A simple learning strategy which realizes robust cooperation better than Pavlov in Iterated Prisoners' Dilemma

Joe Yuichiro Wakano

Norio Yamamura

(Center for Ecological Research, Kyoto Univ.)

Prisoner's Dilemma (PD)

	C	D	
C	3	0	
D	5	1	

C:cooperation
D:defection

Defection is always the best strategy in a single round PD game.

Pavlov

When the game is repeated (Iterated Prisoner's Dilemma or IPD), adaptive strategy is not clear.

Pavlov is proposed as a leading strategy in IPD
(Nowak & Sigmund 1993)

Pavlov := (1 , 0 , 0 , 1)

Probability of playing C after

CC CD DC DD

Performance of Classic Strategies

TFT (1,0,1,0) ends up very poor when noise is introduced.

Pavlov is invaded by All-D but
Pavlov like strategy (1-e,0,0,1-e) is not.
(Nowak & Sigmund 1993)

Is Pavlov like strategy an ESS in a strategy set of (p1,p2,p3,p4) ?

Performance of Classic Strategies

Payoff for (p1,p2,p3,p4) against (q1,q2,q3,q4) can be derived analytically (eigen vector in Markov process).

21^4 strategies are matched each other and payoff is calculated. By investigating the payoff matrix, we obtain ...

Pavlov like strategy (1-e,0,0,1-e) is invaded by (1,0,0,1).

Only two ESSs found: All-D and GRIM (1,0,0,0)

No strategy of type (p1,p2,p3,p4) can establish stable mutual cooperation.

Introduction of Learning Players

Reinforcement learning:

"Repeat when rewarded, change when punished" or

"*Win-Stay, Lose-Shift*"

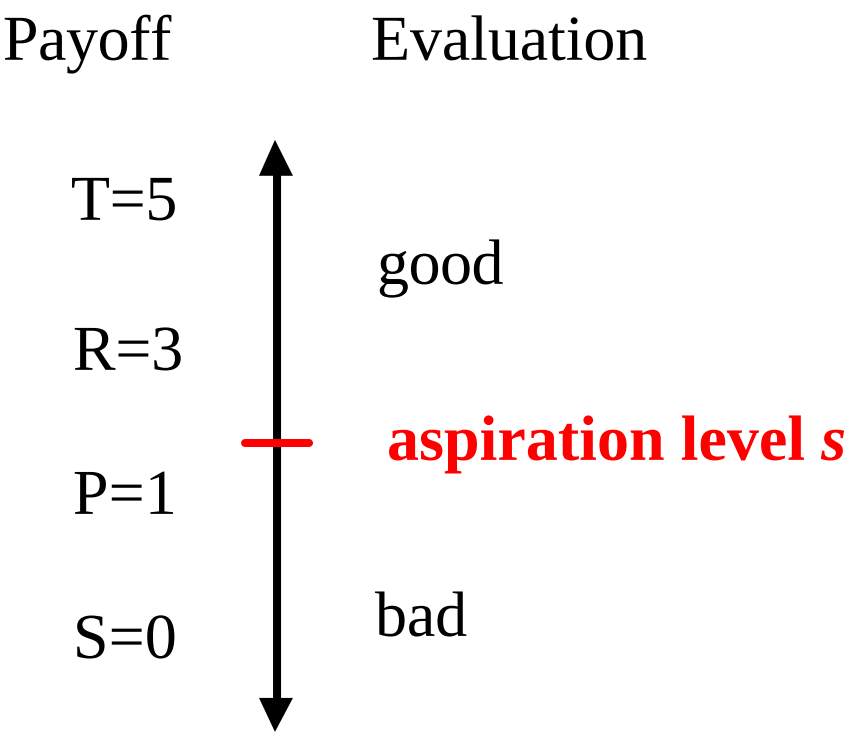
Simple and intuitively understandable rule.

Pavlov can be considered as the most simple player adopting reinforcement learning.
However, Pavlov considers the previous move only.

We will intoduce a learning player who considers past experiences (generalized Pavlov ?).

Aspiration Level

In order to adopt reinforcement learning, standard or aspiration level of evaluation must be defined.



To establish stable mutual cooperation, $1<s<3$ is necessary.

Model of Learning Players

Aspiration level s is genetically fixed.

Internal state h is changed during reptition of PD.

C is played when current internal state $h>0$, otherwise D.

$$\Delta h = a \cdot \mathrm{sgn}(f - s) \cdot \mathrm{sgn}(h)$$

$$\mathrm{sgn}(x) \equiv \begin{cases} 1 (x \geq 0) \\ -1 (x < 0) \end{cases}$$

Examples:

If C is played and the resulting score f is larger than s, cooperation is affirmatively learned. That is, the player increases h to become more cooperative.

If C is played and the resulting score f is smaller than s, cooperation is negatively learned. That is, the player decreases h to become less cooperative.

If D is played and the resulting score f is larger than s, defection is affirmatively learned, leading decrease in h.

Model of Learning Players

Each IPD consists of 10,000 PD games.
Error or noise is introduced at probability *e*.
When error happens, the opposite action expected from *h* value is played.

Internal state *h* is reset to $a/2$ every time the opponent changes, thus the learner always plays D at the first round.

→ The learner can learn to exploit All-C.

Internal state *h* can be considered as ‘faith in cooperation’.
Such faith is altered by past experience.

Payoffs of both players in IPD is determined only by aspiration levels.

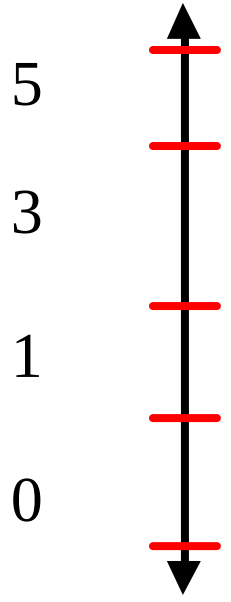
Digital Learner

$$\Delta h = 2 \cdot \mathrm{sgn}(f - s) \cdot \mathrm{sgn}(h)$$

$$\mathrm{sgn}(x) \equiv \begin{cases} 1 (x \geq 0) \\ -1 (x < 0) \end{cases}$$

Only 5 independent variants;

DL(s£0), DL(0<s£1), DL(1<s£3), DL(3<s£5), DL(s>5)



Example of Digital Learner's behavior

Table Comparison of digital learner ($1<s\leq 3, a = 2$) and Pavlov

strategy	opponent	actual realization of game (h)	average payoff
DL($1<s\leq 3$)	All-D	DD(-1), CD(1), DD(-1), CD(1), DD(-1),...	0.5
	All-C	DC(-1), DC(-3), DC(-5), DC(-7), DC(-9),...	5
	Pavlov	DD(-1), CC(1), CC(3), CC(5), CC(7),...	3
	TFT	DC(-1), DD(-3), DD(-1), CD(1), DC(-1),...	1.75
	DL($1<s\leq 3$)	DD(-1), CC(1), CC(3), CC(5), CC(7),...	3
Pavlov	All-D	DD, CD, DD, CD, DD, ...	0.5
	All-C	DC, DC, DC, DC, DC, ...	5
	Pavlov	DD, CC, CC, CC, CC, ...	3
	TFT	DC, DD, CD, DC, DD, ...	2
	DL($1<s\leq 3$)	DD, CC, CC, CC, CC, ...	3

ESS among 5 variants of Digital Learner

able Results of matches among digital learners.

	player (mutant)	s£0	0<s£1	opponent 1<s£3	(wild type) 3<s£5	s>5
a) 1% error	s£0	1.030049	1.030053	2.965461	2.965461	2.985005
	0<s£1	1.030048	1.030052	2.985288	2.965454	2.974166
	1<s£3	0.535146	0.535152	2.984959	0.584596	1.519351
	3<s£5	0.539876	0.53988	2.943325	1.738342	2.502725
	s>5	0.535053	0.545318	3.954128	1.258617	2.010029
b)10% error	s£0	1.290936	1.290994	2.851567	2.709865	2.849876
	0<s£1	1.290922	1.290977	2.851499	2.709804	2.788106
	1<s£3	0.848479	0.848549	2.865511	1.243602	1.848385
	3<s£5	0.888843	0.88891	2.482726	1.910963	2.405803
	s>5	0.850024	0.98887	3.526731	1.601163	2.090827

cooperative variant

ESS among 5 variants of Digital Learner

able Results of matches among digital learners.

	player (mutant)	opponent s£0	0<s£1	1<s£3	(wild type) 3<s£5	s>5
a) 1% error	s£0	1.030049	1.030053	2.965461	2.965461	2.985005
	0<s£1	1.030048	1.030052	2.985288	2.965454	2.974166
	1<s£3	0.535146	0.535152	2.984959	0.584596	1.519351
	3<s£5	0.539876	0.53988	2.943325	1.738342	2.502725
	s>5	0.535053	0.545318	3.954128	1.258617	2.010029
b)10% error	s£0	1.290936	1.290994	2.851567	2.709865	2.849876
	0<s£1	1.290922	1.290977	2.851499	2.709804	2.788106
	1<s£3	0.848479	0.848549	2.865511	1.243602	1.848385
	3<s£5	0.888843	0.88891	2.482726	1.910963	2.405803
	s>5	0.850024	0.98887	3.526731	1.601163	2.090827

ESS

cooperative variant

Cooperative variant DL(1<s<3) is not an ESS.

DL(s<0) is almost same as All-D, thus defection spreads

Analog Learner

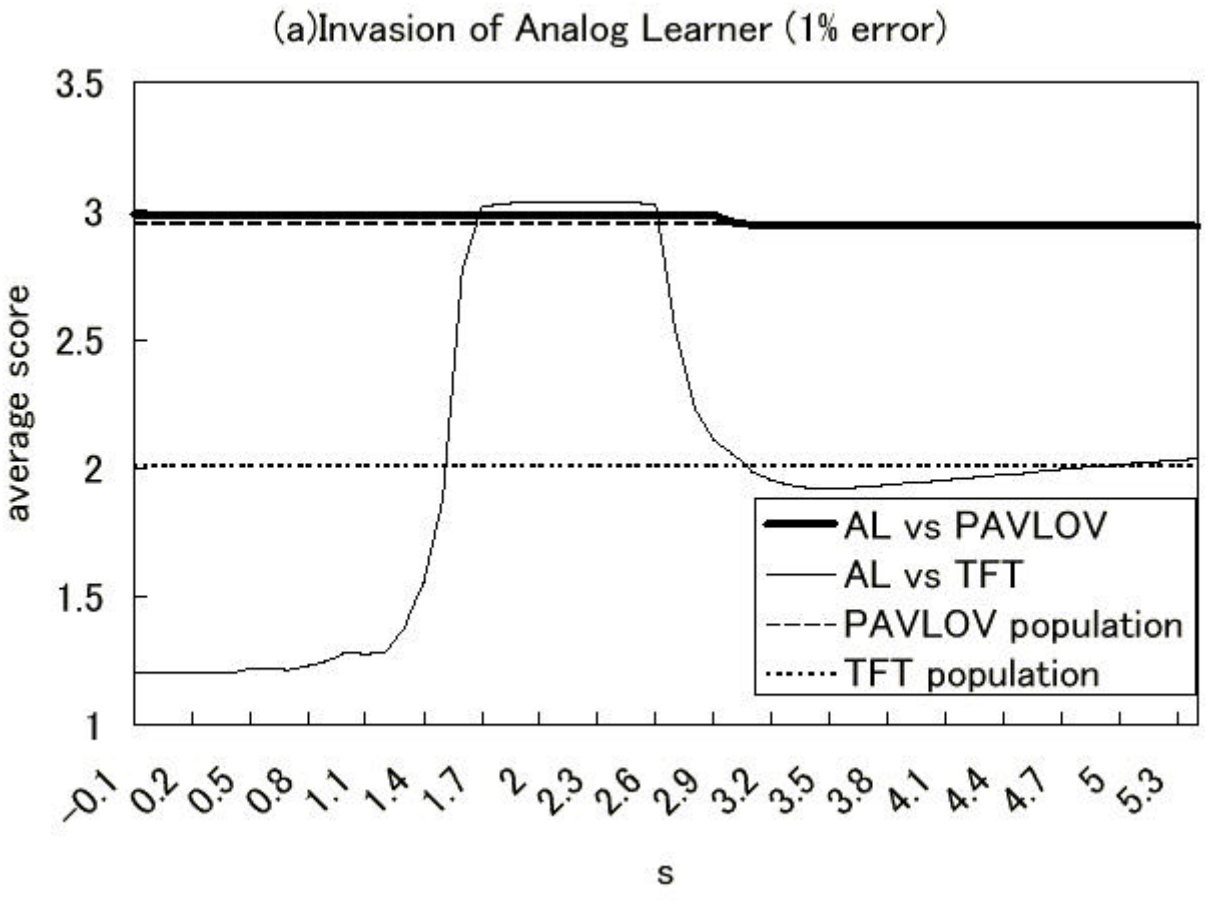
The impact on learning process may depend on the payoff value (instead of a constant as in DL model).
Analog evaluation model is developed.

$$\Delta h = 2 \cdot (f - s) \cdot \mathrm{sgn}(h)$$

Small difference in aspiration level *s* produces different learning process.

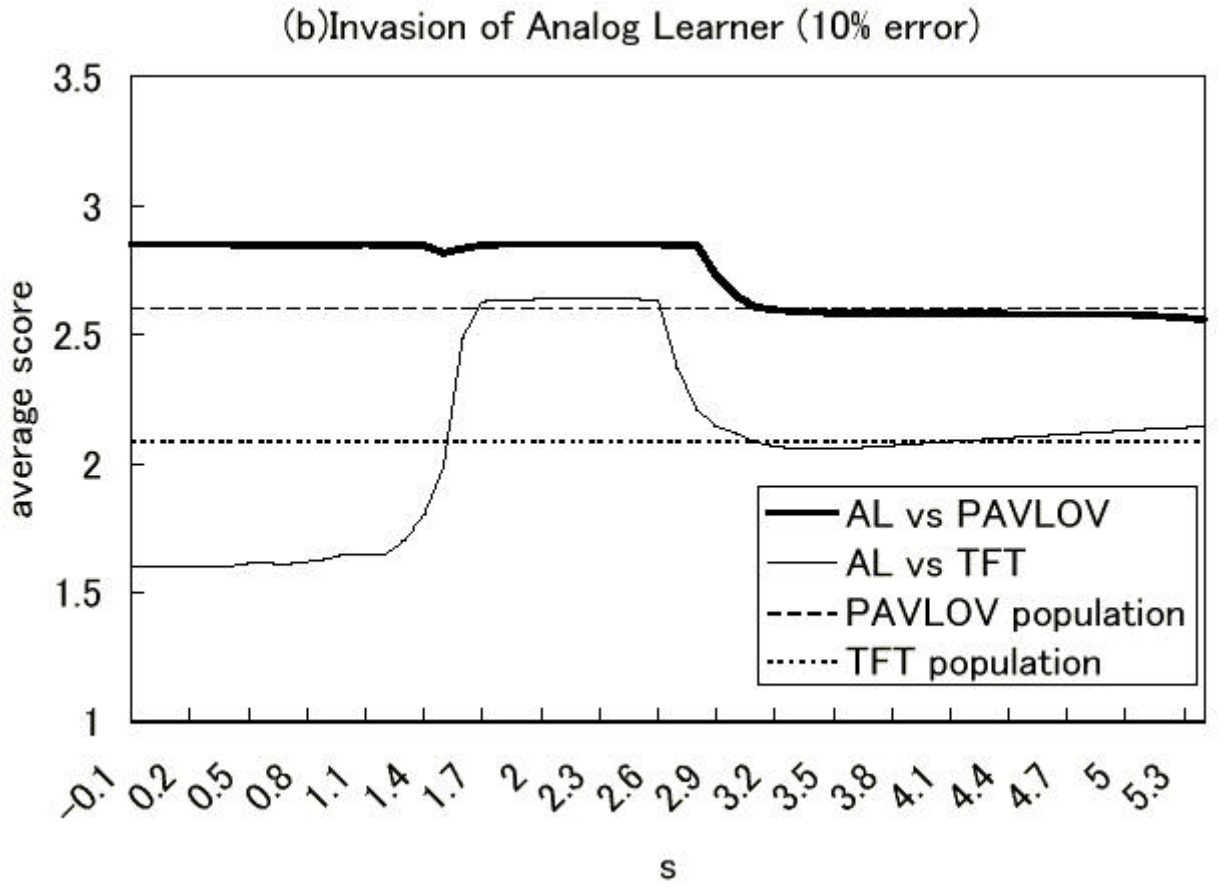
We used 56 different *s* values (s=-0.1, 0, 0.1, 0.2, …,5.4) in computer simulations.

Can Analog Learners invade classic strategies ?



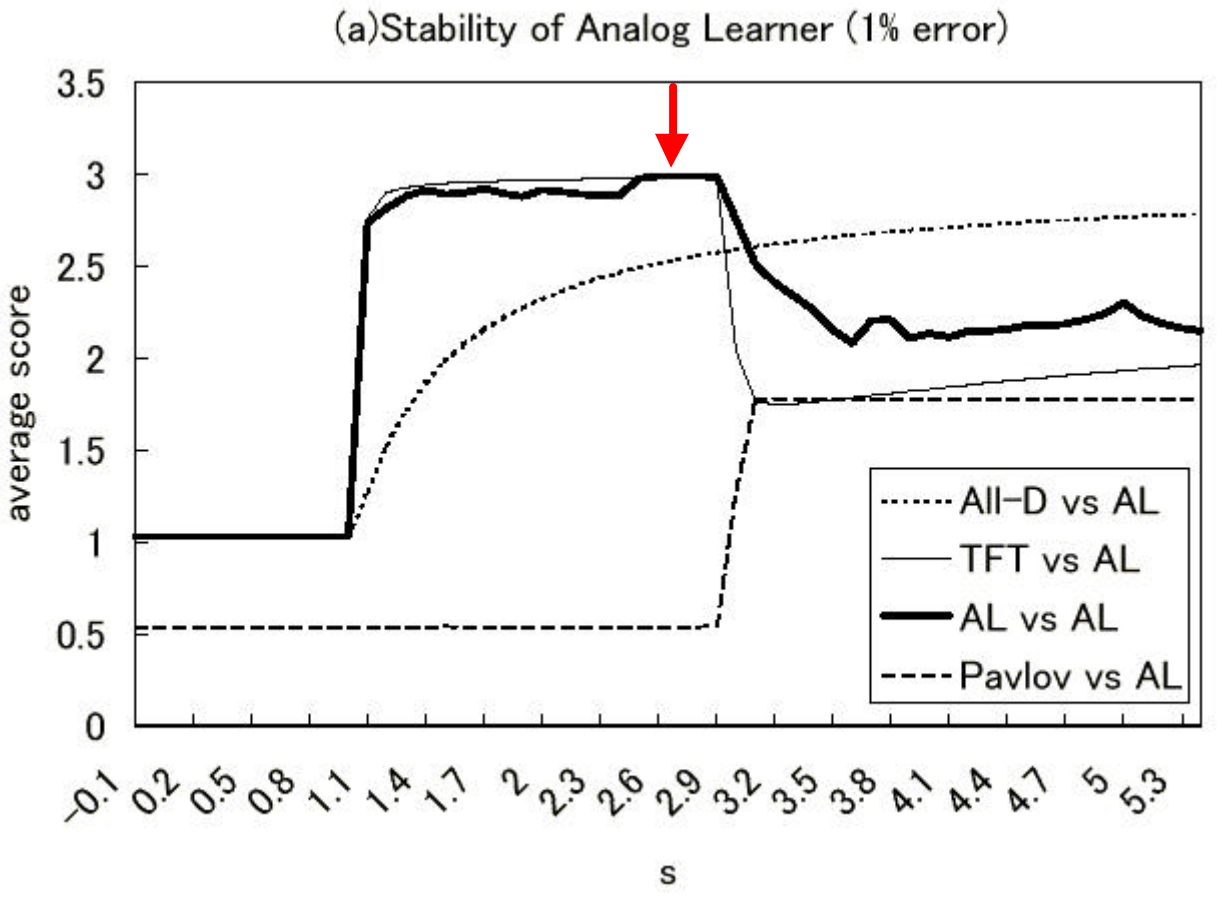
Analog Learner for s=1.6,…,3.0 invades All-C, All-D, TFT and Pavlov.

Can Analog Learners invade classic strategies ?



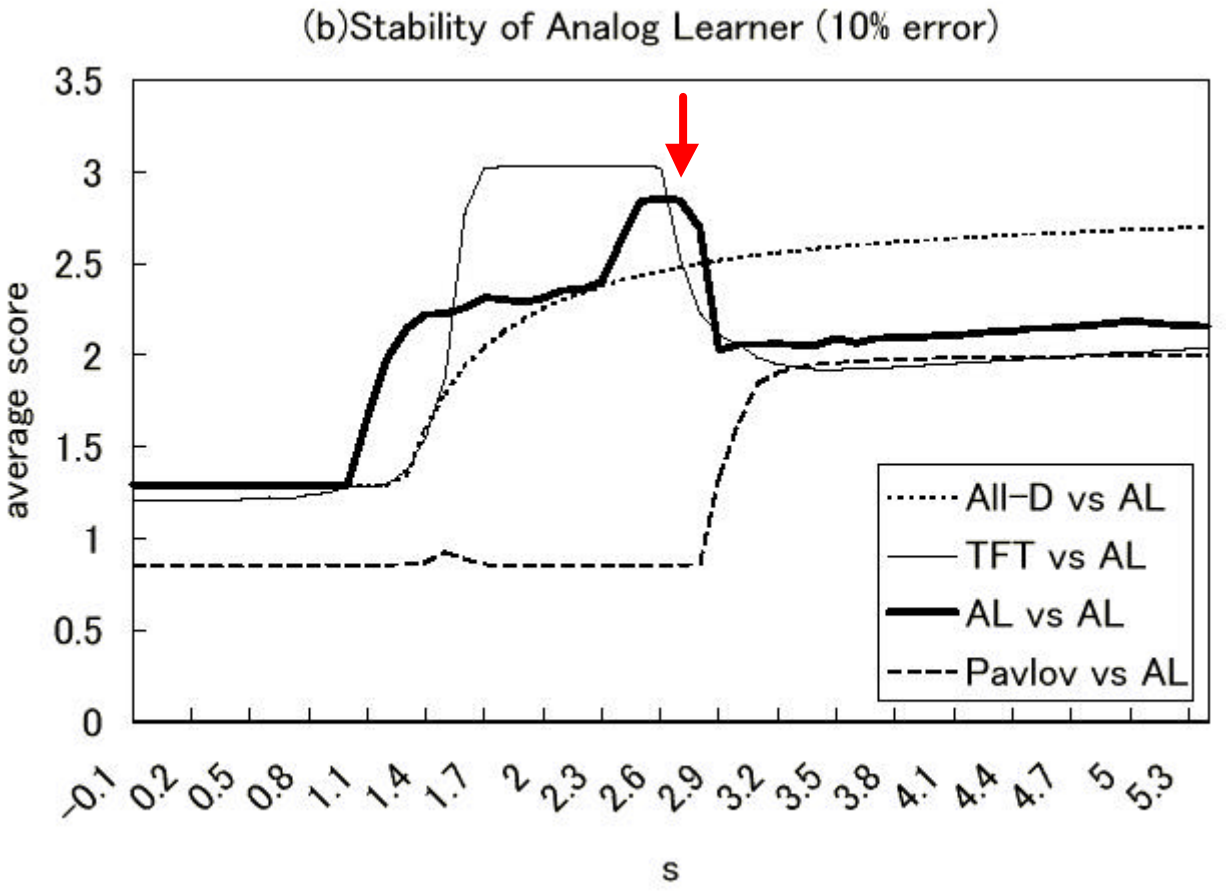
Analog Learner for s=1.6,…,3.0 invades All-C, All-D, TFT and Pavlov.

Can classic strategies invade Analog Learners ?



Analog Learner for s=2.6,…,3.0 is stable against invasion of All-C, All-D, TFT and Pavlov.

Can classic strategies invade Analog Learners ?



Analog Learner for s=2.7,…,2.8 is stable against invasion of All-C, All-D, TFT and Pavlov.

Summary of classic strategies vs. Analog Learner

Though Analog Learner cannot invade All-D population, cooperative variants of Analog Learner do very well against classic strategies.

1% error:

Analog Learner for s=2.6,…,3.0 invades and is stable against All-C, TFT, Pavlov (and All-D).

10% error:

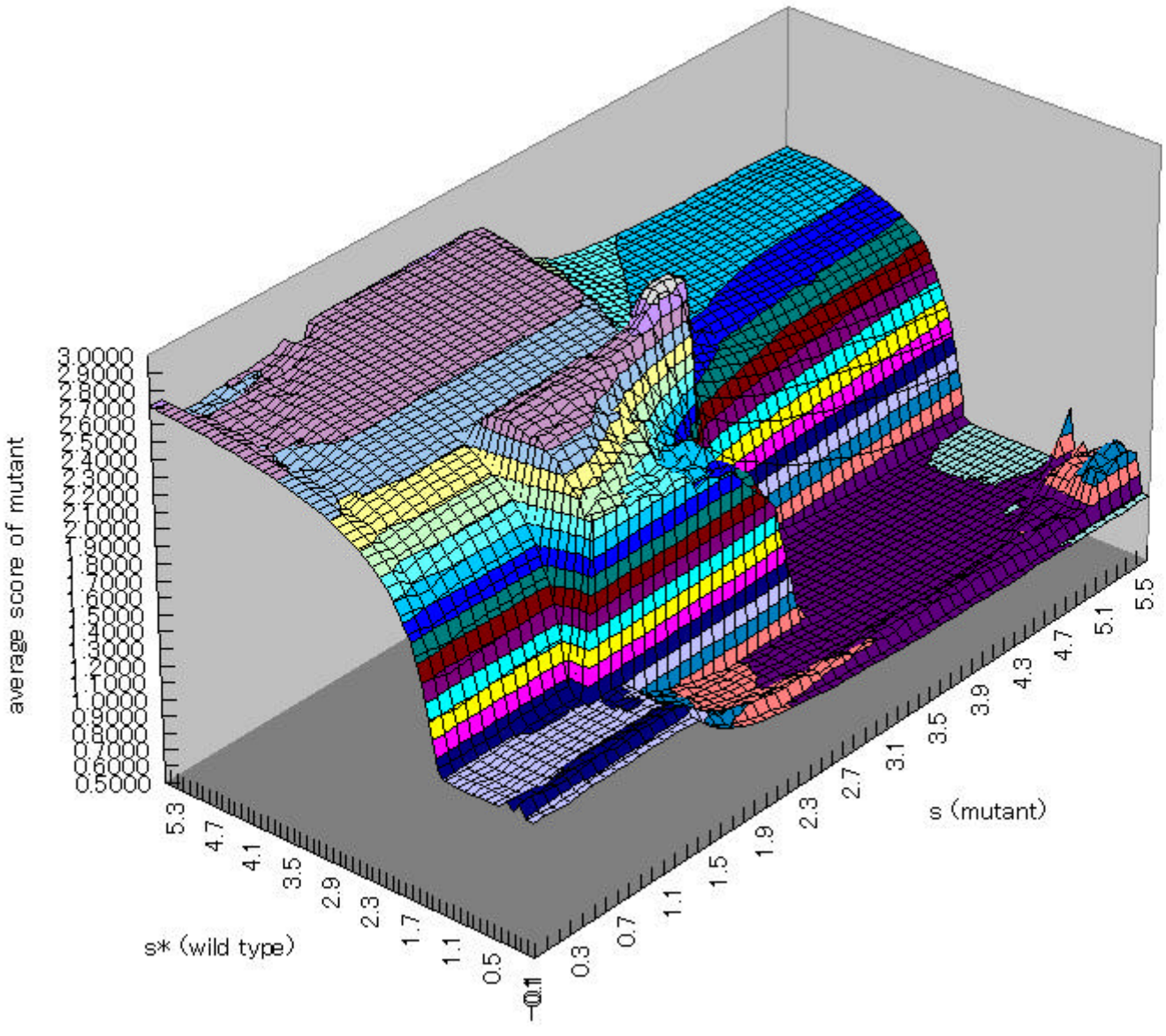
Analog Learner for s=2.7,…,2.8 invades and is stable against All-C, TFT, Pavlov (and All-D).

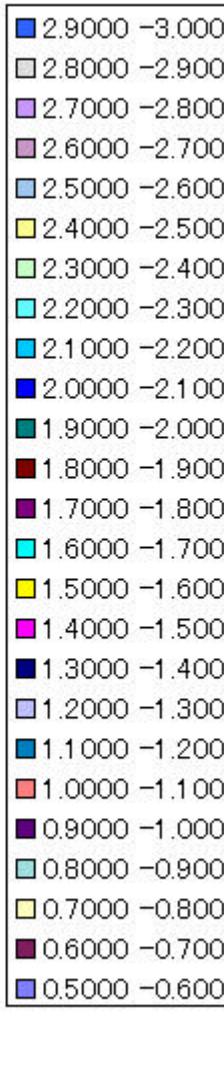
Evolutionary dynamics of aspiration level

F(s,s*):= payoff value of AL(s=s) when matched against AL(s=s*)

F(s,s*) is obtained by compter simulation. By calculating many sets of (s,s*), we obtain the payoff matrix (56 x 56 matrix).

Evolutionary dynamics of aspiration level



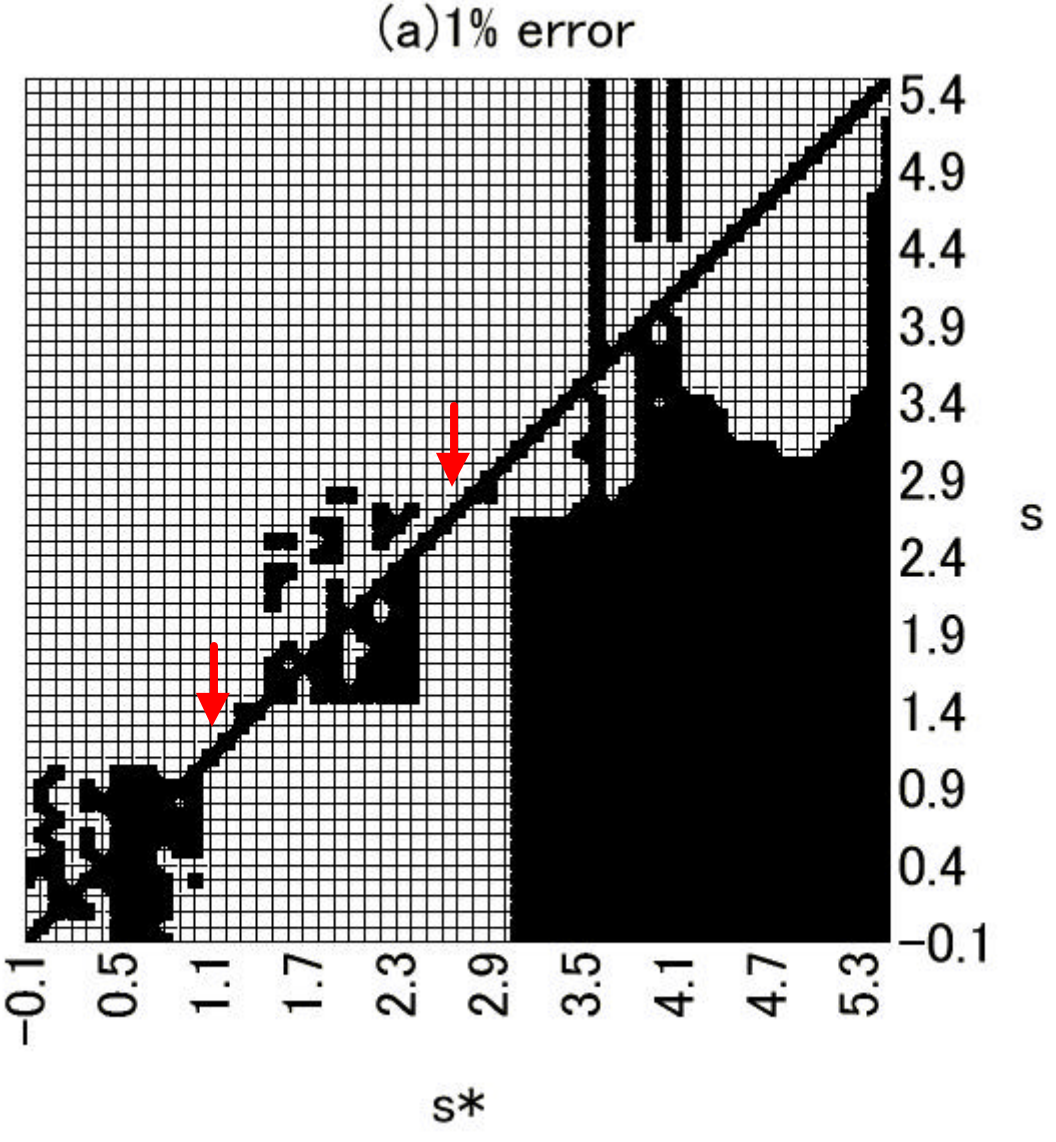


Evolutionary dynamics of aspiration level

From this payoff matrix, Pairwise Invasibility Plot (PIP) is drawn.

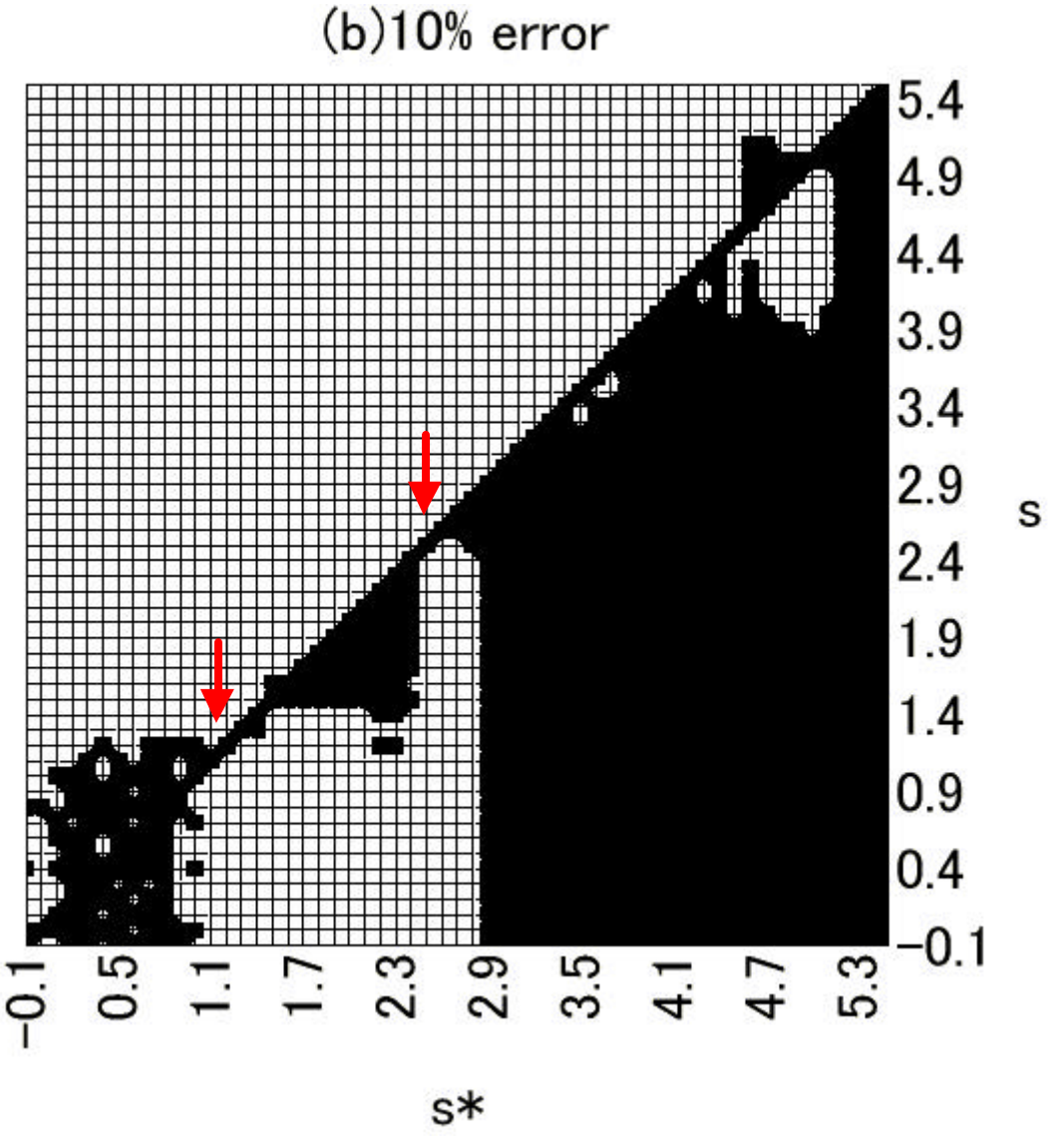
By surveying PIP, it becomes clear whether evolutionarily stable strategy (ESS) exists or not.

ESS among Analog Learners



Pairwise Invasibility Plot

ESS among Analog Learners



Pairwise Invasibility Plot

Analog Learner can establish robust cooperation

- Under both 1% and 10% error rate, Analog Learner with aspiration level around 2.7 can invade All-C, TFT or Pavlov populations and establish very robust mutual cooperation.

- Such population is more cooperative than Pavlov population.

F(2.7,2.7):
Average score of
AL(s=2.7) population

1%	2.9891
10%	2.8491

Average scores of populations for infinitely iterated Prisoners' Dilemma

error rate	All-C (1,1,1,1)	Pavlov (1,0,0,1)	TFT (1,0,1,0)	All-I (0,0,0,
0%	3.000	3.000	3.000	1.00
1%	2.990	2.951	2.010	1.03
5%	2.948	2.777	2.048	1.14
10%	2.890	2.602	2.090	1.29
20%	2.760	2.376	2.160	1.56

Due to strong faith in cooperation, it allows its partner's defection played by error without punishment.

Difference between Digital and Analog Learner

When matched against All-D (or DL(s<1) or AL(s<1));

	-1	1	-1	1	1
DL(1<s<3)	DD	CD	DD	CD	DD
AL(s=2)	DD	CD	DD	DD	CD
	-1	1	-3	-1	1

All-D strategy's payoff = { 3 (Digital Learner); 2.3 (Analog Learner)

∧

Average payoff of AL(s=2) population (=3)

Quantified evaluation ability is essential to avoid being exploited by always defecting strategies.

Future problems

Memory decay;
more recent result might affect internal state more strongly

Analog Learner with non-linear evaluation