Meaning in Costly-Signaling Games

Christina Pawlowitsch

Université Paris–Panthéon–Assas

Based on joint work with Josef Hofbauer

Workshop: What is strategic information?

London School of Economics, April 15–16, 2024

Hofbauer and Pawlowitsch, "Evolutionary Dynamics in Costy Signaling Games" (WP):

$\longrightarrow$ Minimalist model with: 2 types ("high" and "low"), 2 signals (s and "not s"), 2 actions ("accept" and "do not accept") in response to signals

$\longrightarrow$ Classification:

- vary cost parameters (3 typical cases)
- vary prior beliefs (3 critical cases)
- 9 classes

$\longrightarrow$ For each class:

- analyze entire equilibrium structure
  - index theory
  - evolutionary dynamics: replicator and best-response dynamics
  - classical refinements of Bayesian-Nash sequential equilibrium, based on restrictions on beliefs (over types) "off the equilibrium path" (philosophers might call this a "counterfactual situation")

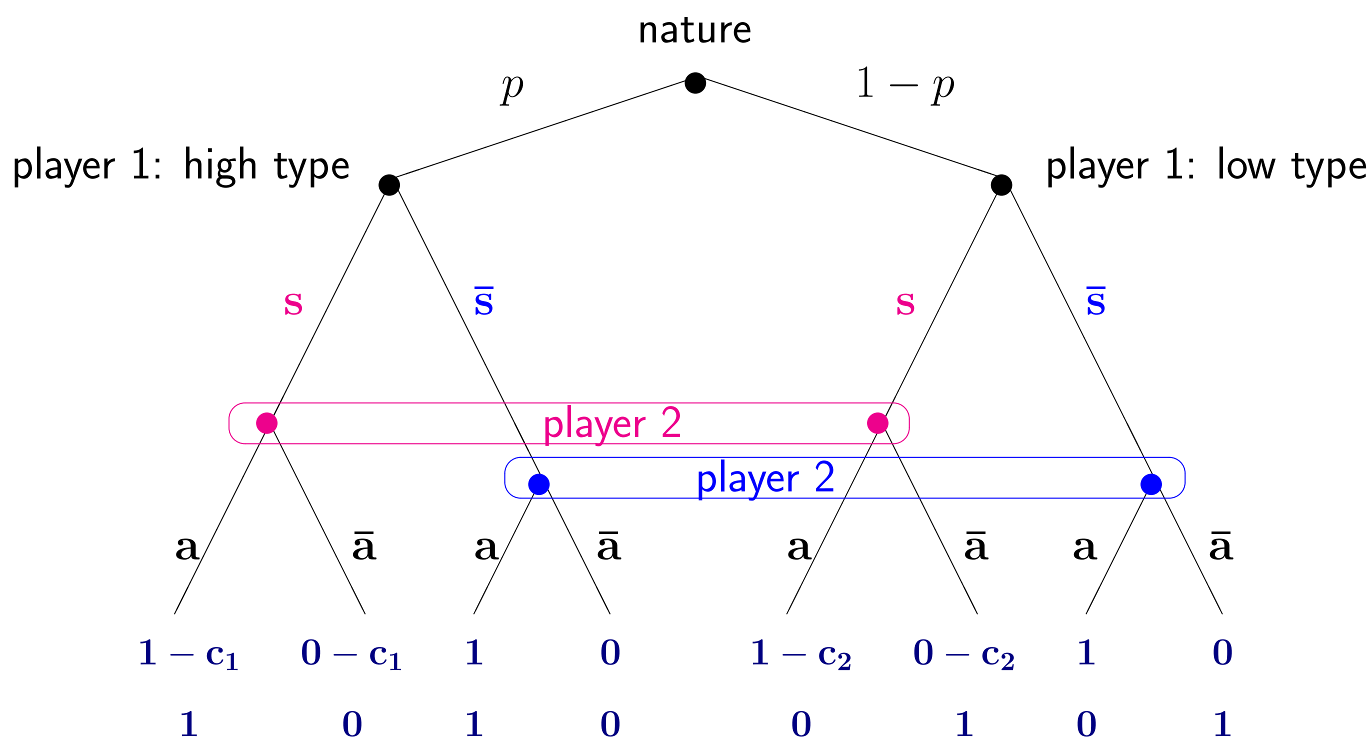
This talk/current project:

Explore potential for applications in the study of language

- Point of depature: These kind of costly-signaling games allow us to say something about the "emergence of meaning" attached to a signal (or its absence) as a function of
  - the costs of the signal carried by various types
  - the prior probability distribution over types

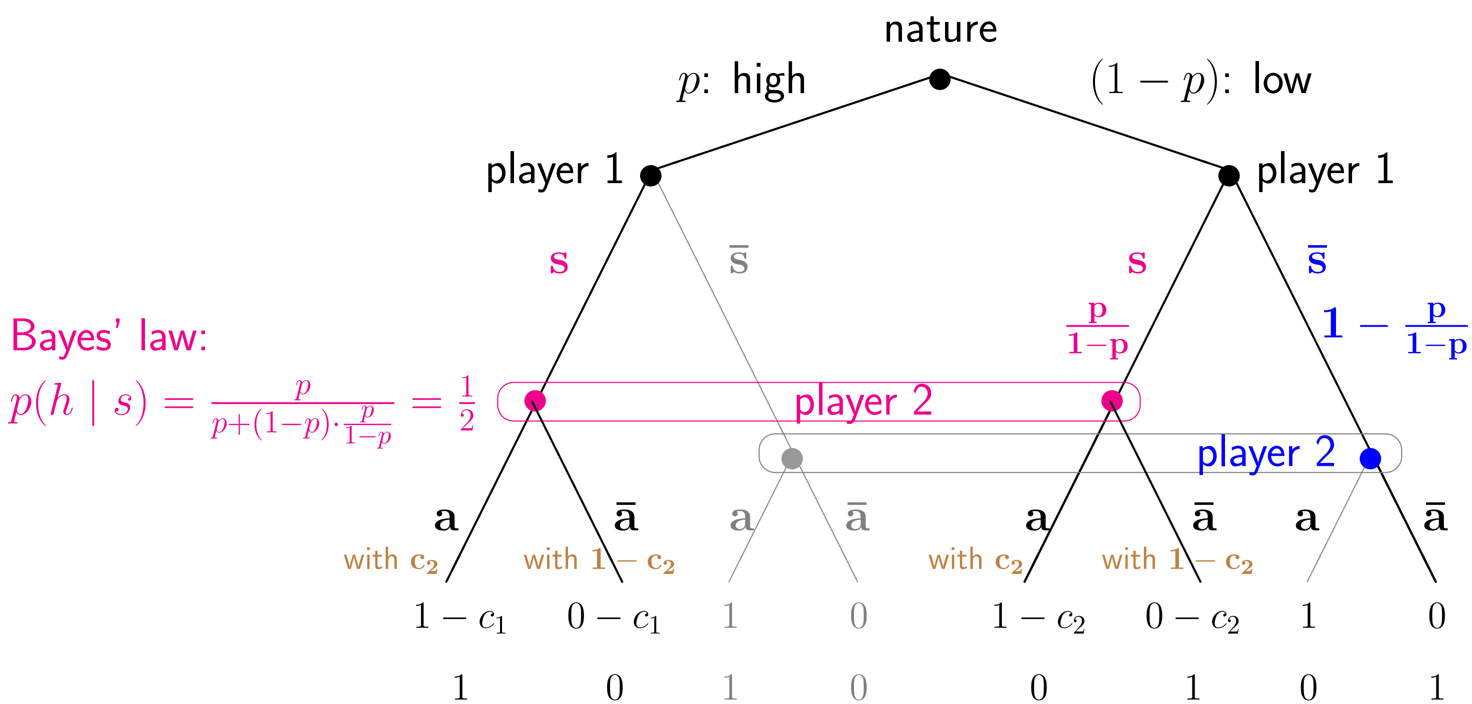
$\longrightarrow$ First: overview of results

Costly-signaling game (discrete version of Spence 1973)



	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	$1,\ p$	$0,\ 1-p$	$1,\ p$	$0,\ 1-p$

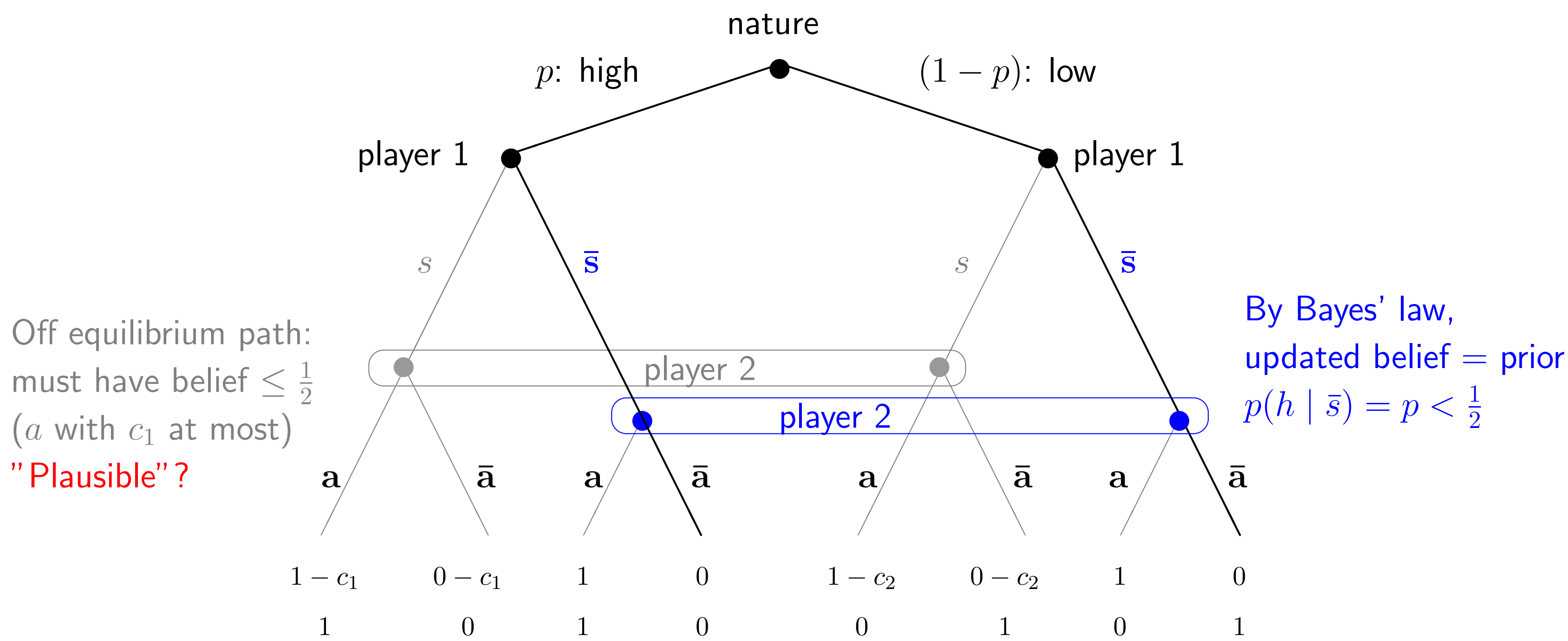
Case $0 \leq c_1 < c_2 < 1$, $p < 1/2$: E1 partially revealing equilibrium



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

- 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.

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



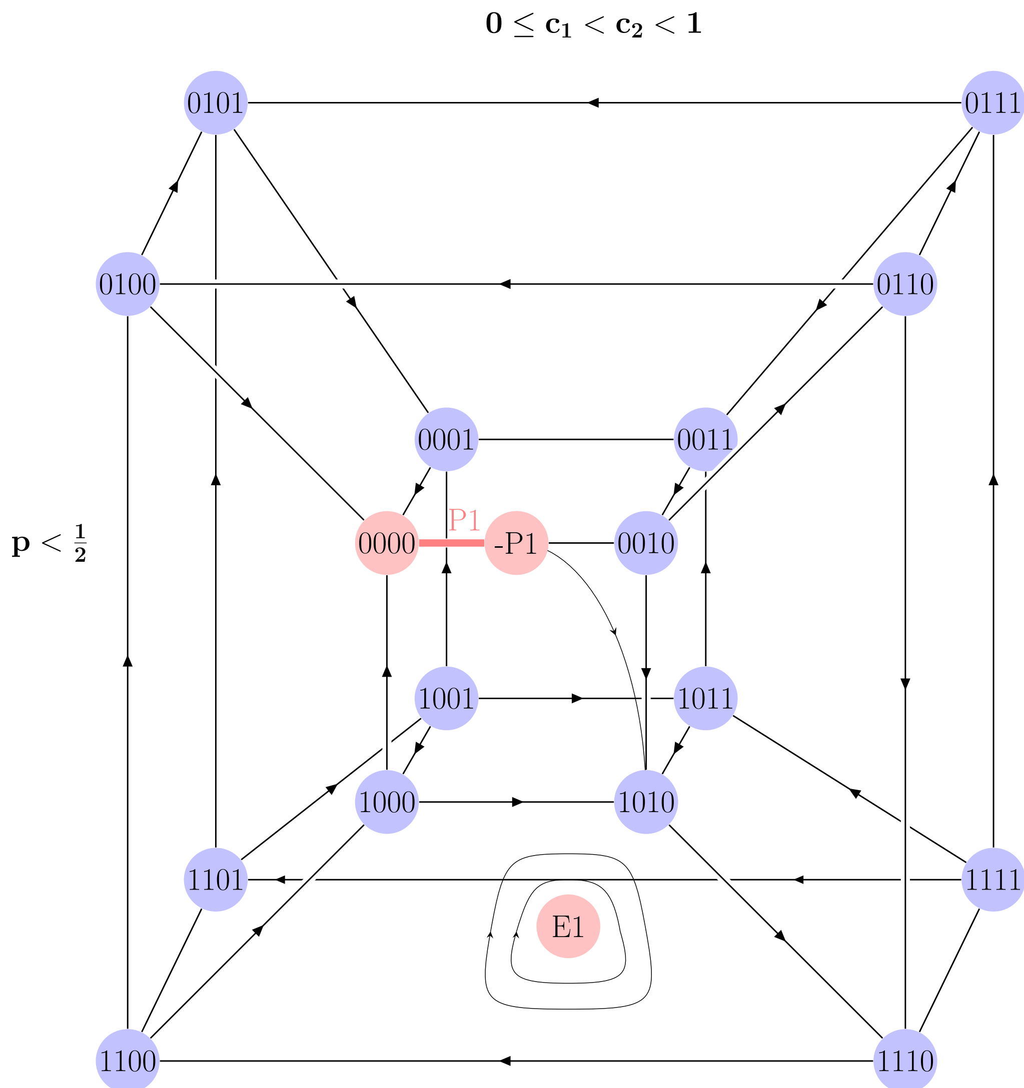
	$\mathbf{aa}$	with $y \in [0, c_1] \to \mathbf{a\bar{a}}$	$\mathbf{\bar{a}a}$	with $1-y \to \mathbf{\bar{a}\bar{a}}$
$\mathbf{ss}$	$1 - pc_1 - (1-p)c_2,\ p$	$\mathbf{1 - pc_1 - (1-p)c_2},\ p$	$-pc_1 - (1-p)c_2,\ \mathbf{1-p}$	$-pc_1 - (1-p)c_2,\ \mathbf{1-p}$
$\mathbf{s\bar{s}}$	$1 - pc_1,\ p$	$p(1-c_1),\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

- 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.

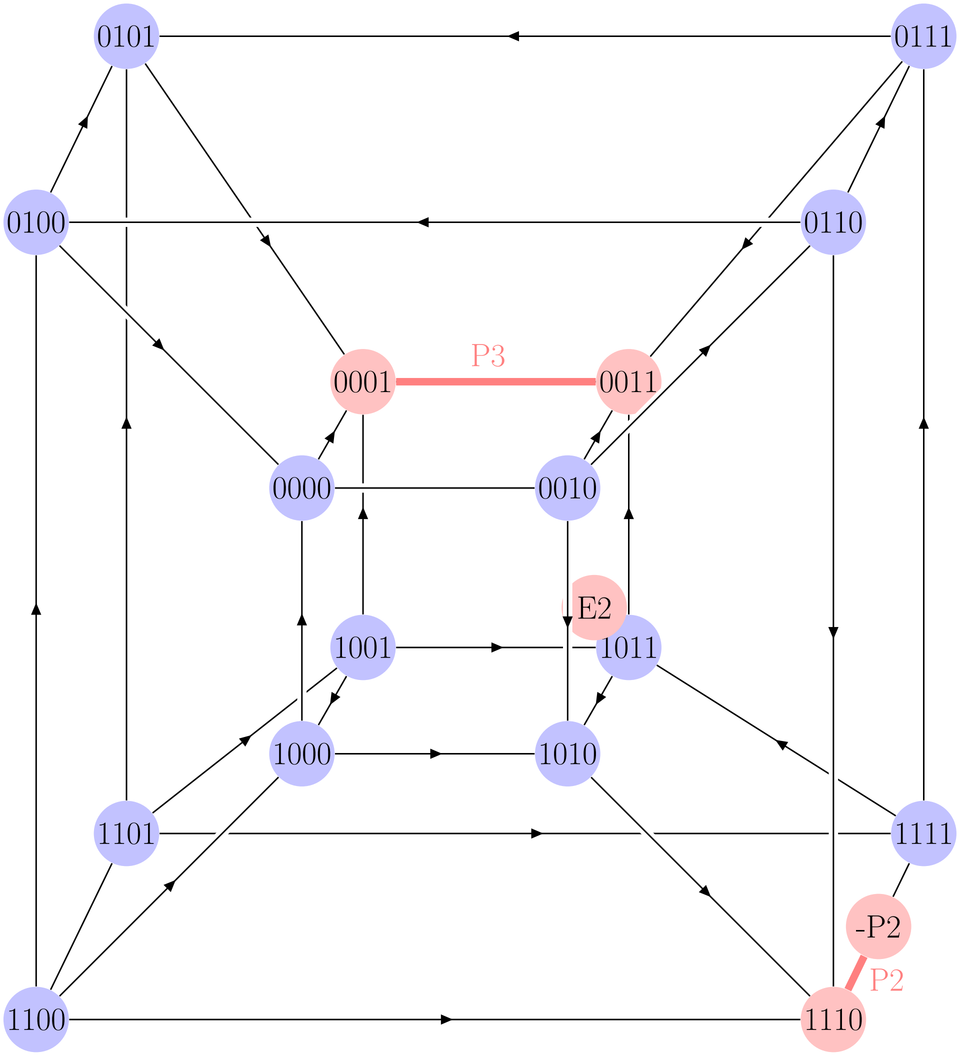
Table 1. Equilibrium structure of the game in Figure 1: $0 \leq c_1 < c_2 < 1$

Prior	Equilibrium component	Index	Rep. dyn.	BR dyn.	NWBR, forward induction	Invariance criterion	Payoffs:
$p < \frac{1}{2}$	(E1): *partially revealing/ partially pooling in* s: $(1, \frac{p}{1-p}, c_2, 0)$	$+1$	stable	as. stable	yes	invariant	$h: c_2 - c_1$ $\ell: 0$ $2: 1-p$
	(P1): *pooling in* $\bar{s}$: $(0, 0, y, 0)$, $y \in [0, c_1]$	0	unstable	unstable	no	not invariant	h: 0 ℓ: 0 $2: 1-p$
$p > \frac{1}{2}$	(E2): *partially revealing/ partially pooling in* $\bar{s}$: $(1 - \frac{1-p}{p}, 0, 1, 1 - c_1)$	-1	unstable	unstable	yes	invariant	h: $1 - c_1$ ℓ: $1 - c_1$ $2: p$
	(P2): *pooling in* s: $(1, 1, 1, y')$, $y' \in [0, 1 - c_2]$	$+1$	stable	as. stable	yes	invariant	h: $1 - c_1$ ℓ: $1 - c_2$ $2: p$
	(P3): *pooling in* $\bar{s}$: $(0, 0, y, 1)$, $y \in [0, 1]$	$+1$	as. stable	as. stable	yes	invariant	h: 1 ℓ: 1 $2: p$
$p = \frac{1}{2}$	(E1'-P2): *pooling in* s: $(1, 1, y, y')$, $y \in [c_2, 1]$, $y' \in [0, y - c_2])$	$+1$	stable	as. stable	yes	invariant	h: $[c_2 - c_1, 1 - c_1]$ ℓ: $[0, 1 - c_2]$ $2: \frac{1}{2}$
	(P1-E2'-P3): *pooling in* $\bar{s}$: $(0, 0, y, y')$, $(y, y') \in [0, 1]^2$, $y \leq y' + c_1$	0	unstable	unstable	only when $y' \in [1 - c_1, 1]$	invariant	h: $[0, 1]$ ℓ: $[0, 1]$ $2: \frac{1}{2}$

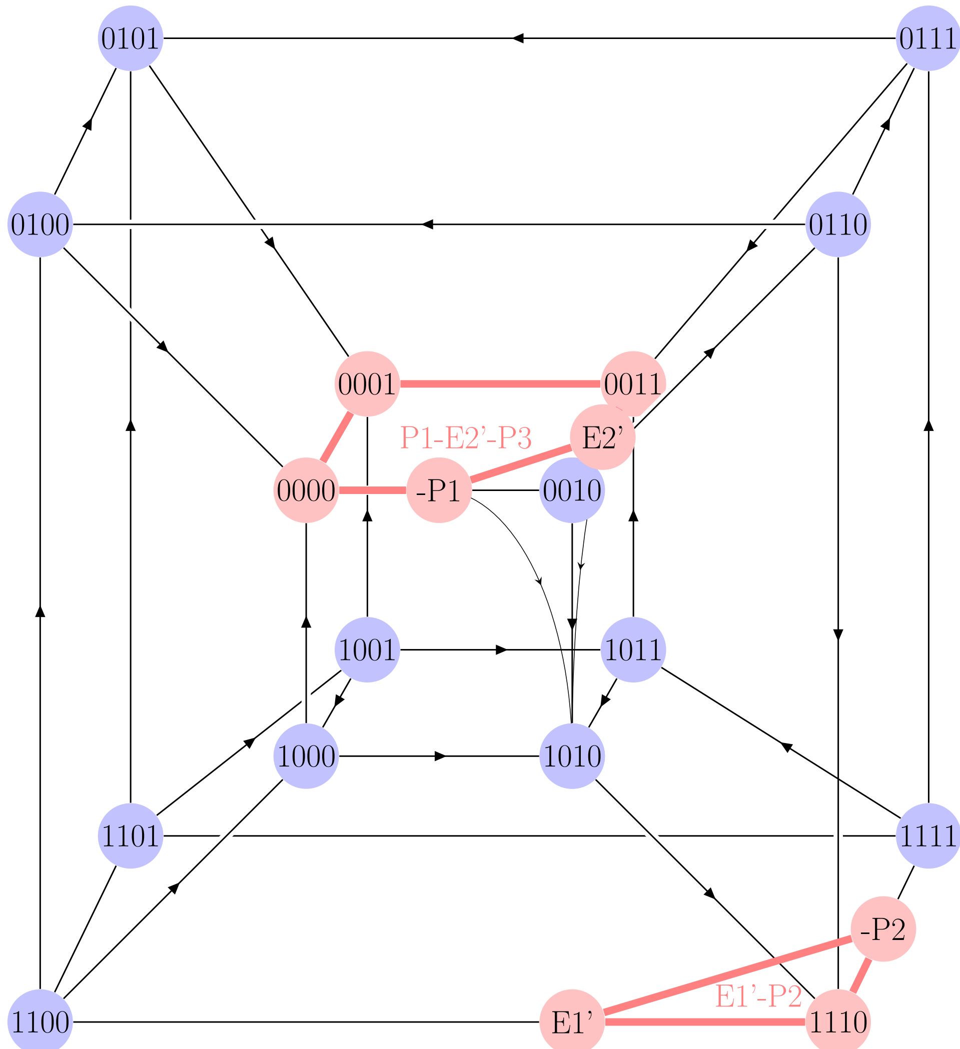
Replicator dynamics

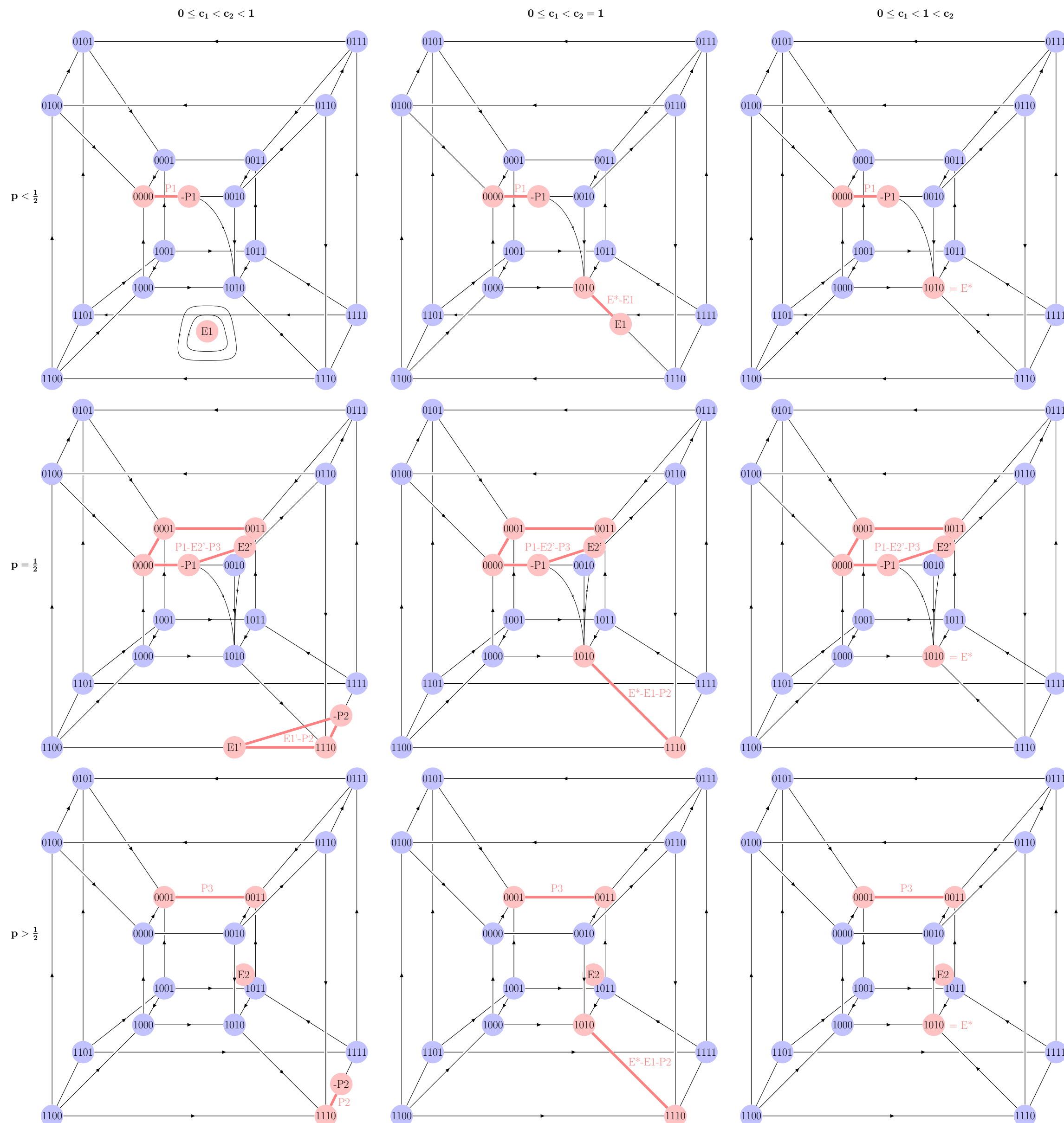


$\mathbf{p > \frac{1}{2}}$



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





0 ≤ c1 < c2 < 1
0 ≤ c1 < c2 = 1
0 ≤ c1 < 1 < c2
p < 1/2
p = 1/2
p > 1/2
0000
0001
0010
0011
0100
0101
0110
0111
1000
1001
1010
1011
1100
1101
1110
1111
P1
-P1
E1
E*-E1
= E*
P1-E2'-P3
E2'
E1'
E1'-P2
-P2
E*-E1-P2
P3
E2
P2

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”

Applications in the study of language:

Costly signal S:	Absence of costly signal $\bar{S}$:
"marked form"	"unmarked form"

- politeness: S: polite form; $\bar{S}$: not polite form
- accents: S: standard; $\bar{S}$: not standard (regional accent)

  $\longrightarrow$ "code switching" and "style shifting"

Phenomena explained:

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

- Partially revealing equilibrium E1 (high always S; low with some probability):

  costly signal becomes a means to shape the belief of the other; specifically: "push the belief of the other up"

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

  $\longrightarrow$ model of "indirect speech"

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.

  P2: Social tragedy: everybody needs to signal, but signal carries no information!

  $\longrightarrow$ overstatement (P2) and understatement (P3)

  $\longrightarrow$ P3 can also be interpreted as "countersignaling"

  $\longrightarrow$ when (P2) or (P3) is linked to some other observable characteristic: possible source of discrimination

A real-world example:

Chers tous,

J'espère que la reprise n'est pas trop rude !

Il n'y aura pas de conseil de département lundi prochain faute d'un ordre du jour suffisamment étoffé.

La DRH nous demande toutefois de faire formellement approuver par le bureau du département le classement des candidats sur le poste LRU en mathématiques que nous avons publié en urgence au mois de juillet.

Cette approbation permettra, après avis favorable des conseils centraux, à la personne recrutée de débuter son service au mois d'octobre.

Le classement a été réalisé au mois de juillet par une commission inter-centres présidée par N ... H .... Voici le classement :

1 H ... C ....
2 Z ... K ...
3 I ... A ...

Amitiés,
Bertrand C ...

Cher tous,
Merci à la commission inter-centres pour ce travail! Avis favorable
Claudine

Chers tous,
avis favorable également!
Bonne journée,
Lucie

OK
AB

Bonjour à tous,
Je suis favorable également.
Amicalement,
Maria

Bonjour à tous,
avis favorable également. Bonne reprise à tous !
Amicalement,
Fabienne

Cher Bertrand, chers tous,
Avis favorable également.
Amitiés, Christina

Merci à tous pour votre réponse rapide !
Amitiés, Bertrand

References

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

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

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

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

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

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

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

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

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

[10] 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.

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

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

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

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

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

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

1) The index of equilibria

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 of the replicator dynamics

Ritzberger (1994, 2002): extends this to equilibrium components:

- Index as an integer, such that the sum over all components is again $+1$
- Robust under payoff perturbations: Let C be a component and U an open neighborhood of C such that all equilibria in the closure of U are already in C. Let C^ε be the set of all equilibria of the perturbed game that lie in U—the finite union of connected components $C_1^\varepsilon, \ldots, C_k^\varepsilon$. By Brouwer's degree theory, the sum of the indices of $C_1^\varepsilon, \ldots, C_k^\varepsilon$ equals the index of C. (C^ε might be empty—but only if C has index 0.)

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 $+1$.

In our game (based on Hofbauer and Pawlowitsch 2023):

$p < 1/2$:

- 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 $+1$

2) Restricting beliefs "off the equilibrium path"

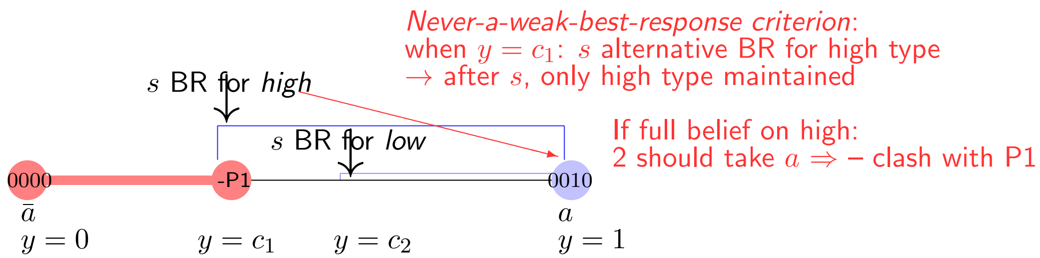
In signaling games: "off the equilibrium path" $=$ after an unused signal

- Cho and Kreps (1987): "never-a-weak-best-response" criterion
- Banks and Sobel (1987): "divinity"
- Govindan and Wilson (2009): "forward induction"

$\rightarrow$ 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 $p > 1/2$).

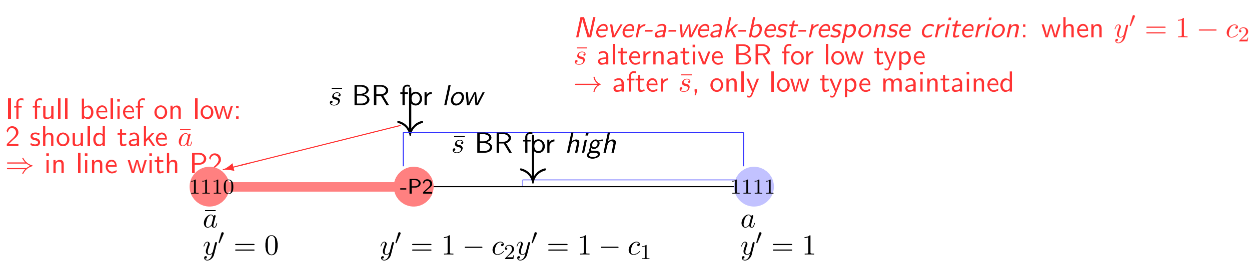
$p < 1/2$:

P1 ($\bar{s}\bar{s} \to \bar{a}$): NOT robust against "belief-based" refinements: responses of player 2 to the off-the-equilibrium-path signal s:



$p > 1/2$:

P2 ($ss \to a$): robust against "belief-based" refinements: responses of player 2 to the off-the-equilibrium-path signal $\bar{\mathrm{s}}$:



3) Invariance

Kohlberg and Mertens (1986) : a Nash equilibrium should be selected only if it corresponds to a sequential Bayesian Nash equilibrium in every extensive-form game that maps to the same (reduced) normal form.

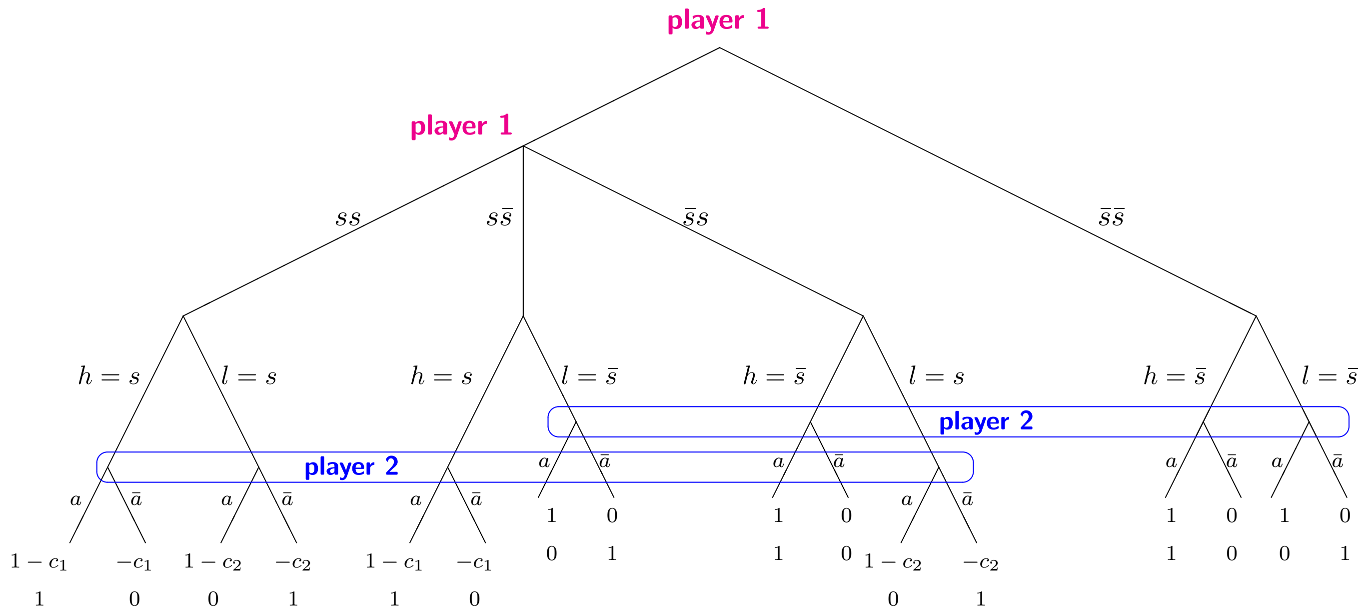
Govindan and Wilson (2009):

invariance $\Rightarrow$ forward induction $\Rightarrow$ never-a-weak-best-response, “divinity”

For the game studied here:

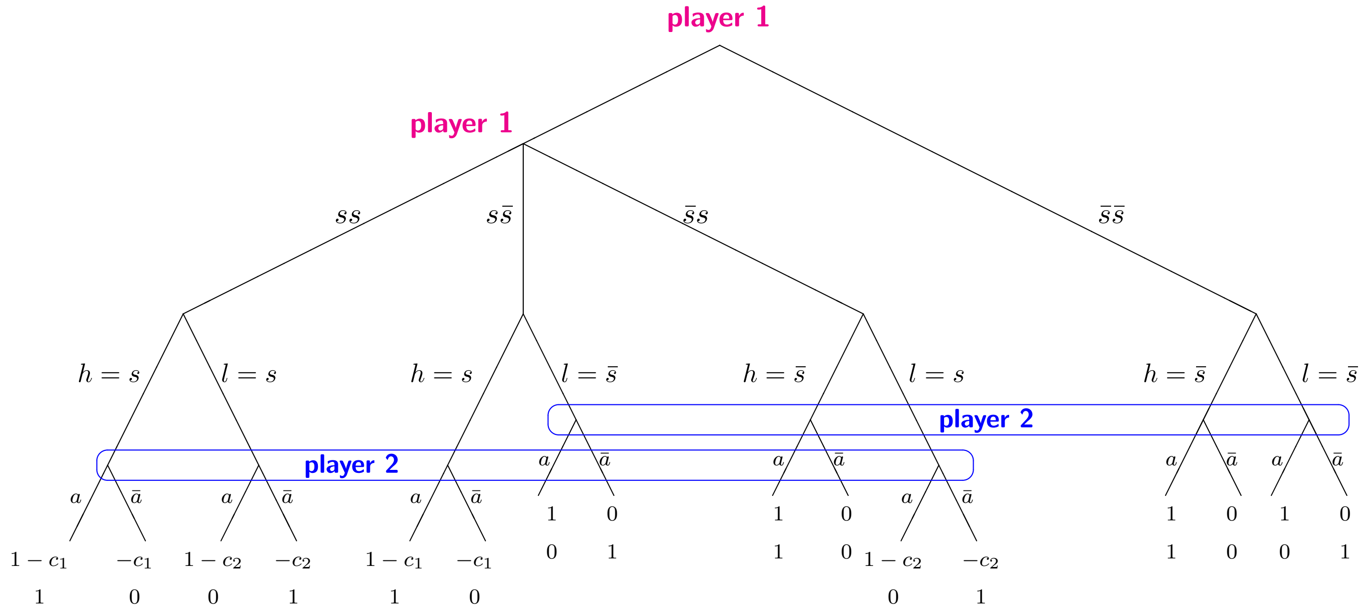
invariance $\Rightarrow$ forward induction $\Leftrightarrow$ never-a-weak-best-response, “divinity”

$\mathbf{p < 1/2}$: P1 (index 0) not forward induction $\Rightarrow$ not invariant



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

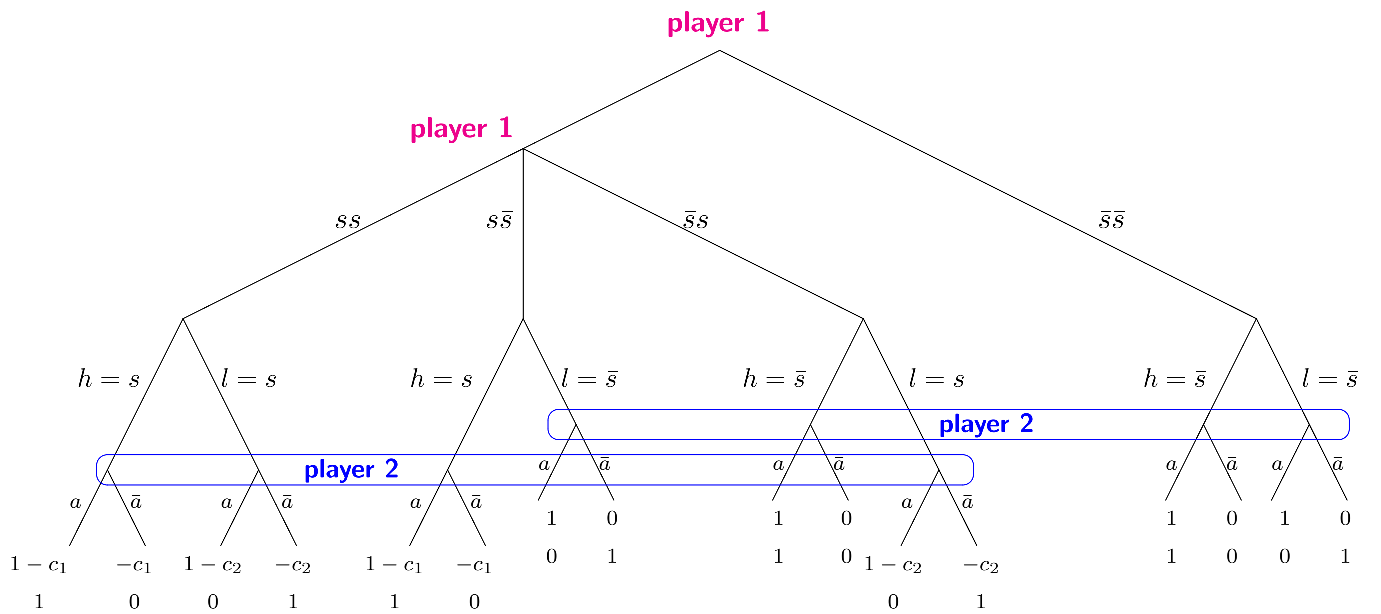
$\mathbf{p < 1/2}$: P1 (index 0) not forward induction $\Rightarrow$ not invariant



	aa	**aā**	**āa**	**āā**
ss	$1 - pc_1 - (1-p)c_2,\ p$	$1 - pc_1 - (1-p)c_2,\ p$	$-pc_1 - (1-p)c_2,\ 1-p$	$-pc_1 - (1-p)c_2,\ 1-p$
s$\bar{\text{s}}$	$1 - pc_1,\ p$	$p(1-c_1),\ 1$	$-pc_1 + (1-p),\ 0$	$-pc_1,\ 1-p$
$\bar{\text{s}}$s	$1 - (1-p)c_2,\ p$	$(1-p)(1-c_2),\ 0$	$p - (1-p)c_2,\ 1$	$-(1-p)c_2,\ 1-p$

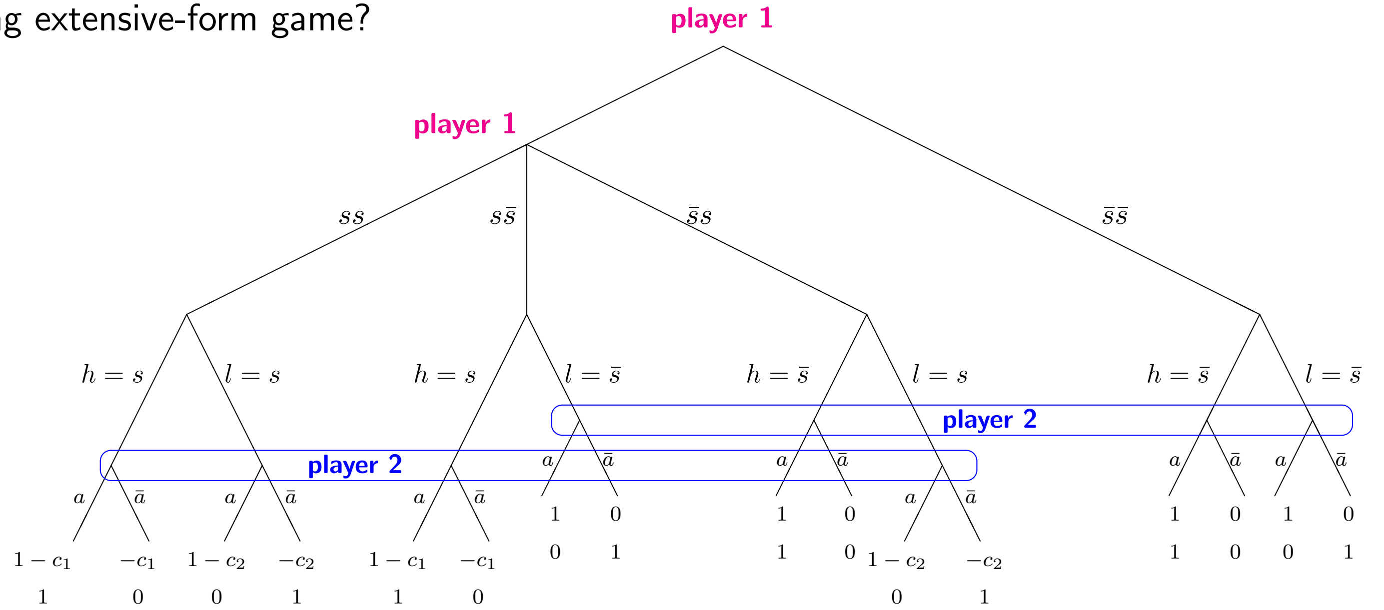
	aa	**aā**	**āa**	**āā**
$\bar{\text{s}}\bar{\text{s}}$	$1,\ p$	$0,\ 1-p$	$1,\ p$	$0,\ 1-p$

Case $\mathbf{p > 1/2}$, $2p-1 < (1-p)c_2 - pc_1$, which guarantees that $\bar{s}s$ (high $\bar{s}$, low s) is strictly dominated by $s\bar{s}$: equilibrium outcome P3, in which both types of player 1 use $\bar{s}$ and player 2 accepts (a) (and can have any reaction to s) has index $+1$ and satisfies forward induction ... Is it a sequential equilibrium in the following extensive-form game?



	aa	a$\bar{\text{a}}$	$\bar{\text{a}}$a	$\bar{\text{a}}\bar{\text{a}}$
ss	$1-pc_1-(1-p)c_2,\ p$	$1-pc_1-(1-p)c_2,\ p$	$-pc_1-(1-p)c_2,\ 1-p$	$-pc_1-(1-p)c_2,\ 1-p$
s$\bar{\text{s}}$	$1-pc_1,\ p$	$p(1-c_1),\ 1$	$-pc_1+(1-p),\ 0$	$-pc_1,\ 1-p$
$\bar{\text{s}}$s	$1-(1-p)c_2,\ p$	$(1-p)(1-c_2),\ 0$	$p-(1-p)c_2,\ 1$	$-(1-p)c_2,\ 1-p$
$\bar{\text{s}}\bar{\text{s}}$	$1,\ p$	$0,\ 1-p$	$1,\ p$	$0,\ 1-p$

Case $\mathbf{p > 1/2}$, $2p - 1 < (1-p)c_2 - pc_1$, which guarantees that $\bar{s}s$ (high $\bar{s}$, low s) is strictly dominated by $s\bar{s}$: equilibrium outcome P3, in which both types of player 1 use $\bar{s}$ and player 2 accepts (a) (and can have any reaction to s) has index $+1$ and satisfies forward induction ... Is it a sequential equilibrium in the following extensive-form game?



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

	$\mathbf{aa}$	$\mathbf{a\bar{a}}$	$\mathbf{\bar{a}a}$	$\mathbf{\bar{a}\bar{a}}$
$\mathbf{\bar{s}\bar{s}}$	$1,\ p$	$0,\ 1-p$	$1,\ p$	$0,\ 1-p$