Costly Signaling: rationality and evolution

Josef Hofbauer
University of Vienna

Christina Pawlowitsch
Université Paris–Panthéon–Assas

2nd Workshop on Dynamic Games in Paris
October 26–27, 2023

50 years on: Michael Spence, “Job Market Signaling” (1973)

“The term ‘market signaling’ is not exactly a part of the well-defined, technical vocabulary of the economist ... In fact, it is part of my purpose to outline a model in which signaling is implicitly defined and to explain why one can, and perhaps should, be interested in it.”

Today ‘market signaling’ is part of the well-defined, technical vocabulary of the economist – thanks to Spence!

Dynamics in costly-signaling games: relatively unexplored

- Spence (1973) appeals to a "dynamic" story as a foundation of his analysis (not fully closed in a game-theoretic sense; abstracts from mixed equilibria; elements of partial equilibrium analysis)
- Nöldeke and Samuelson (1997): study in more detail Spence's dynamic model and introduce perturbations
- Wagner (2013): replicator dynamics in "truncated" version of Spence's model
- Zollman, Bergstrom, and Hutteger (2013): replicator dynamics in discrete version of Spence's model (limited to certain parameter constellations; do not study global convergence

Costly-signaling theory: wide range of applications

Miller and Rock (1985): dividend payments as a costly signal

Milgrom and Roberts (1986): advertising as a costly signal

Zahavi (1975): "The Handicap Principle." Grafen (1990): formal model

Caro (1986): costly signals in predator–prey interaction

Archetti (2008): costly signals in parasite-host interaction

Bliege Bird and Smith: inefficient foraging strategies, gift-giving, communal sharing as costly signals

Van Rooy (2003): "Politeness is a Handicap"

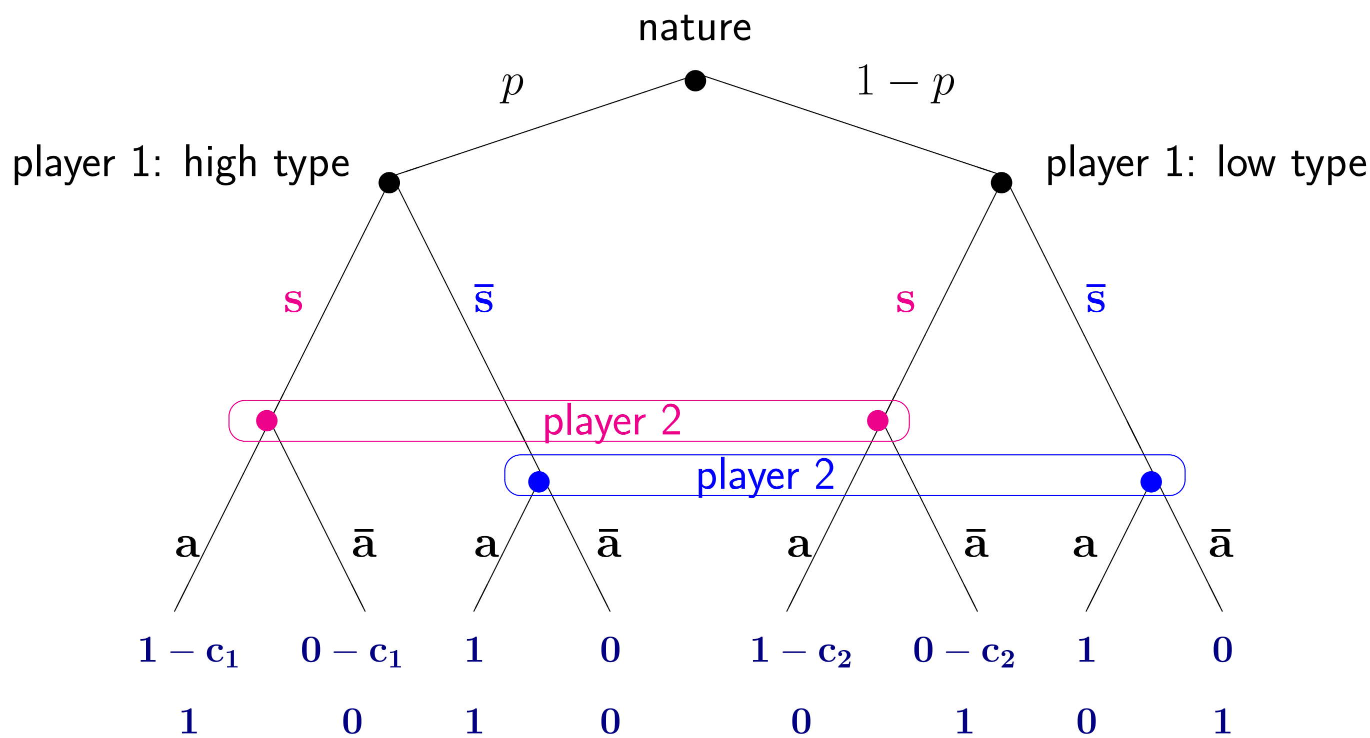
... Veblen (1899), *Theory of the Leisure Class*, Mauss (1924): "The Gift: Forms and Functions of Exchange in Archaic Societies"

Approach taken here:

- Minimal, discrete model: 2 states of nature (high and low), 2 signals (costly signal or not), 2 actions (accept or not). Two classes:
  - (I) production of the costly signal is of different costs for different types (as in Spence 1973)
  - (II) production of the costly signal is of the same cost for different types, but types have different benefits if the signal has the desired effect (as in models of advertising)

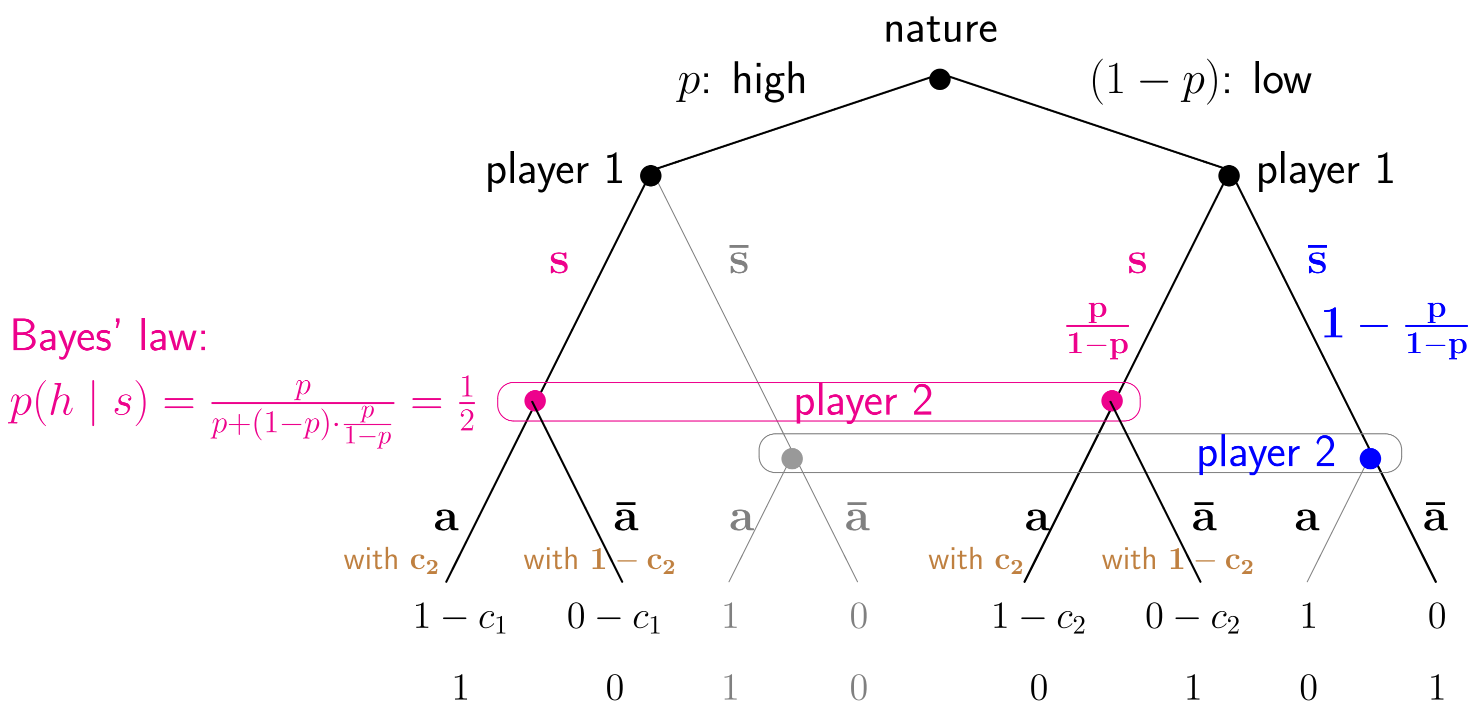
  Further classification:
  - signaling costs in relation to relative rewards for different types (3 paradigmatic cases)
  - prior belief (3 relevant cases)
- Equilibrium refinement:
  - index
  - evolutionary dynamics: replicator dynamics and BR dynamics
  - classical refinements (restrictions on beliefs off the equilibrium path): "never-a-weak-best-response," "divinity," "intuitive" criterion.

Class I: different costs in producing the signal



	aa	**aā**	**āa**	**āā**
ss	$1 - pc_1 - (1-p)c_2,\ p$	$1 - pc_1 - (1-p)c_2,\ p$	$-pc_1 - (1-p)c_2,\ 1-p$	$-pc_1 - (1-p)c_2,\ 1-p$
ss̄	$1 - pc_1,\ p$	$p(1-c_1),\ 1$	$-pc_1 + (1-p),\ 0$	$-pc_1,\ 1-p$
s̄s	$1-(1-p)c_2,\ p$	$(1-p)(1-c_2),\ 0$	$p-(1-p)c_2,\ 1$	$-(1-p)c_2,\ 1-p$
s̄s̄	$1,\ p$	$0,\ 1-p$	$1,\ p$	$0,\ 1-p$

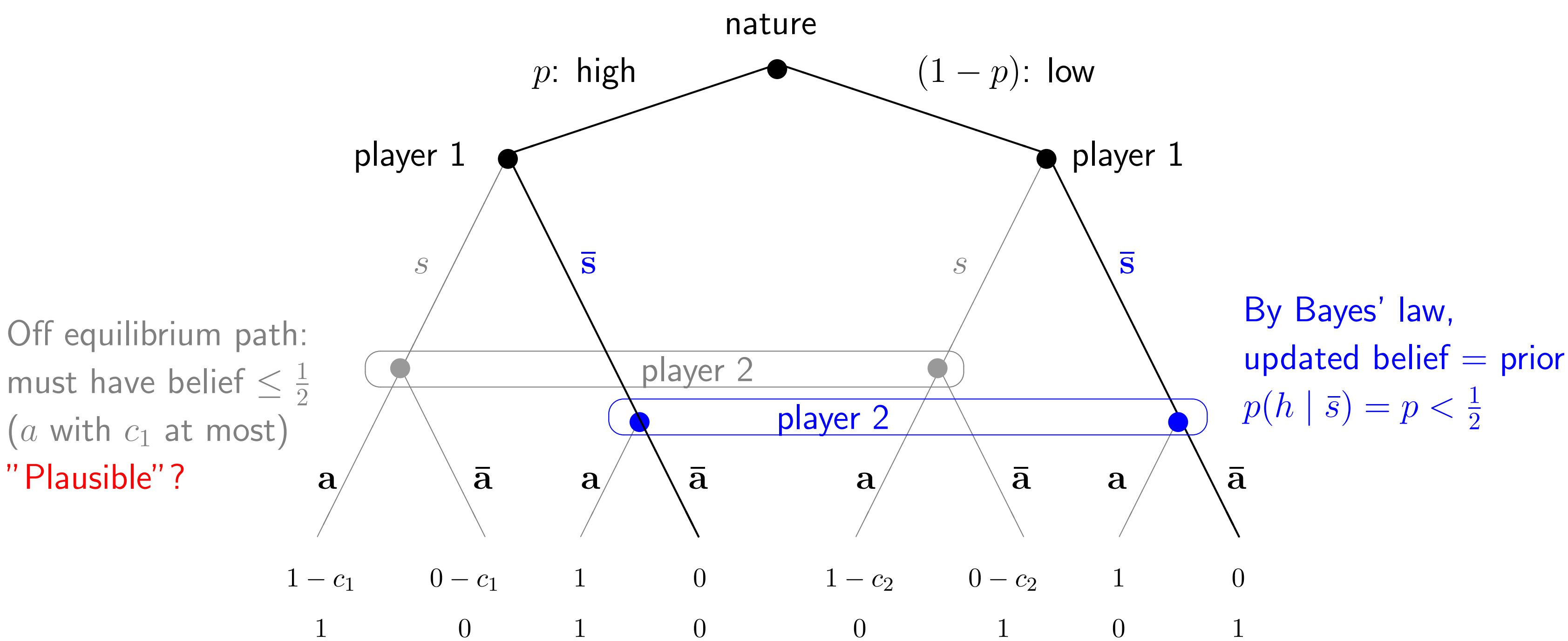
Class I, $0 \le c_1 < c_2 < 1$, $p < 1/2$: E1 partially revealing equilibrium



		$\mathbf{aa}$	$\overset{c_2}{\downarrow}$ $\mathbf{a\bar{a}}$	$\mathbf{\bar{a}a}$	$\overset{1-c_2}{\downarrow}$ $\mathbf{\bar{a}\bar{a}}$
$\frac{p}{1-p} \to$	$\mathbf{ss}$	$1 - pc_1 - (1-p)c_2,\ p$	$\mathbf{1 - pc_1 - (1-p)c_2},\ p$	$-pc_1 - (1-p)c_2,\ \mathbf{1-p}$	$-pc_1 - (1-p)c_2,\ \mathbf{1-p}$
$1 - \frac{p}{1-p} \to$	$\mathbf{s\bar{s}}$	$1 - pc_1,\ p$	$p(1-c_1),\ \mathbf{1}$	$-pc_1 + (1-p),\ 0$	$-pc_1,\ 1-p$
	$\mathbf{\bar{s}s}$	$1 - (1-p)c_2,\ p$	$(1-p)(1-c_2),\ 0$	$p - (1-p)c_2,\ \mathbf{1}$	$-(1-p)c_2,\ 1-p$
	$\mathbf{\bar{s}\bar{s}}$	$\mathbf{1},\ p$	$0,\ \mathbf{1-p}$	$\mathbf{1},\ p$	$\mathbf{0},\ \mathbf{1-p}$

- E1: 1 mixes between ss and $s\bar{s}$ with $\frac{p}{1-p}$ on first; 2 between $a\bar{a}$ and $\bar{a}\bar{a}$, with c_2 on first.

Class I, $0 \leq c_1 < c_2 < 1$, $p < 1/2$: P1 "no-signaling" equilibrium outcome



	$\mathbf{aa}$	with $y \in [0, c_1] \rightarrow \mathbf{a\bar{a}}$	$\mathbf{\bar{a}a}$	with $1-y \rightarrow \mathbf{\bar{a}\bar{a}}$
$\mathbf{ss}$	$1 - pc_1 - (1-p)c_2,\ p$	$\mathbf{1 - pc_1 - (1-p)c_2},\ p$	$-pc_1 - (1-p)c_2,\ \mathbf{1-p}$	$-pc_1 - (1-p)c_2,\ \mathbf{1-p}$
$\mathbf{s\bar{s}}$	$1 - pc_1,\ p$	$p(1-c_1),\ \mathbf{1}$	$-pc_1 + (1-p),\ 0$	$-pc_1,\ 1-p$
$\mathbf{\bar{s}s}$	$1 - (1-p)c_2,\ p$	$(1-p)(1-c_2),\ 0$	$p - (1-p)c_2,\ \mathbf{1}$	$-(1-p)c_2,\ 1-p$
$\mathbf{\bar{s}\bar{s}}$	$\mathbf{1},\ p$	$0,\ \mathbf{1-p}$	$\mathbf{1},\ p$	$\mathbf{0},\ \mathbf{1-p}$

- P1: No-signaling: 1 takes $\bar{s}\bar{s}$; 2 mix between $a\bar{a}$ and $\bar{a}\bar{a}$ with $y \in [0, c_1]$ on first.

Equilibrium structure

$p < \frac{1}{2}$:	(E1) : *partially revealing*	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow p^\star = \frac{1}{2}$: $\mathbf{a}$ with c_2
		$\mathbf{l} \longrightarrow \mathbf{s}$ with $\frac{\mathbf{p}}{\mathbf{1-p}}$	$\bar{\mathbf{s}} \longrightarrow$ low for sure : $\bar{\mathbf{a}}$
	(P1): *both use $\bar{s}$*	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathbf{s} \longrightarrow a$ with prob $\leq c_1$
		$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = p < \frac{1}{2}$: $\bar{\mathbf{a}}$
$p > \frac{1}{2}$:	(E2) : *partially revealing*	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$ with $\frac{\mathbf{1-p}}{\mathbf{p}}$	$\mathbf{s} \longrightarrow$ high for sure : $\mathbf{a}$
		$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = \frac{1}{2}$: $\mathbf{a}$ with $1 - c_1$
	(P2): *both use s*	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow p^\star = p > \frac{1}{2}$: $\mathbf{a}$
		$\mathbf{l} \longrightarrow \mathbf{s}$	$\bar{\mathbf{s}} \longrightarrow \mathbf{a}$ with prob $\leq 1 - c_2$
	(P3): *both use $\bar{s}$*	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathbf{s} \longrightarrow \mathbf{a}$ with any prob
		$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = p > \frac{1}{2}$: $\mathbf{a}$
$p = \frac{1}{2}$:	(E1-P2): *both use s*	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow p^\star = p = \frac{1}{2}$: $\mathbf{a}$ with $y \in [c_2, 1]$
		$\mathbf{l} \longrightarrow \mathbf{s}$	$\bar{\mathbf{s}} \longrightarrow \mathbf{a}$ with $y' \in [0, y - c_2]$
	(P1-E2-P3): *both use $\bar{s}$*	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathbf{s} \longrightarrow \mathbf{a}$ with $y \in [0, \min\{y' + c_1, 1\}]$
		$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = p = \frac{1}{2}$: $\mathbf{a}$ with $y' \in [0, 1]$

The index: a rough guide to evolutionary stability

Shapley (1974): Index, $+1$ or -1, to every regular equilibrium

- Strict equilibrium has index $+1$.
- Removing or adding unused strategies does not change the index.
- *Index Theorem*: the sum of the indices of all equilibria is $+1$.

Hofbauer and Sigmund (1988, 1998): index as the sign of the determinant of the negative Jacobian

Ritzberger (1994, 2002): index of an equilibrium component is:

- an integer
- robust under payoff perturbations

Demichelis and Ritzberger (2003):

- If an equilibrium component is asymptotically stable under some evolutionary dynamics, then its index equals its Euler characteristics.

  If it is convex or contractible, then its index is $+1$.

Equilibrium structure

$p < \frac{1}{2}$:	(E1) : *partially revealing*	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow p^\star = \frac{1}{2}$: $\mathbf{a}$ with c_2
	Index: $+1$. FI	$\mathbf{l} \longrightarrow \mathbf{s}$ with $\frac{\mathbf{p}}{\mathbf{1-p}}$	$\bar{\mathbf{s}} \longrightarrow$ low for sure : $\bar{\mathbf{a}}$
	(P1): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathbf{s} \longrightarrow a$ with prob $\leq c_1$
	Index: 0. Not FI	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = p < \frac{1}{2} : \bar{\mathbf{a}}$
$p > \frac{1}{2}$:	(E2) : *partially revealing*	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$ with $\frac{\mathbf{1-p}}{\mathbf{p}}$	$\mathbf{s} \longrightarrow$ high for sure : $\mathbf{a}$
	Index: -1. FI	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = \frac{1}{2}$: $\mathbf{a}$ with $1 - c_1$
	(P2): *both use* s	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow p^\star = p > \frac{1}{2}$: $\mathbf{a}$
	Index: $+1$. FI	$\mathbf{l} \longrightarrow \mathbf{s}$	$\bar{\mathbf{s}} \longrightarrow \mathbf{a}$ with prob $\leq 1 - c_2$
	(P3): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathbf{s} \longrightarrow \mathbf{a}$ with any prob
	Index: $+1$. FI	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = p > \frac{1}{2}$: $\mathbf{a}$
$p = \frac{1}{2}$:	(E1-P2): *both use* s	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow p^\star = p = \frac{1}{2}$: $\mathbf{a}$ with $y \in [c_2, 1]$
	Index: $+1$ FI	$\mathbf{l} \longrightarrow \mathbf{s}$	$\bar{\mathbf{s}} \longrightarrow \mathbf{a}$ with $y' \in [0, y - c_2]$
	(P1-E2-P3): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathbf{s} \longrightarrow \mathbf{a}$ with $y \in [0, \min\{y' + c_1, 1\}]$
	Index: 0. Not all FI	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^\star = p = \frac{1}{2}$: $\mathbf{a}$ with $y' \in [0, 1]$

Evolutionary dynamics in costly-signaling games

The Replicator Dynamics (Taylor and Jonker 1978; Hofbauer, Schuster, and Sigmund 1979)

Game played repeatedly in a large population. Growth rate of a strategy proportional to its fitness-difference relative to the average fitness in the population.

For a two-population game:

$$\begin{aligned}\dot{x}_i &= x_i(u_i^1 - \bar{u}^1), \quad i = 1, \dots n^1,\\ \dot{y}_j &= y_j(u_j^2 - \bar{u}^2), \quad j = 1, \dots n^2,\end{aligned}$$

where u_i^k is the payoff of player k playing strategy i, and $\bar{u}^k$ the average payoff of player k.

The Replicator Dynamics for our game in normal form

Payoffs

$$\begin{aligned} u^1(ss,\mathbf{y}) &= y - pc_1 - (1-p)c_2 \\ u^1(s\bar{s},\mathbf{y}) &= p(y-c_1) + (1-p)y' \\ u^1(\bar{s}s,\mathbf{y}) &= (1-p)(y-c_2) + py' \\ u^1(\bar{s}\bar{s},\mathbf{y}) &= y' \end{aligned} \tag{1}$$

Where $\mathbf{y} = (y(aa), y(a\bar{a}), y(\bar{a}a), y(\bar{a}\bar{a}))$, a mixed strategy of player 2, and

$y = y(aa) + y(a\bar{a})$
$y' = y(aa) + y(\bar{a}a)$

We observe:

$$u^1(ss) + u^1(\bar{s}\bar{s}) = u^1(s\bar{s}) + u^1(\bar{s}s) \tag{2}$$

Similarly:

$$\begin{aligned} u^2(aa, \mathbf{x}) &= p \\ u^2(a\bar{a}, \mathbf{x}) &= px_h + (1-p)(1-x_\ell) \\ u^2(\bar{a}a, \mathbf{x}) &= p(1-x_h) + (1-p)x_\ell \\ u^2(\bar{a}\bar{a}, \mathbf{x}) &= 1-p \end{aligned} \tag{3}$$

$\mathbf{x} = (x(ss), x(s\bar{s}), x(\bar{s}s), x(\bar{s}\bar{s}))$,

$x_h = x(ss) + x(s\bar{s})$,
$x_\ell = x(ss) + x(\bar{s}s)$

And we observe also that:

$$u^2(aa) + u^2(\bar{a}\bar{a}) = 1 = u^2(a\bar{a}) + u^2(\bar{a}a) \tag{4}$$

Eqs. (2) and (4): for any game with the same extensive form.

Gaunersdorfer, Hofbauer, and Sigmund (1991):

If $u_1 + u_4 = u_2 + u_3$, then $\frac{x_1 x_4}{x_2 x_3}$ is a constant of motion for the replicator dynamics $\rightarrow$ foliation of state space $\Delta_4 \times \Delta_4$ into 4-dimensional invariant manifold.

The 'central' invariant manifold, given by $x_1 x_4 = x_2 x_3$, the *Wright manifold*, can be parameterized:

$$\begin{aligned} x_1 &= x x', \\ x_2 &= x(1 - x'), \\ x_3 &= (1 - x) x', \\ x_4 &= (1 - x)(1 - x'), \end{aligned}$$

with $(x, x') \in [0, 1]^2$: $x = x_1 + x_2, x' = x_1 + x_3$.

On this invariant manifold, the replicator dynamics can be written as:

$$\begin{aligned} \dot{x} &= x(1 - x)(u_1 - u_3) \\ \dot{x}' &= x'(1 - x')(u_1 - u_2) \end{aligned} \tag{5}$$

In our game:

On the 'central' invariant manifold:

$$x(ss)x(\bar{s}\bar{s}) = x(s\bar{s})x(\bar{s}s), \quad y(aa)y(\bar{a}\bar{a}) = y(a\bar{a})y(\bar{a}a)$$

with $x_h = x(ss) + x(s\bar{s})$, $x_\ell = x(ss) + x(\bar{s}s)$
and $y = y(aa) + y(a\bar{a}), y' = y(aa) + y(\bar{a}a)$:

$$\begin{aligned}\dot{x}_h &= x_h(1 - x_h)(y - c_1 - y')p \\ \dot{x}_\ell &= x_\ell(1 - x_\ell)[y - c_2 - y'](1 - p) \\ \dot{y} &= y(1 - y)[px_h - (1 - p)x_\ell] \\ \dot{y}' &= y'(1 - y')[p(1 - x_h) - (1 - p)(1 - x_\ell)]\end{aligned} \tag{6}$$

This system of differential equations on the hypercube $[0, 1]^4$ can be derived directly from the extensive form, as the

$\longrightarrow$ *replicator dynamics for behavior strategies*.

Replicator dynamics for behavior strategies

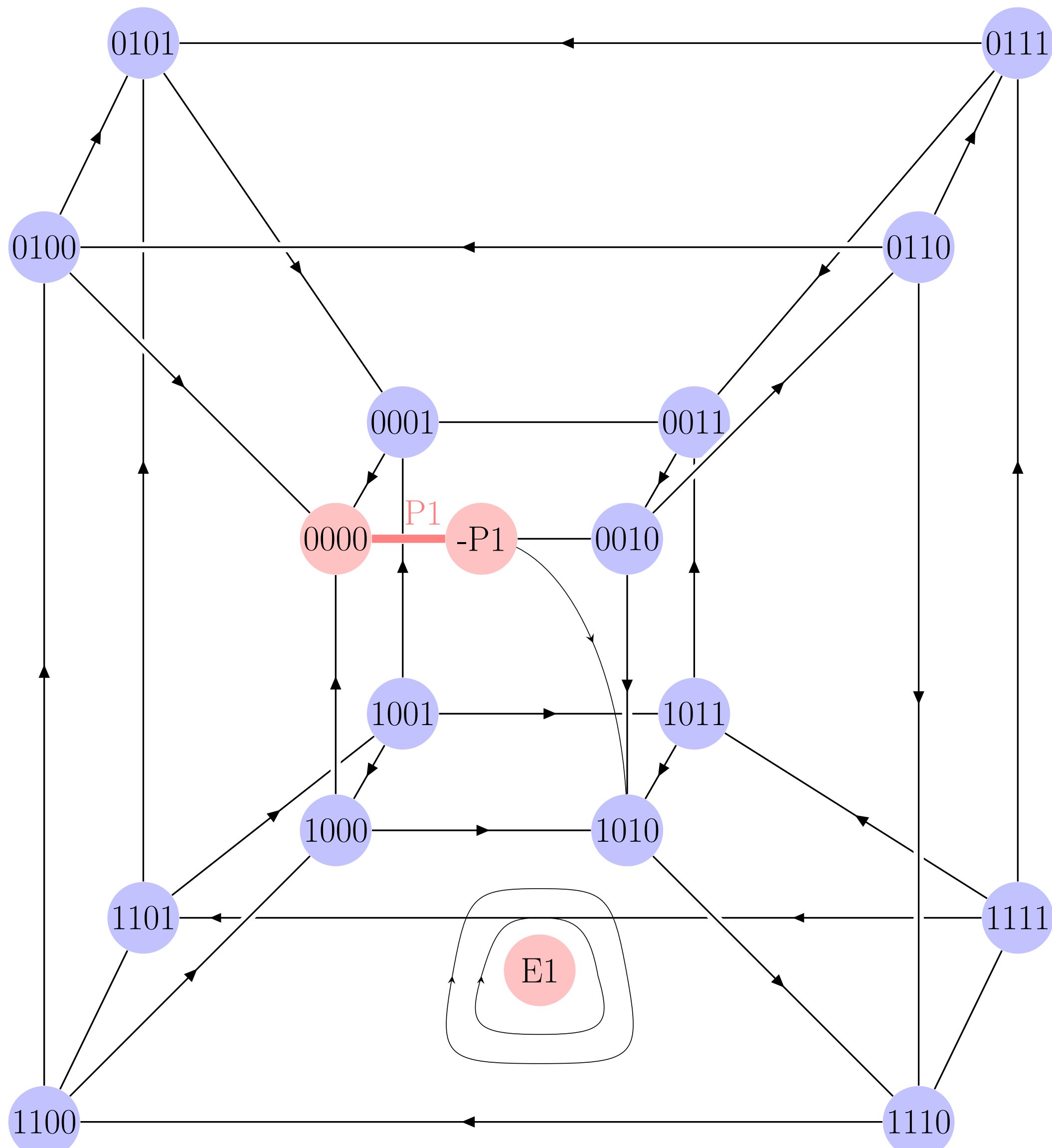
$x_h = \text{prob}(s|\text{high})$, $x_\ell = \text{prob}(s|\text{low})$,
$y = \text{prob}(a|s)$, $y' = \text{prob}(a|\bar{s})$.

State space: (x_h, x_ℓ, y, y') in hypercube $[0,1]^4$

$$\begin{aligned}
\dot{x}_h &= x_h(1 - x_h)(y - c_1 - y')p \\
\dot{x}_\ell &= x_\ell(1 - x_\ell)[y - c_2 - y'](1 - p) \\
\dot{y} &= y(1 - y)[px_h - (1 - p)x_\ell] \\
\dot{y}' &= y'(1 - y')[p(1 - x_h) - (1 - p)(1 - x_\ell)]
\end{aligned} \tag{7}$$

$\mathbf{0 \leq c_1 < c_2 < 1}$

$\mathbf{p < \frac{1}{2}}$



Case: $p < \frac{1}{2}$

Replicator dynamics near the partially revealing E1 $= (1, \frac{p}{1-p}, c_2, 0)$:

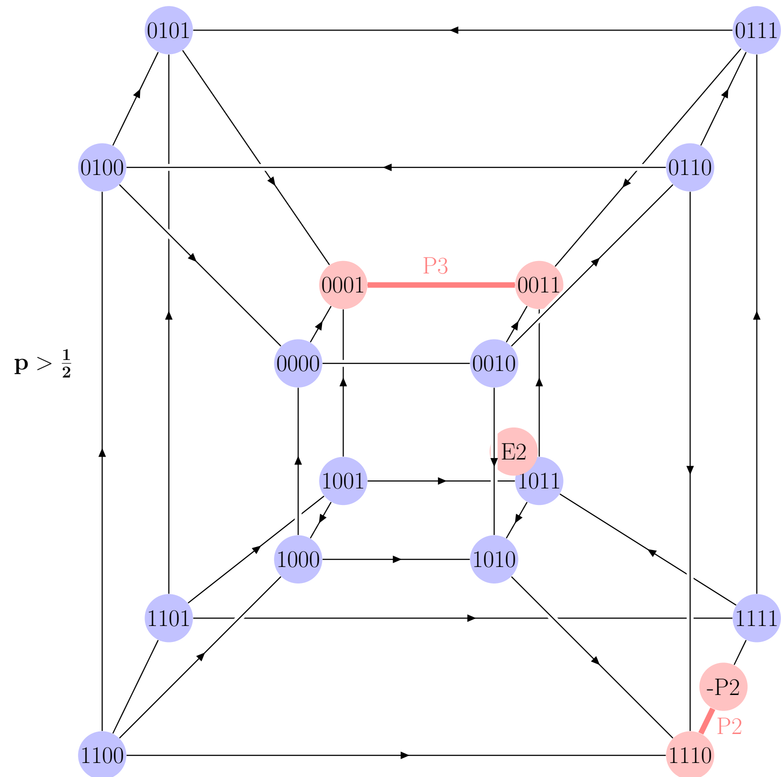
In the supporting boundary face, replicator dynamics for a cyclic 2×2 game, with closed orbits around E1. Each of these periodic solutions attracts a 3-dimensional manifold of solutions $\rightarrow$ Boundary face $x_h = 1, y' = 0$ attracts an open set of initial conditions.

Replicator dynamics near the edge containing P1, $(0, 0, y, 0)$:

The basin of attraction of the whole component P1 contains an open set. The endpoint -P1$=$ $(0, 0, c_1, 0)$ is unstable: one orbit converges to -P1, and one orbit, with -P1 as α–limit, converges to the corner $(1, 0, 1, 0)$. Hence, the component P1 is unstable.

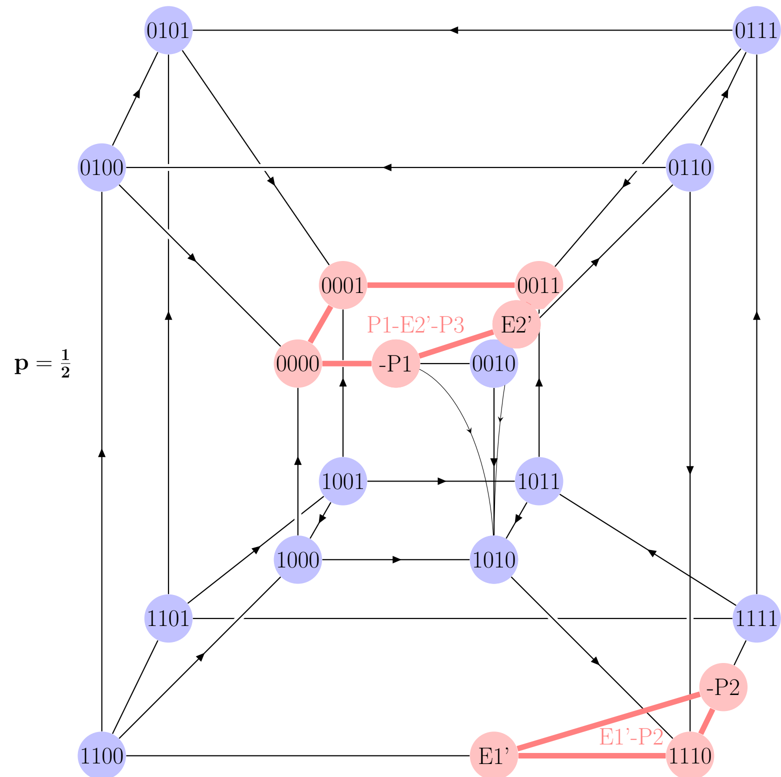
Global convergence: all orbits in the interior converge to the union of the lower front and the inner front boundary face: the high type sends the costly signal or the low type does not; and in no costly signal, 2 never accepts.

(Best-response dynamics: E1 is asymptotically stable; P1 is not. All orbits to one of the Nash equilibria.)



$p > 1/2$: P3 asymptotically stable; P2 stable and interior attracting, but not asymptotically stable. Convergence: every orbit in interior to a Nash equilibrium. E2 is a saddle.

(Best-response dynamics: both P3 and P2 asymptotically stable.)



$p = 1/2$: P1-E2'-P3 is unstable; nevertheless interior attracting.
E1'-P2 is stable and interior attracting but not asymptotically stable.
Convergence: every orbit in interior to a Nash equilibrium.

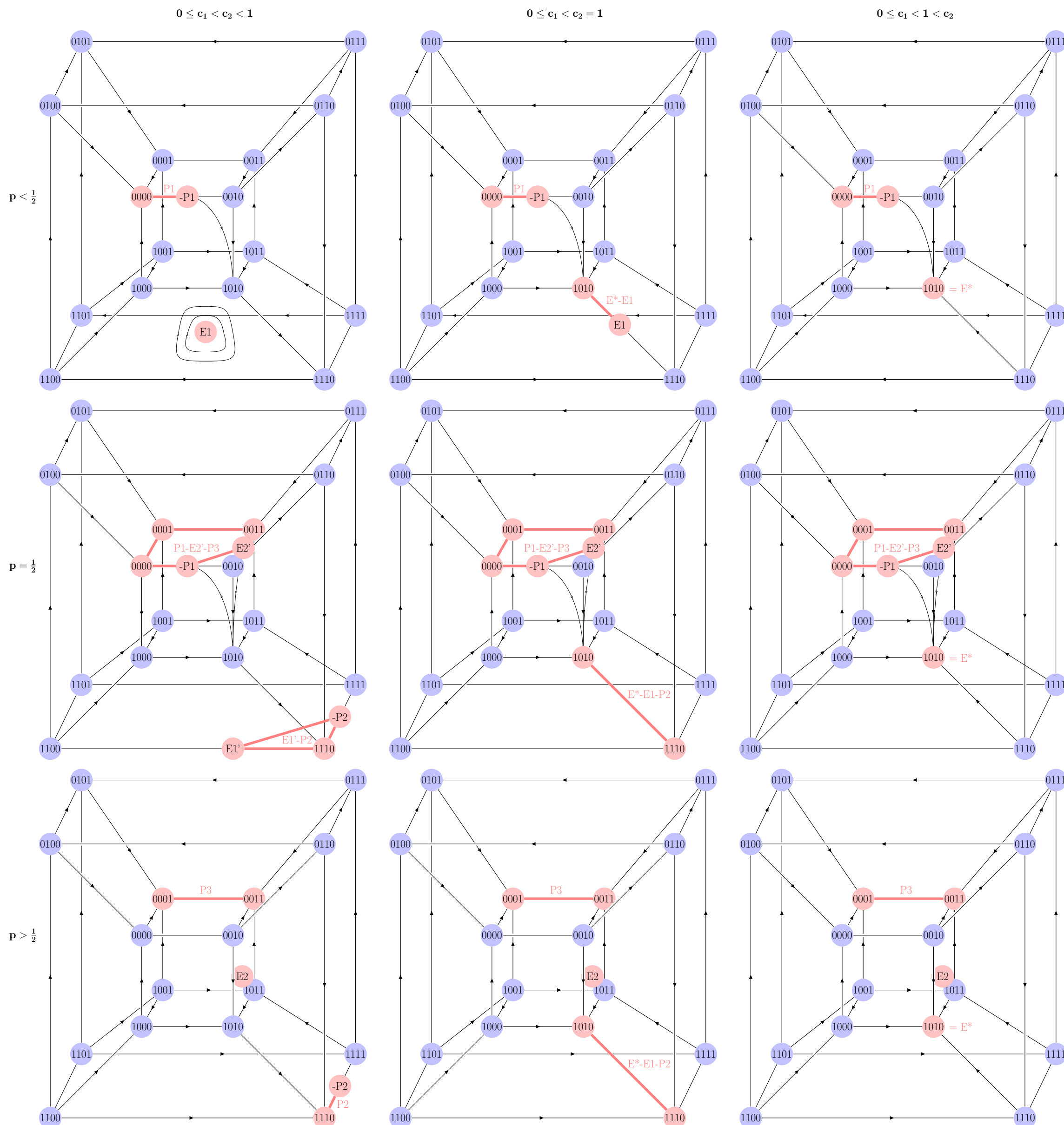
(Best-response dynamics: both E1'-P2 asymptotically stable; P1-E2'-P3 unstable but interior attracting.)

Equilibrium structure, $0 \leq c_1 < c_2 = 1$

$p < \frac{1}{2}$:	(E*-E1) : *fully–part revealing*	$\mathbf{h} \longrightarrow \mathrm{s}$	$\mathrm{s} \longrightarrow \mathbf{a}$
	Index: $+1$. Fwd Ind	$\mathbf{l} \longrightarrow \mathrm{s}$ with prob $\leq \frac{\mathbf{p}}{\mathbf{1-p}}$	$\bar{\mathrm{s}} \longrightarrow \bar{\mathbf{a}}$
	(P1): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathrm{s}}$	$\mathrm{s} \longrightarrow a$ with prob $\leq c_1$
	Index: 0. Not Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathrm{s}}$	$\bar{\mathrm{s}} \longrightarrow \bar{\mathbf{a}}$
$p > \frac{1}{2}$:	(E2) : *partially revealing*	$\mathbf{h} \longrightarrow \bar{\mathrm{s}}$ with prob $\frac{\mathbf{1-p}}{\mathbf{p}}$	$\mathrm{s} \longrightarrow \mathbf{a}$
	Index: -1. Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathrm{s}}$	$\bar{\mathrm{s}} \longrightarrow \mathbf{a}$ with prob $1 - c_1$
	(E*-E1'-P2):	$\mathbf{h} \longrightarrow \mathrm{s}$	$\mathrm{s} \longrightarrow \mathbf{a}$
	Index: $+1$. Fwd Ind	$\mathbf{l} \longrightarrow \mathrm{s}$ with any prob	$\bar{\mathrm{s}} \longrightarrow \bar{\mathbf{a}}$
	(P3): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathrm{s}}$	$\mathrm{s} \longrightarrow \mathrm{a}$ with any prob
	Index: $+1$. Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathrm{s}}$	$\bar{\mathrm{s}} \longrightarrow \mathbf{a}$
$p = \frac{1}{2}$:	(E*-E1'-P2):	$\mathbf{h} \longrightarrow \mathrm{s}$	$\mathrm{s} \longrightarrow \mathbf{a}$
	Index: $+1$ Fwd Ind	$\mathbf{l} \longrightarrow \mathrm{s}$ with any prob	$\bar{\mathrm{s}} \longrightarrow \mathbf{a}$
	(P1-E2-P3): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathrm{s}}$	$\mathrm{s} \longrightarrow \mathbf{a}$ with $y \in [0, \min \{y' + c_1, 1$
	Index: 0. Not all Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathrm{s}}$	$\bar{\mathrm{s}} \longrightarrow p^{\star} = p = \frac{1}{2}$: $\mathbf{a}$ with $y' \in [0,$

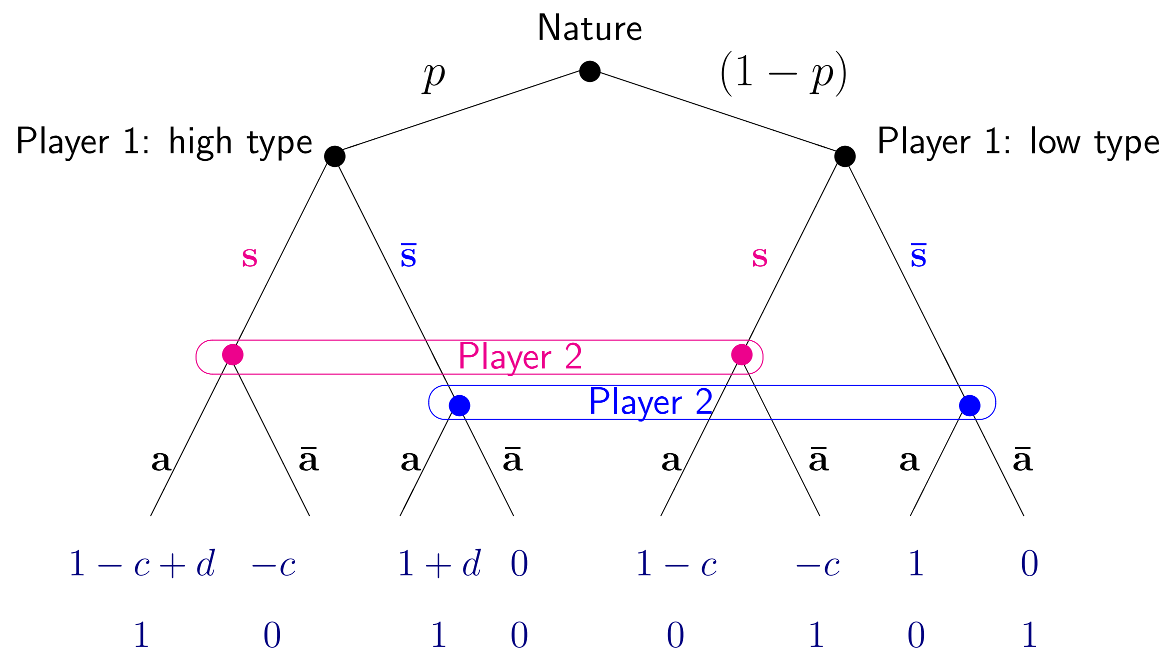
Equilibrium structure, $0 \leq c_1 < 1$, $c_2 > 1$

$p < \frac{1}{2}$:	(E*) : *fully revealing*	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow \mathbf{a}$
	Index: $+1$. Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow \bar{\mathbf{a}}$
	(P1): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathrm{s} \longrightarrow a$ with prob $\leq c_1$
	Index: 0. Not Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow \bar{\mathbf{a}}$
$p > \frac{1}{2}$:	(E2) : *partially revealing*	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$ with prob $\frac{\mathbf{1-p}}{\mathbf{p}}$	$\mathbf{s} \longrightarrow \mathbf{a}$
	Index: -1. Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow \mathbf{a}$ with prob $1 - c_1$
	(E*): *fully revealing*	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow \mathbf{a}$
	Index: $+1$. Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow \bar{\mathbf{a}}$
	(P3): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathrm{s} \longrightarrow$ a with any prob
	Index: $+1$. Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow \mathbf{a}$
$p = \frac{1}{2}$:	(E*): *fully revealing*	$\mathbf{h} \longrightarrow \mathbf{s}$	$\mathbf{s} \longrightarrow \mathbf{a}$
	Index: $+1$ Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow \mathbf{a}$
	(P1-E2-P3): *both use* $\bar{s}$	$\mathbf{h} \longrightarrow \bar{\mathbf{s}}$	$\mathrm{s} \longrightarrow \mathbf{a}$ with $y \in [0, \min\{y' + c_1, 1\}]$
	Index: 0. Not all Fwd Ind	$\mathbf{l} \longrightarrow \bar{\mathbf{s}}$	$\bar{\mathbf{s}} \longrightarrow p^{\star} = p = \frac{1}{2}$: $\mathbf{a}$ with $y' \in [0, 1]$



0 ≤ c1 < c2 < 1
0 ≤ c1 < c2 = 1
0 ≤ c1 < 1 < c2
p < 1/2
p = 1/2
p > 1/2
P1
-P1
E1
E*-E1
= E*
P1-E2'-P3
E2'
E1'-P2
E1'
-P2
E*-E1-P2
P3
E2
P2

Class II: uniform costs, differential gains



	$\mathbf{aa}$	$\mathbf{a\bar{a}}$	$\mathbf{\bar{a}a}$	$\mathbf{\bar{a}\bar{a}}$
$\mathbf{ss}$	$1-c+pd,\ p$	$1-c+pd,\ p$	$-c,\ 1-p$	$-c,\ 1-p$
$\mathbf{s\bar{s}}$	$1-pc+pd,\ p$	$p(1-c)+pd,\ 1$	$-pc+(1-p),\ 0$	$-pc,\ 1-p$
$\mathbf{\bar{s}s}$	$1-(1-p)c+pd,\ p$	$(1-p)(1-c),\ 0$	$p-(1-p)c+pd,\ 1$	$-(1-p)c,\ 1-p$
$\mathbf{\bar{s}\bar{s}}$	$1+pd,\ p$	$0,\ 1-p$	$1+pd,\ p$	$0,\ 1-p$

Class II: Same equilibrium structure as class I: replace c_1 by $\frac{c}{1+d}$

Combination of class I and II: replace c_1 by $\frac{c_1}{1+d}$

References

[1] Akerlof, G. A. 1970. The market for "lemons": quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84 (3): 488–500.

[2] Archetti, M. 2000. The origin of autumn colours by coevolution. *Journal of Theoretical Biology* 205: 652–630.

[3] Banks, J. S., and J. Sobel. 1987. Equilibrium selection in signaling games, *Econometrica* 55 (3): 647–661.

[4] Berger, U. 2005. Fictitious play in $2 \times n$ games. *Journal of Economic Theory* 120 (2): 139–154.

[5] Bergstrom, C. T., Lachmann M. 1997. Signalling among relatives I. Is costly signalling *too* costly? *Philosophical Transactions of the Royal Society London B*, 352: 609–617.

[6] Bergstrom, C. T., Lachmann M. 2001. Alarm calls as costly signals of anti-predator vigilance: the watchful babbler game. *Animal Behavior* 61: 535–543.

[7] Bliege Bird, R., Smith E. A. 2005. Signaling theory, strategic interaction and symbolic capital. *Current Anthropology* 46 (2): 221–248.

[8] Bliege Bird, R., Smith E. A., Bird, D. W. 2001. The hunting handicap: costly signaling in human foraging strategies. *Behavioral Ecology and Sociobiology* 50: 9–19.

[9] Bourdieu, P. 1982. *Ce que parler veut dire: l'économie des échanges linguistiques*, Paris: Fayard.

[10] Bourdieu, P. 1991 *Language and Symbolic Power*, ed. by J. B. Thompson, transl. by G. Raymond and M. Adamson. Cambridge, MA: Harvard University Press.

[11] Brown, B., Levinson C.S. 1987. *Politeness: Some Universals in Language Usage*. Cambridge/New York: Cambridge University Press.

[12] Caro, T. M. 1986a. The functions of stotting in Thomson's gazelles: a review of the hypotheses. *Animal Behavior* 34: 649–662.

[13] Caro, T. M. 1986b. The functions of stotting in Thomson's gazelles: some tests of the predictions. *Animal Behavior* 34: 663–684.

[14] Cho, I-K. and D. M. Kreps. 1987. Signaling games and stable equilibria. *Quarterly Journal of Economics*, 102 (2): 179–221.

[15] Cressman, R. 2003. *Evolutionary Dynamics and Extensive Form Games*. Cambridge MA: MIT Press.

[16] Dawkins, R., Krebs, J. R. 1978. Animal signals: information and manipulation. In: Eds. Krebs J. R. and Davies N. B. (Eds.) *Behavioral Ecology: An Evolutionary Approach*. Oxford: Blackwell, pp. 282–309.

[17] Demichelis, S., Ritzberger K. 2003. From evolutionary to strategic stability. *Journal of Economic Theory* 113 (1): 51–75.

[18] Gaunersdorfer, A. Hofbauer J., Sigmund K. 1991. On the dynamics of asymmetric games, *Theoretical Population Biology*, 39: 345–357.

[19] Godfray, H. C. J. 1991 Signaling of need by offspring to their parents. *Nature* 352: 328–330.

[20] Govindan, S., Wilson, R., 2009. On forward induction. *Econometrica* 77 (1): 1–28.

[21] Grafen, A. 1990. Biological signals as handicaps. *Journal of Theoretical Biology*, 144 (4): 517–546.

[22] Harsanyi, J. C. 1967. Games with incomplete information played by 'Bayesian' players. *Management Science*, 14 (3): 159–182.

[23] Hawkes K., Bliege Bird R. 2002. Showing off, handicap signaling, and the evolution of men's work. *Evolutionary Anthropology* 11: 58–67.

[24] Hofbauer, J. Sigmund, K. 1988. *The Theory of Evolution and Dynamical Systems*, Cambridge UK: Cambridge University Press.

[25] Hofbauer, J., Sigmund, K. 1998. *Evolutionary Games and Population Dynamics*, Cambridge UK: Cambridge University Press.

[26] Hofbauer, J., P. Schuster, Sigmund, K., 1979. A note on evolutionarily stable strategies and game dynamics. *Journal of Theoretical Biology* 81: 609–612.

[27] Huttegger, S. M., Zollman, K. J. S. 2010. Dynamic stability and basins of attraction in the Sir Philip Sidney game. *Proceedings of the Royal Society London B*, 277: 1915–1922.

[28] Huttegger, S. M., Zollman, K. J. S. 2016. The robustness of hybrid equilibria in costly signaling games. *Dynamic Games and Applications*, 6: 347–358.

[29] Kohlberg, E., Mertens J.-F. 1986. On the strategic stability of equilibria. *Econometrica* 54(5): 1003–1037.

[30] Krebs, J. R., Dawkins, R. 1984. Animal signals: mind-reading and manipulation. In: Krebs J. R. and Davies N. B. (Eds.) *Behavioral Ecology: An Evolutionary Approach*, 2nd Edition. Oxford: Blackwell, pp. 380–402.

[31] Kreps, D. M., Sobel, J. 1994. Signalling. In: Aumann, R. J, Hart, S. (ed.), *Handbook of Game Theory*, Vol. 2. Amsterdam/New York: Elsevier, pp. 849–867.

[32] Kreps, D. M., Wilson, R. 1982. Sequential equilibria. *Econometrica*, 50 (4): 863–894.

[33] Kuhn, H. W. 1950. Extensive games. *Proceedings of the National Academy of Sciences*, 36: 570–576.

[34] Kuhn, H. W. 1953. Extensive games and the problem of information. In: H. W. Kuhn and A. W. Tucker (Eds.), *Contributions to the Theory of Games*, Vol. II, Princeton, Princeton University Press, 193–216.

[35] Lachmann, M., Bergstrom, C. T. 1998. Signalling among relatives II. Beyond the Tower of Babel. *Theoretical Population Biology*, 54: 146–160.

[36] Lachmann, M., Bergstrom, C. T., Számado, S. 2001. Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences* 98 (23): 13189–13194.

[37] Mauss, Marcel. 1924. *The gift: Forms and functions of exchange in archaic societies*. London: Cohen and West.

[38] Maynard Smith, J. 1991. Honest signalling: The Philip Sidney game. *Animal Behavior*, 42: 1034–1035.

[39] Maynard Smith, J., Price, G. R. 1973. The logic of animal conflict. *Nature*, 246: 15–18.

[40] Milgrom P., Roberts, J. 1986. Price and advertising signals of product quality. *Journal of Political Economy*, 94(4): 796–821.

[41] Miller, M. H., Rock, K. 1985. Dividend policy under asymmetric information. *The Journal of Finance*, XL (4), 1031–1051.

[42] Nash, J. 1950. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36: 48–49.

[43] Nash, J. 1951. Noncooperative games. *The Annals of Mathematics*, 54 (2): 286–295.

[44] Nöldeke, G., Samuelson, L. 1997. A dynamic model of equilibrium selection in signaling games. *Journal of Economic Theory*, 73 (1): 118–156.

[45] Pinker, S. 2007. *The Stuff of Thought. Language as a Window into Human Nature*. New York: Viking.

[46] Pinker, S., Nowak, M. A., Lee, J.J. 2008. The logic of indirect speech. *Proceedings of the National Academy of Sciences* 105, 833–838.

[47] Ritzberger, K. 1994. The theory of normal form games from the differentiable viewpoint. *International Journal of Game Theory* 23: 207–236.

[48] Ritzberger, K. 2002. *Foundations of Non-Cooperative Game Theory*, Oxford University Press.

[49] Shapley, L. S. 1974. A note on the Lemke-Howson algorithm. *Mathematical Programming Study* 1: 175–189.

[50] Sobel, J. 2009. Signaling Games. In: R. Meyers (Ed.) *Encyclopedia of Complexity and System Science*. New York: Springer, pp. 8125–8139.

[51] Spence, M. 1973. Job market signaling. *Quarterly Journal of Economics*, 87 (3): 355–374.

[52] Számadó, S. 2011. The cost of honesty and the fallacy of the handicap principle. *Animal Behavior*, 81: 3–10.

[53] Taylor, P., Jonker, L., 1978. Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 40, 145–156.

[54] Van Rooy, R. 2003. Being polite is a handicap: towards a game theoretical analysis of polite linguistic behavior. *Proceedings of TARK 9*.

[55] Veblen, T. 1899. *The Theory of the Leisure Class: An Economic Study of Institutions*. New York: The Macmillan Company.

[56] von Stengel, B. 2021. Finding Nash equilibria of two-person games. Working Paper, London School of Economics.

[57] Wagner, E. O. 2013. The dynamics of costly signaling. *Games*, 4: 163–181.

[58] Weibull, J. 1995. *Evolutionary game theory*. Cambridge, MA: MIT Press.

[59] Zahavi, A. 1975. Mate selection—a selection for a handicap. *Journal of Theoretical Biology*, 53 (1): 205–214.

[60] Zollman, K. J. S., Bergstrom, C. T., Huttegger, S. M. 2013. Between cheap and costly signals: the evolution of partially honest communication. *Proceedings of the Royal Society London B* 280: 20121878.

In our game:

$p < 1/2$:

- E1: Isolated and quasistrict $\longrightarrow$ regular
  - removing unused strategies $\longrightarrow$ 2×2 cyclic game
  - in this game, E1 only equilibrium $\longrightarrow$ index $+1$

  $\Rightarrow$ candidate for asymptotically stable equilibrium
- P1: by Index Theorem $\longrightarrow$ index 0

  $\Rightarrow$ not asymptotically stable, under no evolutionary dynamics

$p > 1/2$:

- P2: by robustness $\longrightarrow$ index $+1$
- E2: Isolated and quasistrict $\longrightarrow$ regular
  - removing unused strategies $\longrightarrow$ 2×2 coordination game with 3 equilibria: E2 and two strict equilibria (index $+1$)
  - by Index Theorem $\longrightarrow$ index -1.
- P3: by Index Theorem $\longrightarrow$ index $+1$

Phenomena explained:

When prior is low, $p < 1/2$:

- Partially revealing equilibrium (E1):

  costly signal becomes a means to shape the belief of the other; specifically: "push the belief of the other up" $\longrightarrow$ for of "indirect speech"

- (E1) welfare-improving over "no-signaling" equilibrium outcome (P1).

When prior is high, $p > 1/2$:

- both routinely using the costly signal (P2) and routinely not using costly signal (P3) are strategically and evolutionarily stable equilibrium outcomes
  - overstatement (P2) and understatement (P3)
  - P2: Social tragedy: everybody needs to signal, but signal carries no information!
  - P3 can also be interpreted as "countersignaling"
- co-existence of these two equilibrium outcomes $\rightarrow$ possible source of discrimination: when (P2) or (P3) is linked to some other observable characteristic

Equilibrium refinement

In classical game theory: restrictions on beliefs "off the equilibrium path" (= after an unused signal)

- Kohlberg and Mertens (1986): "never-a-weak-best-response" criterion
- Banks and Sobel (1987): "divinity"
- Govindan and Wilson (2009): "forward induction" (FI)

$\rightarrow$ all coincide here. Quite weak selection force: discard the no-signaling equilibrium outcome P1; all other equilibria survive (for the two generic cases $p < 1/2$ and $p > 1/2$).

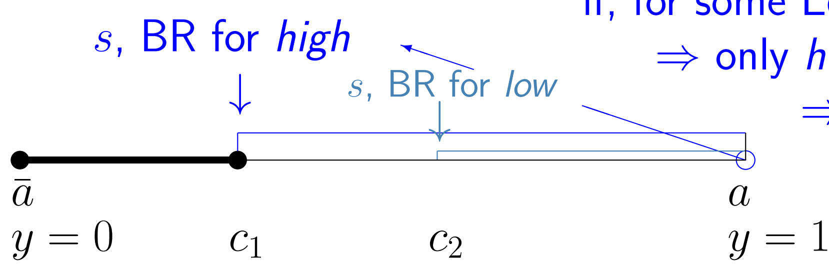
The argument:

P1: both types of player 1 take $\bar{s}$; player 2 in response to $\bar{s}$ takes $\bar{a}$.

Off equilibrium path: in response to the unused costly signal s, player 2 takes a with a prob of c_1 at most $\longrightarrow$ implies that 2 attributes to the high type a belief of $\frac{1}{2}$ at most!

But not "plausible" (by various criteria) $\rightarrow$ equilibrium discarded!

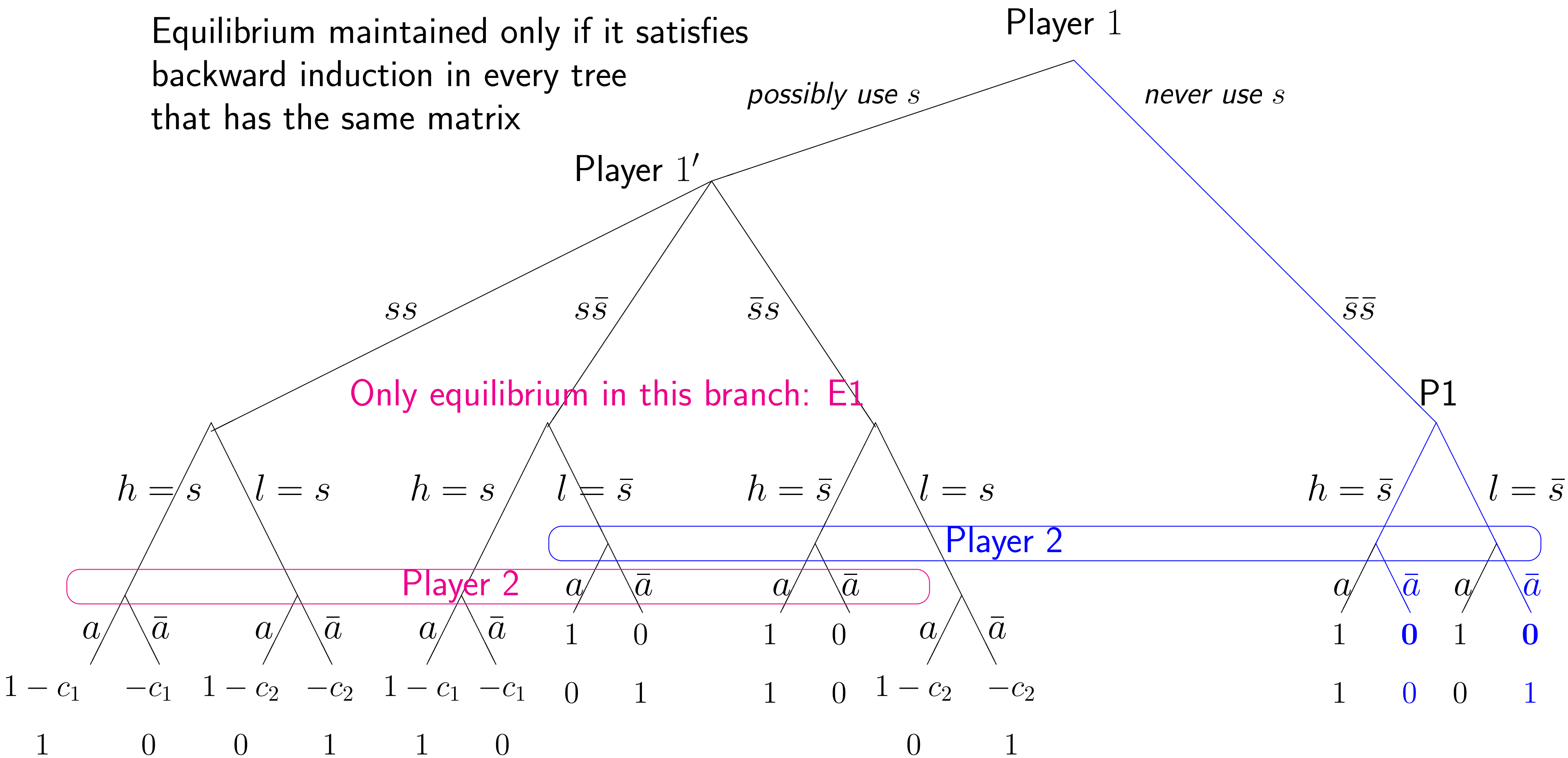
Forward induction (Govindan and Wilson 2009)
eq. to "never-a-weak BR": after s, type maintained only if, for some Eq point, s is alternative BR for him
$\Rightarrow$ only *high* survives
$\Rightarrow$ after s, 2 should take a!
$\Rightarrow$ clash with equilibrium P1!



Divinity (Banks and Sobel 1987):
after s, type maintained only
if there is no other type who has a larger better-off set

Forward induction (Govindan & Wilson 2009): foundation in "invariance + sequentiality"

Equilibrium maintained only if it satisfies backward induction in every tree that has the same matrix



This tree has the same matrix as class I. But P1 (both use $\bar{s}$), not backward induction! $\longrightarrow$ E1 only backward-induction equilibrium!