off-path flexibility plays role in other direction
  $\rightarrow$ construction of a "bad" eqm

Proposition

A symmetric PBE exists.

Main Result

(Eventual) Full Effort

Definition

An equilibrium attains full effort if $\mathbb{P}(a_t = 1) = 1$ for all $t \geq 0$.

(Eventual) Full Effort

Definition

An equilibrium attains full effort if $\mathbb{P}(a_t = 1) = 1$ for all $t \geq 0$.

Definition

An equilibrium attains eventual full effort if a.s. there is some τ s.t. $a_t = 1$ for all $t \geq \tau$.

(Eventual) Full Effort

Definition

An equilibrium attains full effort if $\mathbb{P}(a_t = 1) = 1$ for all $t \geq 0$.

Definition

An equilibrium attains eventual full effort if a.s. there is some τ s.t. $a_t = 1$ for all $t \geq \tau$.

Note: negative result also under weaker notions of eventual full effort

Equivalence Theorem

If an eqm attains FE, it has no learning; so FE must also be attainable absent good type

Intuitively, this requires conjunction of

low-enough effort cost; sufficiently-informative monitoring; enough patience

Formally on next slide → Condition FE-I

Equivalence Theorem

If an eqm attains FE, it has no learning; so FE must also be attainable absent good type

Intuitively, this requires conjunction of

low-enough effort cost; sufficiently-informative monitoring; enough patience

Formally on next slide → Condition FE-I

Theorem

The following are equivalent:

1. $\{\kappa, f, \delta\}$ violates Condition FE-I
2. No equilibrium attains FE

Equivalence Theorem

If an eqm attains FE, it has no learning; so FE must also be attainable absent good type

Intuitively, this requires conjunction of

low-enough effort cost; sufficiently-informative monitoring; enough patience

Formally on next slide → Condition FE-I

Theorem

The following are equivalent:

1. $\{\kappa, f, \delta\}$ violates Condition FE-I
2. No equilibrium attains FE
3. All equilibria attain Eventual FE
4. $\mathbb{P}(a_0 = 1)$ bounded away from 1 across eqa

Equivalence Theorem

If an eqm attains FE, it has no learning; so FE must also be attainable absent good type

Intuitively, this requires conjunction of

low-enough effort cost; sufficiently-informative monitoring; enough patience

Formally on next slide → Condition FE-I

Theorem

The following are equivalent:

1. $\{\kappa, f, \delta\}$ violates Condition FE-I
2. No equilibrium attains FE
3. All equilibria attain Eventual FE
4. $\mathbb{P}(a_0 = 1)$ bounded away from 1 across eqa

So, always some eqm that attains EFE

But there is a tension between

- Eventual FE in all eqa
- FE (or even FE in first period) in some eqm

Condition FE-I

Condition FE-I

There is a vector $(v(s))_{s\in S} \in [0,1]^S$ such that

$$(1-\delta)(1-\kappa) + \delta \sum_{s\in S} f(s|1)v(s) \geq (1-\delta)1 + \delta \sum_{s\in S} f(s|0)v(s) \qquad (\mathsf{IC_{FE}})$$

and

$$(1-\delta)(1-\kappa) + \delta \sum_{s\in S} f(s|1)v(s) \geq \max_{s\in S} v(s). \qquad (\mathsf{PK_{FE}})$$

Condition FE-I

Condition FE-I

There is a vector $(v(s))_{s\in S} \in [0,1]^S$ such that

$$(1-\delta)(1-\kappa) + \delta \sum_{s\in S} f(s|1)v(s) \geq (1-\delta)1 + \delta \sum_{s\in S} f(s|0)v(s) \qquad (\mathsf{IC_{FE}})$$

and

$$(1-\delta)(1-\kappa) + \delta \sum_{s\in S} f(s|1)v(s) \geq \max_{s\in S} v(s). \qquad (\mathsf{PK_{FE}})$$

Nec and suff for FE eqm

Holds if and only if

- κ is small enough; *and*
- f is sufficiently informative; *and*
- δ is large enough

What Prevents Good Long-Run Outcomes?

Theorem says some eqa do not attain Eventual FE when FE-I holds

What causes the failure of Eventual FE in those equilibria?

What Prevents Good Long-Run Outcomes?

Theorem says some eqa do not attain Eventual FE when FE-I holds

What causes the failure of Eventual FE in those equilibria?

Let $\pi(h^t)$ denote the incumbent's reputation at history h^t. (Probability of good type.)

Proposition

In any eqm that does not attain Eventual FE,

$\exists$ pos-prob history h^t at which
$\pi(h^t) > \pi_0$ and yet incumbent is replaced with pos prob.

$\Longrightarrow$ replacement despite positive track record
($\because$ correctly expect lower effort than from new officeholder)

What Prevents Good Long-Run Outcomes?

Theorem says some eqa do not attain Eventual FE when FE-I holds

What causes the failure of Eventual FE in those equilibria?

Let $\pi(h^t)$ denote the incumbent's reputation at history h^t. (Probability of good type.)

Proposition

In any eqm that does not attain Eventual FE,

$\exists$ pos-prob history h^t at which

$\pi(h^t) > \pi_0$ and yet incumbent is replaced with pos prob.

$\Longrightarrow$ replacement despite positive track record

($\because$ correctly expect lower effort than from new officeholder)

Contrast with failure of Eventual FE in Myerson (2006)

- There, replacement cost is too high and there is never replacement

(EFEproof)

What Assures Good Long-Run Outcomes?

Theorem says all eqa attain Eventual FE when FE-I fails

It turns out this obtains via eventual selection

What Assures Good Long-Run Outcomes?

Theorem says all eqa attain Eventual FE when FE-I fails

It turns out this obtains via eventual selection

Let π_t denote probability that period-t incumbent is a good type

Proposition

In every eqm:

1. Every opportunistic type is eventually replaced with probability 1.
2. If FE-I fails, every good type is retained forever with positive probability.

Hence, if FE-I fails, then with probability 1, some good type is never replaced.

What Assures Good Long-Run Outcomes?

Theorem says all eqa attain Eventual FE when FE-I fails

It turns out this obtains via eventual selection

Let π_t denote probability that period-t incumbent is a good type

Proposition

In every eqm:

1. Every opportunistic type is eventually replaced with probability 1.
2. If FE-I fails, every good type is retained forever with positive probability.

Hence, if FE-I fails, then with probability 1, some good type is never replaced.

Logic for the first part is related to CMS (2004), "impermanent reputations"

(EFEproof)

Proofs Snippets

Plan

Theorem

The following are equivalent:

1. $\{\kappa, f, \delta\}$ violates Condition FE-I
2. No equilibrium attains FE
3. All equilibria attain Eventual FE
4. $\mathbb{P}(a_0 = 1)$ bounded away from 1 across eqa

Plan

Theorem

The following are equivalent:

1. $\{\kappa, f, \delta\}$ violates Condition FE-I
2. No equilibrium attains FE
3. All equilibria attain Eventual FE
4. $\mathbb{P}(a_0 = 1)$ bounded away from 1 across eqa

I will discuss:

- not (1) $\implies$ not (3)
- (1) $\implies$ (3)

Omit:

- (1) $\implies$ (2)
- (2) [hence (4)] $\implies$ (1)
- (1) $\implies$ (4)

Equilibria that Fail Eventual FE

Why is it not easy to construct eqa that fail Eventual FE?

Naive attempt:

- opportunistic incumbent never works
- voter always replaces (no replacement cost)

Equilibria that Fail Eventual FE

Why is it not easy to construct eqa that fail Eventual FE?

Naive attempt:

- opportunistic incumbent never works
- voter always replaces (no replacement cost)

Problem:

- not optimal for voter after "good" signals $\rightarrow$ learns too much

Equilibria that Fail Eventual FE

Why is it not easy to construct eqa that fail Eventual FE?

Naive attempt:

- opportunistic incumbent never works
- voter always replaces (no replacement cost)

Problem:

- not optimal for voter after "good" signals $\rightarrow$ learns too much

Solution when Condition FE-I holds: An eqm in which

- there isn't too much learning
- partial shirking in first period
- every incumbent gets replaced with prob 1

Hence Eventual FE violated

▸ Construction

FE-I Fails $\Longrightarrow$ Eventual FE in All Equilibria

- When Condition FE-I fails, cannot incentivize full effort—in particular, not in first period $\Longrightarrow$ there has to be some learning

FE-I Fails $\implies$ Eventual FE in All Equilibria

- When Condition FE-I fails, cannot incentivize full effort—in particular, not in first period $\implies$ there has to be some learning
- An opportunistic type gets replaced with prob 1 (build on CMS-style argument)
  - Martingale + replacement only incentive tool

FE-I Fails $\implies$ Eventual FE in All Equilibria

- When Condition FE-I fails, cannot incentivize full effort—in particular, not in first period $\implies$ there has to be some learning
- An opportunistic type gets replaced with prob 1 (build on CMS-style argument)
  - Martingale + replacement only incentive tool
- If belief is high enough, incumbent will not be replaced

FE-I Fails $\implies$ Eventual FE in All Equilibria

- When Condition FE-I fails, cannot incentivize full effort—in particular, not in first period $\implies$ there has to be some learning
- An opportunistic type gets replaced with prob 1 (build on CMS-style argument)
  - Martingale + replacement only incentive tool
- If belief is high enough, incumbent will not be replaced
- $\forall \varepsilon > 0$, there is a pos prob that a new incumbent will get to belief above $1 - \varepsilon$
  - Loose intuition: If $\sup$[on-path beliefs] < 1, then incumbent would have to be working when near that $\sup$, contrary to FE-I not holding
  - Much of the work in the proof is here

FE-I Fails $\implies$ Eventual FE in All Equilibria

- When Condition FE-I fails, cannot incentivize full effort—in particular, not in first period $\implies$ there has to be some learning
- An opportunistic type gets replaced with prob 1 (build on CMS-style argument)
  - Martingale + replacement only incentive tool
- If belief is high enough, incumbent will not be replaced
- $\forall \varepsilon > 0$, there is a pos prob that a new incumbent will get to belief above $1 - \varepsilon$
  - Loose intuition: If $\sup$[on-path beliefs] < 1, then incumbent would have to be working when near that $\sup$, contrary to FE-I not holding
  - Much of the work in the proof is here
- Using Doob's upcrossing ineq, every good type has pos prob of sticking at high-enough beliefs s.t. never replaced
- So eventually some good type is never replaced

FE-I Fails $\implies$ Eventual FE in All Equilibria

- When Condition FE-I fails, cannot incentivize full effort—in particular, not in first period $\implies$ there has to be some learning
- An opportunistic type gets replaced with prob 1 (build on CMS-style argument)
  - Martingale + replacement only incentive tool
- If belief is high enough, incumbent will not be replaced
- $\forall \varepsilon > 0$, there is a pos prob that a new incumbent will get to belief above $1 - \varepsilon$
  - Loose intuition: If $\sup$[on-path beliefs] < 1, then incumbent would have to be working when near that $\sup$, contrary to FE-I not holding
  - Much of the work in the proof is here
- Using Doob's upcrossing ineq, every good type has pos prob of sticking at high-enough beliefs s.t. never replaced
- So eventually some good type is never replaced

These arguments also lead to the two Propositions stated earlier. ▸ TrackRecord ▸ Selection

Conclusion

Conclusion

A model of accountability with moral hazard and adverse selection

- Each politician is either good or opportunistic
- Replacement is instrument to provide incentives and to select good politicians

Takeaways

- Replacement makes long-run good outcomes always possible (in some eqm) and, sometimes guarantees it (in all eqa)
- Mechanism for good long-run outcomes varies
- In incentive-sufficient environs, can discipline moral hazard, but requires good norms
  $\rightarrow$ novel source of long-run inefficiencies: replacement despite good track record
- In incentive-deficient environs, discipline fails, but selection necessarily prevails

Appendix

Condition FE-I Fails $\implies$ $\mathbb{P}(a_0 = 1) < 1$ in every eqm

If $\mathbb{P}(a_0 = 1) = 1$, then the opportunistic type must be incentivized to play $a_0 = 1$

(Back)

Condition FE-I Fails $\implies$ $\mathbb{P}(a_0 = 1) < 1$ in every eqm

If $\mathbb{P}(a_0 = 1) = 1$, then the opportunistic type must be incentivized to play $a_0 = 1$ (Back)

Consider any $\varepsilon > 0$ small enough.

Because FE-I fails, there exists $s_0 \in S$ such that $V(s_0) \geq V(\emptyset) + \varepsilon$.

Condition FE-I Fails $\implies$ $\mathbb{P}(a_0 = 1) < 1$ in every eqm

If $\mathbb{P}(a_0 = 1) = 1$, then the opportunistic type must be incentivized to play $a_0 = 1$ (Back)

Consider any $\varepsilon > 0$ small enough.

Because FE-I fails, there exists $s_0 \in S$ such that $V(s_0) \geq V(\emptyset) + \varepsilon$.

Hence, at history s_0, the incumbent is retained with pos prob

- Voter optimality $\implies$ the opportunistic type must again be incentivized to play $a_1 = 1$
- There exists $s_1 \in S$ such that $V(s_0, s_1) \geq V(s_0) + \varepsilon$.

Condition FE-I Fails $\implies$ $\mathbb{P}(a_0 = 1) < 1$ in every eqm

If $\mathbb{P}(a_0 = 1) = 1$, then the opportunistic type must be incentivized to play $a_0 = 1$ (Back)

Consider any $\varepsilon > 0$ small enough.

Because FE-I fails, there exists $s_0 \in S$ such that $V(s_0) \geq V(\emptyset) + \varepsilon$.

Hence, at history s_0, the incumbent is retained with pos prob

- Voter optimality $\implies$ the opportunistic type must again be incentivized to play $a_1 = 1$
- There exists $s_1 \in S$ such that $V(s_0, s_1) \geq V(s_0) + \varepsilon$.

Iterating, for any $t \geq 0$, there is a pos-prob history $(s_0, \ldots, s_t)$ s.t. $V(s_0, \ldots, s_t) \geq V(\emptyset) + t\varepsilon$

Contradiction, as $V(\cdot) \leq 1$

FE-I $\implies$ Equilibria that Attain FE

Consider the following strategy profile

- Incumbent always exerts effort on the eqm path.
- Voter t replaces the incumbent if and only signal s_{t-1} has a LR below some threshold.
- Off path, voter never retains incumbent and incumbent always shirks.

FE-I $\implies$ Equilibria that Attain FE

Consider the following strategy profile

- Incumbent always exerts effort on the eqm path.
- Voter t replaces the incumbent if and only signal s_{t-1} has a LR below some threshold.
- Off path, voter never retains incumbent and incumbent always shirks.

We show that FE-I $\implies$ there is a threshold s.t. this profile is an eqm.

- Only thing to check is incumbent's incentive

Lemma

Assume Condition FE-I. There exist $\overline{v} > 0$ and $x \in (0, 1]$ such that (IC_{FE}) holds and

$$(1-\delta)(1-\kappa) + \delta \sum_{s \in S} f(s|1) v(s) = \overline{v},$$

with $v(s) = \overline{v} \cdot \mathbf{1}\left\{ \frac{f(s|1)}{f(s|0)} \geq x \right\}$

Incentive Lemma when FE-I Fails

Lemma

If Condition FE-I fails, then $\exists \varepsilon > 0$ s.t. $\forall \left(v(s)\right)_{s \in S} \in [0,1]^S$ that satisfies ($\mathsf{IC_{FE}}$),

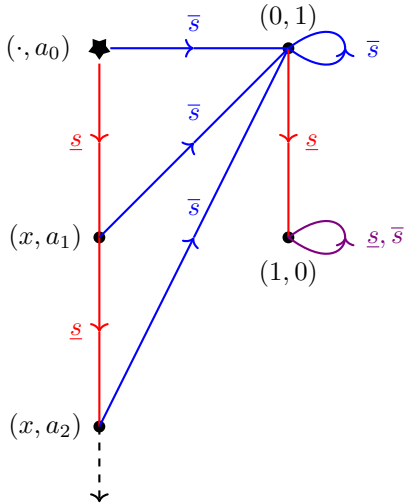
$$\max_{s \in S} v(s) \geq (1-\delta)(1-\kappa) + \delta \sum_{s \in S} f(s|1) v(s) + \varepsilon.$$

FE-I $\implies$ Equilibria that Fail Eventual FE

Construction with two signals, $S = \{\underline{s}, \overline{s}\}$.

FE-I $\implies$ Equilibria that Fail Eventual FE

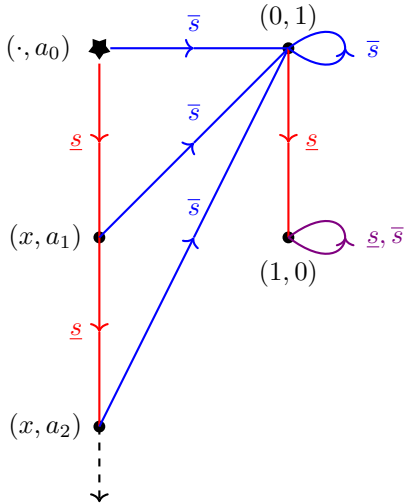
Construction with two signals, $S = \{\underline{s}, \overline{s}\}$. The first number in the vector is the prob of replacement, and the second number is the prob of choosing $a = 1$. It holds that $0 < a_0 < a_1 < \cdots < 1$ and $x \in (0, 1]$.



(Back)

FE-I $\implies$ Equilibria that Fail Eventual FE

Construction with two signals, $S = \{\underline{s}, \overline{s}\}$. The first number in the vector is the prob of replacement, and the second number is the prob of choosing $a = 1$. It holds that $0 < a_0 < a_1 < \cdots < 1$ and $x \in (0, 1]$.



(Back)

Caveat:
use weak PBE to sustain $(1,0)$ profiles