Adaptive Learning Models

References

- Cheung, Y. and D. Friedman: Individual Learning in Normal Form Games: Some Laboratory Results, *Games and Economic Behavior*, 19 (1997), 46-76.
- Camerer, C. and T. Ho: Experience Weighted Attraction Learning in Normal Form Games, *Econometrica*, 67 (1999), 827-63.
- Ioannou, C., and J. Romero: A Generalized Approach to Belief Learning in Repeated Games, *Games and Economic Behavior*, 87 (2014), 178-203.

Motivation

- Which models describe human behavior best?
- Why is this important?
  - Design effective and efficient market mechanisms.
  - Apply counterfactual analysis.
  - Refine strategy sets.

Types of Adaptive Action Learning Models

- Belief-based Models
  - Fictitious Play - Brown (1951)
  - Cournot Best Response - Cournot (1960)
  - γ-Weighted Beliefs - Cheung & Friedman (1997)
- Reinforcement-based Models
  - Cumulative Reinforcement - Harley (1981), Erev & Roth (1995, 1998)
  - Averaged Reinforcement - Mookerjhee & Sopher (1994, 1997)
- Hybridized Models
  - Experience Weighted Attraction (EWA) - Camerer & Ho (1999)
  - Inertia, Sampling And Weighting (I-SAW) - Nevo & Erev (2012)

Preliminaries

- The set of players is denoted by $I = \{1, ..., n\}$.
- Each player $i \in I$ has an *action set* denoted by $\mathcal{A}_i$. An *action profile* $a = (a_i, a_{-i})$ consists of the action of player i and the actions of the other players, denoted by $a_{-i} = (a_1, ..., a_{i-1}, a_{i+1}, ..., a_n) \in \mathcal{A}_{-i}$.
- In addition, each player i has a real-valued, stage-game, payoff function $g_i : \mathcal{A} \to \mathbb{R}$, which maps every action profile $a \in \mathcal{A}$ into a payoff for i, where $\mathcal{A}$ denotes the cartesian product of the action spaces $\mathcal{A}_i$, written as $\mathcal{A} \equiv \underset{i=1}{\overset{I}{\times}} \mathcal{A}_i$.
- The indicator function $I\left(a_i^j, a_i(t)\right)$ equals 1 if $a_i^j = a_i(t)$ and 0 otherwise.
- In the infinitely-repeated game with *perfect monitoring*, the stage game in each time period $t = 0, 1, ...$ is played with the action profile chosen in period t publicly observed at the end of that period.

Preliminaries

- The *history* of play at time t is denoted by $h^t = (a^0, ..., a^{t-1}) \in \mathcal{A}^t$, where $a^r = (a_1^r, ..., a_n^r)$ denotes the actions taken in period r.
- The set of histories is given by

$$\mathcal{H} = \bigcup_{t=0}^{\infty} \mathcal{A}^t,$$

  where we define the initial history to the null set $\mathcal{A}^0 = \{\emptyset\}$.
- A *strategy* $s_i \in S_i$ for player i is, then, a function $s_i : \mathcal{H} \to \mathcal{A}_i$, where the strategy space of i consists of K_i discrete strategies; that is, $S_i = \{s_i^1, s_i^2, ..., s_i^{K_i}\}$.
- The set of joint-strategy profiles is denoted by $S = S_1 \times \cdots \times S_n$.
- Each player i has a payoff function $\pi_i^t : S \to \mathbb{R}$, which represents the average payoff per period.

Experience Weighted Attraction (EWA)

- Attraction for player i's action is updated as follows:

$$N(t) = \rho \cdot N(t-1) + 1,$$

$$A_i^j(t) = \frac{\phi \cdot N(t-1)\, A_i^j(t-1) + \left[\delta + (1-\delta)\, \mathbb{I}\left(a_i^j, a_i(t)\right)\right] g_i\left(a_i^j, a_{-i}(t)\right)}{N(t)}$$

  - ϕ - discount factor on previous attractions
  - ρ - discount factor on previous experience
  - δ - weight on forgone (hypothetical) payoffs

- Players choose an action each period with

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}}.$$

Nested Models

ϕ	δ	ρ	$N(0)$	**Learning Model**
1	1	1	-	Fictitious Play
0	1	0	-	Cournot Best-Response
$\phi \in (0,1)$	1	ϕ	-	Weighted Fictitious Play
$\phi \in [0,1]$	0	0	1	Cumulative Reinforcement
$\phi \in [0,1]$	0	ϕ	$\frac{1}{1-\phi}$	Averaged Reinforcement

- EWA fits experimental data better than belief-learning and reinforcement-learning models in:
  - constant-sum games with a unique mixed equilibrium,
  - coordination games with multiple Pareto-ranked equilibria, and
  - beauty contest games with a unique dominance-solvable equilibrium.

Self-Tuning EWA

- Attraction for player i's action is updated as follows:

$$N(t) = \phi_i(t) \cdot N(t-1) + 1,$$

$$A_i^j(t) = \frac{\phi_i(t) \cdot N(t-1)\, A_i^j(t-1) + \left[\delta_i^j + \left(1 - \delta_i^j\right) I\left(a_i^j, a_i(t)\right)\right] g_i\left(a_i^j, a_{-i}(t)\right)}{N(t)}$$

- The model makes parameters ϕ, δ, and ρ self-tuning functions.
- Players choose an action each period with

$$P_i^j(t+1) = \frac{e^{\lambda \cdot A_i^j(t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(t)}}.$$

Self-Tuning Functions

- The attention function, $\delta_i^j(t)$, is

$$\delta_i^j(t) = \begin{cases} 1 & \text{if } g_i\left(a_i^j, a_{-i}(t)\right) \geq g_i(t) \\ 0 & \text{otherwise.} \end{cases}$$

- The decay function, $\phi_i(t)$, consists of
  - the cumulative history vector across the other players' actions k, which records the historical frequencies (including the last period t) of the choices by other players

$$\sigma(a_{-i}^k, t) = \frac{1}{t}\sum_{\tau=1}^{t} I\left(a_{-i}^k, a_{-i}(\tau)\right),$$

  - so that the surprise index is

$$\mathcal{S}_i(t) = \sum_{k=1}^{m_{-i}} (\sigma(a_{-i}^k, t) - r(a_{-i}^k, t))^2.$$

  - Finally, the decay function is

$$\phi_i(t) = 1 - \frac{1}{2}\mathcal{S}_i(t).$$

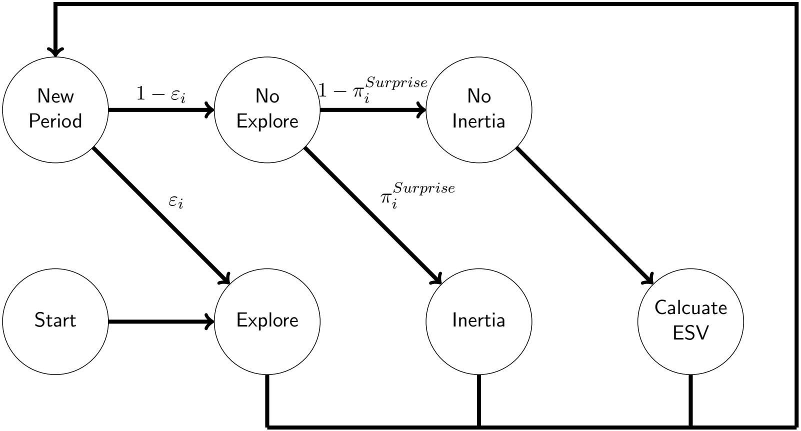
Inertia, Sampling and Weighting (I-SAW)

- I-SAW is an instance-based model, which allows for three response modes: *exploration, inertia and exploitation.*
- In each period, a player enters one of the modes with different probabilities.
- There are n players in the game.
- Each player has a set of parameters $(p_A, \varepsilon_i, \pi_i, \mu_i, \rho_i, \omega_i)$.
- The parameter $p_A \in [0, 1]$ is the same for all agents.
- The other parameters are idiosyncratic with $\varepsilon_i \sim U\,[0, \varepsilon]$, $\pi_i \sim U\,[0, \pi]$, $\mu_i \sim U\,[0, \mu]$, $\rho_i \sim U\,[0, \rho]$ and $\omega_i \sim U\,[0, \omega]$.

Inertia, Sampling and Weighting (I-SAW)

- For simplicity, assume two actions.
- Player i's action set is $\mathcal{A}_i = \{A, B\}$.
- Let a_i^t be the action of player i that was played in period t, where $h_i\left(t_1, t_2\right) = \left\{a_i^{t_1}, a_i^{t_1+1}, \ldots, a_i^{t_2}\right\}$ for $t_1 \leq t_2$.
- Similarly, let a_{-i}^t be the actions of players other than i in period t, where $h_{-i}\left(t_1, t_2\right) = \left\{a_{-i}^{t_1}, a_{-i}^{t_1+1}, \ldots, a_{-i}^{t_2}\right\}$ for $t_1 \leq t_2$.
- We explain next the three response modes.

Schematic of I-SAW



Exploration

- In exploration, each player chooses action A with probability p_A and action B with probability $1 - p_A$. The probabilities are the same for all players.

Inertia

- A player might enter the inertia mode after period 2 with probability $\pi_i^{Surprise(t)}$, where $\pi_i \in [0, 1]$ and $Surprise(t) \in [0, 1]$. Specifically,

$$Gap(t) =$$

$$\frac{1}{4}\left[\sum_{j=1}^{2} |Obtained_j(t-1) - Obtained_j(t)| + \sum_{j=1}^{2} |GMean_j(t) - Obtained_j(t)$$

$$Surprise(t) = \frac{Gap(t)}{MeanGap(t) + Gap(t)}$$

$$MeanGap(t+1) = MeanGap(t)(1 - \frac{1}{r}) + Gap(t)\frac{1}{r}$$

where r is the expected number of trials in the experiment.

Exploitation

- In exploitation trials, an individual selects the action with the highest Estimated Subjective Value (ESV).
- To determine the ESV, player i randomly selects μ_i elements from $h_{-i}(0, t-1)$ with replacement; let us call this set $M_{-i}(0, t-1)$.
- This set is chosen according to the following: with probability ρ_i player chooses a_{-i}^{t-1} and with probability $1-\rho_{-i}$ player chooses uniformly over $h_{-i}(0, t-1)$.
- The same set M_{-i} is used for each $a_i \in \mathcal{A}_i$.
- The sample mean for action a_i' is then defined as

$$SampleM(a_i', t) = \frac{1}{|M_{-i}(0, t-1)|} \sum_{a_{-i} \in M_{-i}(0,t-1)} g_i\left(a_i', a_{-i}\right).$$

Exploitation

- The $GrandM(a_i', t)$ is defined as

$$GrandM(a_i', t) = \frac{1}{|h_{-i}(0, t-1)|} \sum_{a_{-i} \in h_{-i}(0,t-1)} g_i(a_i', a_{-i}).$$

- Then, player i's ESV of action a_i' is

$$ESV(a_i') = (1 - \omega_i) \cdot SampleM(a_i', t) + \omega_i \cdot GrandM(a_i', t),$$

  where ω is the weight assigned on the payoff based on the entire history ($GrandM$) and $1 - \omega$ is the weight assigned on the payoff based on the sample from the history ($SampleM$).

- Then, the player simply chooses the a_i' that maximizes ESV (and chooses randomly in ties).

γ-Weighted Beliefs

- A player updates his beliefs on the opponent's actions with parameter γ; that is,

$$b_i\left(a_{-i}^k, t\right) = \frac{\mathbb{I}\left(a_{-i}^k = a_{-i}(t)\right) + \sum_{r=1}^{t-1} \gamma^{t-r}\mathbb{I}\left(a_{-i}^k = a_{-i}(r)\right)}{1 + \sum_{r=1}^{t-1} \gamma^{t-r}}.$$

- Setting $\gamma = 0$ yields the Cournot learning rule, and setting $\gamma = 1$ yields fictitious play.

- We have adaptive learning when $0 < \gamma < 1$.

- He then calculates the expected payoff of action a_i^j as

$$E_i(a_i^j, t) = \sum_{k=1}^{m_{-i}} b_i\left(a_{-i}^k, t\right) g_i\left(a_i^j, a_{-i}^k\right).$$

γ-Weighted Beliefs

- Finally, an action is selected via the logit specification with parameter λ; that is, the probability of choosing action a_i^j in period $t+1$ is

$$\mathbb{P}_i\left(a_i^j, t+1\right) = \frac{e^{\lambda \cdot E_i\left(a_i^j, t\right)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot E_i\left(a_i^k, t\right)}}.$$

Why Not Repeated-Game Strategy Learning Models?

> *Incorporating a richer specification of strategies is important because stage-game strategies are not always the most natural candidates for the strategies that players learn about.*
>
> *Camerer and Ho (1999)*

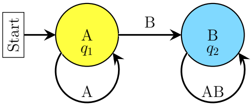
1. The set of possible strategies in repeated games is infinite.
2. Beliefs become more complex because several different strategies can lead to the same history.
3. Repeated-game strategies need several periods to be evaluated.

A Generalized Approach

- We propose novel rules to facilitate operability of belief learning models with repeated-game strategies.
- The impact of the rules is assessed on a self-tuning Experience Weighted Attraction model, a γ-Weighted Beliefs model and an Inertia, Sampling and Weighting model.
- The predictions of the three models with strategy learning are validated with experimental data in four symmetric games.
- We also validate the predictions of their respective models with actions.
- The strategy learning models approximate subjects' behavior substantially better than their respective action learning models.

Deterministic Finite Automata (DFA)

- A *Moore Machine* is a four-tuple $\left(Q_i, q_i^0, f_i, \tau_i\right)$ where
  - Q_i is a finite set of states,
  - q_i^0 is the initial state,
  - $f_i : Q_i \rightarrow \mathcal{A}_i$ is an output function, and
  - $\tau_i : Q_i \times \mathcal{A}_{-i} \rightarrow Q_i$ is a transition function.



$$Q_i = \{q_1, q_2\}$$

$$q_i^0 = q_1$$

$$f_i(q) = \begin{cases} A & \text{if } q = q_1 \\ B & \text{if } q = q_2 \end{cases}$$

$$\tau_i(q, a_{-i}) = \begin{cases} q_1 & \text{if } (q, a_{-i}) = (q_1, A) \\ q_2 & \text{otherwise} \end{cases}$$

Fitness Function

- Define the fitness function $\mathcal{F} : S_{-i} \times \mathbb{N} \to [0, \mathcal{T}_i(\chi)]$ as

$$\mathcal{F}\left(s_{-i}, \chi\right) = \max\left\{ t' | s_{-i} \text{ is consistent with } h^{\mathcal{T}_i(\chi)} \text{ for the last } t' \text{ periods} \right\}.$$

- Define the belief function $\mathcal{B} : S_{-i} \times \mathbb{N} \to [0, 1]$ as

$$\mathcal{B}(s_{-i}, \chi) = \frac{\mathcal{F}(s_{-i}, \chi)}{\sum_{r \in S_{-i}} \mathcal{F}(r, \chi)},$$

- which can be interpreted as player i's belief that the other player was using strategy s_{-i} at the end of block χ.

Asynchronous Updating of Strategies

- Players update their strategies with the completion of a block of periods.

- The probability of updating depends on the expected block length, which is calculated *at the end of each period*.

    - The probability that player i updates his strategy in period t, $\frac{1}{\mathcal{P}_i^t}$, is therefore determined endogenously via the expected length of the block term, $\mathcal{P}_i^t$, which is updated recursively:

$$\mathcal{P}_i^t = \mathcal{P}_i^{t-1} - \frac{1}{\mathcal{P}_i^{t-1}} \frac{\left| \frac{1}{t-\underline{t}(\chi(t))} \sum_{s=\underline{t}(\chi(t))}^{t-1} g_i(a_i^s, a_{-i}^s) - \mathcal{E}_i^{s_i(\chi(t))}(\chi(t)) \right|}{\bar{g} - \underline{g}}.$$

Self-tuning EWA with Strategies

The model consists of two variables:

1. $N_i(\chi)$ is interpreted as the number of observation-equivalents of past experience in block χ of player i, and
2. $A_i^j(\chi)$ indicates player i's attraction to strategy j *after* the χ^{th} block of periods.

The evolution of learning over the χ^{th} block with $\chi \geq 1$ is governed by the following rules:

$$N_i\left(\chi\right) = \phi_i(\chi) \cdot N_i\left(\chi - 1\right) + 1,$$

and

$$A_i^j\left(\chi\right) = \frac{\phi_i(\chi) \cdot N_i\left(\chi - 1\right) \cdot A_i^j\left(\chi - 1\right) + \mathbb{I}(s_i^j, s_i\left(\chi\right)) \cdot R_i\left(\chi\right) + \delta_i^j(\chi) \cdot \mathcal{E}_i^j\left(\chi\right.}{N_i\left(\chi\right)}$$

- Attractions are updated by

$$A_i^j(\chi) = \frac{\phi_i(\chi) \cdot N_i(\chi - 1) \cdot A_i^j(\chi - 1) + \mathbb{I}(s_i^j, s_i(\chi)) \cdot R_i(\chi) + \delta_i^j(\chi) \cdot \mathcal{E}_i^j(\chi}{\phi_i(\chi) \cdot N_i(\chi - 1) + 1}$$

  - χ is the block index,
  - $R_i(\chi)$ is the reinforcement payoff from block χ,
  - $\mathcal{E}_i^j(\chi)$ is the forgone payoff from block χ,
  - $\delta_i^j(\chi)$ is the self-tuning attention function, and
  - $\phi_i(\chi)$ is the self-tuning decay function.

- Players choose strategy at the beginning of block $\chi + 1$ with

$$\mathbb{P}_i^j(\chi + 1) = \frac{e^{\lambda \cdot A_i^j(\chi)}}{\sum_k^K e^{\lambda \cdot A_i^k(\chi)}}.$$

Reinforcement and Foregone Payoffs

- Reinforcement Payoffs

$$R_i\left(\chi\right)=\frac{1}{T_i(\chi)}\sum_{a\in h(\chi)}g_i\left(a\right)$$

- Forgone Payoffs
  - Define the fitness function

$$\mathcal{F}\left(s_{-i},\chi\right)=\max\left\{t'|s_{-i}\text{ is consistent with }h(\chi)\text{ for the last }t'\text{ periods}\right\}.$$

  - The belief of player i that the other played s_{-i} when the history was h and player i played s_i, is

$$\mathcal{B}(s_{-i},\chi)=\frac{\mathcal{F}(s_{-i},\chi)}{\sum_{r\in S_{-i}}\mathcal{F}(r,\chi)}.$$

  - The forgone payoff is

$$\mathcal{E}_i^j\left(\chi\right)=\sum_{s_{-i}\in S_{-i}}\pi_i(s_i^j,s_{-i})\cdot\mathcal{B}(s_{-i},\chi).$$

Self-tuning Functions

- The attention function is

$$\delta_i^j(\chi) = \begin{cases} 1 & \text{if } \mathcal{E}_i^j(\chi) \geq R_i(\chi) \text{ and } s_i^j \neq s_i(\chi) \\ 0 & \text{otherwise.} \end{cases}$$

- The decay function, $\phi_i(\chi)$, consists of
  - the averaged belief which is

$$\sigma(s_{-i}, \chi) = \frac{1}{\chi} \sum_{j=1}^{\chi} \mathcal{B}(s_{-i}, j),$$

  - so that the surprise index is

$$\mathcal{S}_i(\chi) = \sum_{s_{-i} \in S_{-i}} (\sigma(s_{-i}, \chi) - \mathcal{B}(s_{-i}, \chi))^2.$$

  - The decay function is

$$\phi_i(\chi) = 1 - \frac{1}{2} \mathcal{S}_i(\chi).$$

γ-Weighted Beliefs with Strategies

- The attractions in this model evolve according to the following two rules and parameter γ:

$$N_i(\chi) = \gamma \cdot N_i(\chi - 1) + 1,$$

and

$$A_i^j(\chi) = \frac{\gamma \cdot N_i(\chi - 1) \cdot A_i^j(\chi - 1) + \mathcal{E}_i^j(\chi)}{N_i(\chi)}.$$

- A player chooses strategy at the beginning of block $\chi + 1$ with

$$\mathbb{P}_i^j(\chi + 1) = \frac{e^{\lambda \cdot A_i^j(\chi)}}{\sum_k^K e^{\lambda \cdot A_i^k(\chi)}}.$$

I-SAW with Strategies

- Exploration: A strategy is randomly selected.
- Inertia: The probability is $(1 - \frac{1}{\mathcal{P}_i^t})$ rather than $\pi_i^{Surprise}$.
- Exploitation: The grand mean is

$$GrandM_i\left(s^j, \chi\right) = \frac{1}{\chi} \sum_{k=1}^{\chi} \mathcal{E}_i^j(k).$$

Let $M_i(\chi)$ be a set of μ_i numbers drawn with replacement from $\{1, 2, \ldots, \chi\}$. Then, the sample mean is

$$SampleM_i\left(s^j, \chi\right) = \frac{1}{|M_i(\chi)|} \sum_{k \in M_i(\chi)} \mathcal{E}_i^j(k).$$

Finally, the ESV is calculated as

$$ESV_i\left(s^j, \chi\right) = (1 - \omega_i) \cdot SampleM_i\left(s^j, \chi\right) + \omega_i \cdot GrandM\left(s^j, \chi\right).$$

SIMULATIONS

- In the first (pre-experimental) phase, players engage in a lengthy process of learning among strategies.
- Each pair of agents stay matched until the average payoff of the given pair converges.
- The second (experimental) phase consists of a fixed-pair matching of 100 periods.
- The results are averaged over 1,000 simulations.

	A	B
A	3,3	1,4
B	4,1	2,2

Prisoner's Dilemma

	A	B
A	1,1	2,4
B	4,2	1,1

Battle of the Sexes

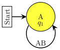
	A	B
A	3,3	0,2
B	2,0	1,1

Stag-Hunt

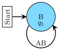
	A	B
A	3,3	1,4
B	4,1	0,0

Chicken

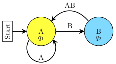
Strategy Set



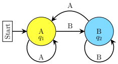
Automaton 1



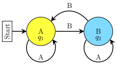
Automaton 2



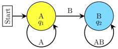
Automaton 3



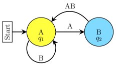
Automaton 4



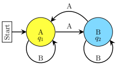
Automaton 5



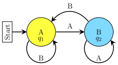
Automaton 6



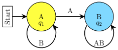
Automaton 7



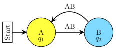
Automaton 8



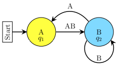
Automaton 9



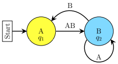
Automaton 10



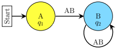
Automaton 11



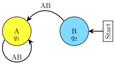
Automaton 12



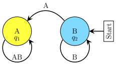
Automaton 13



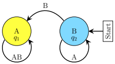
Automaton 14



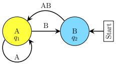
Automaton 15



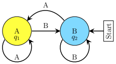
Automaton 16



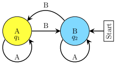
Automaton 17



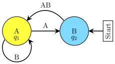
Automaton 18



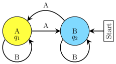
Automaton 19



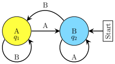
Automaton 20



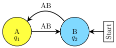
Automaton 21



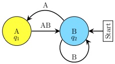
Automaton 22



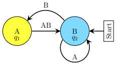
Automaton 23



Automaton 24



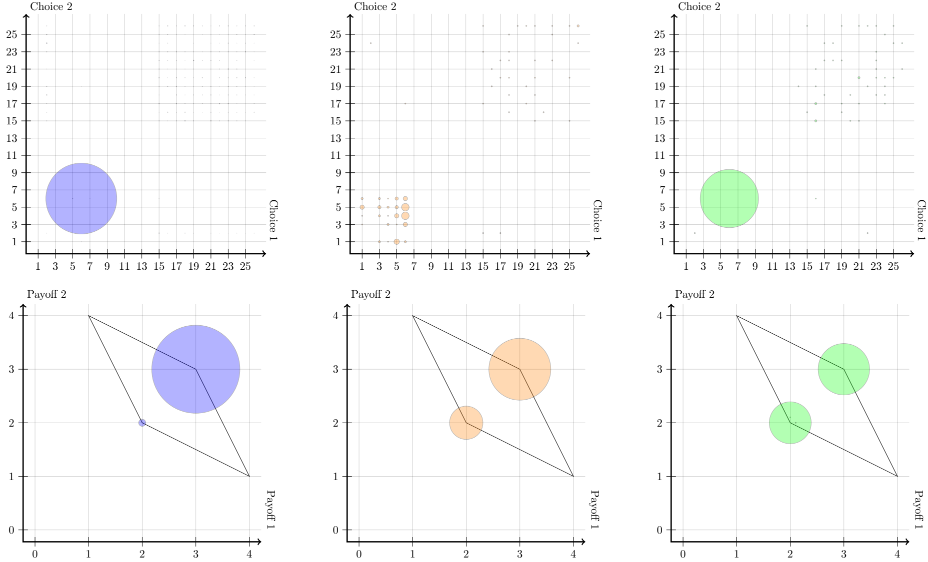
Automaton 25



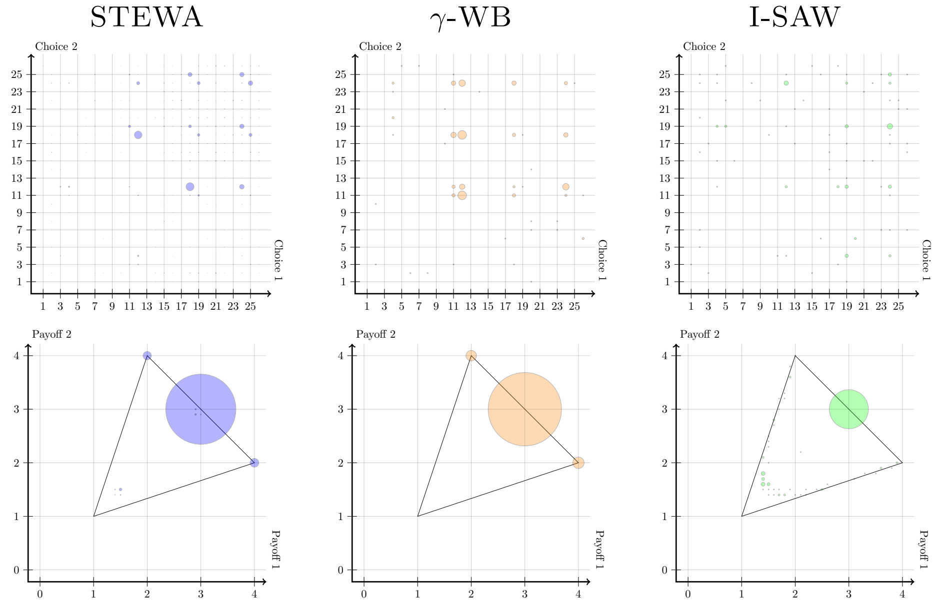
Automaton 26

Prisoner's Dilemma

STEWA | γ-WB | I-SAW

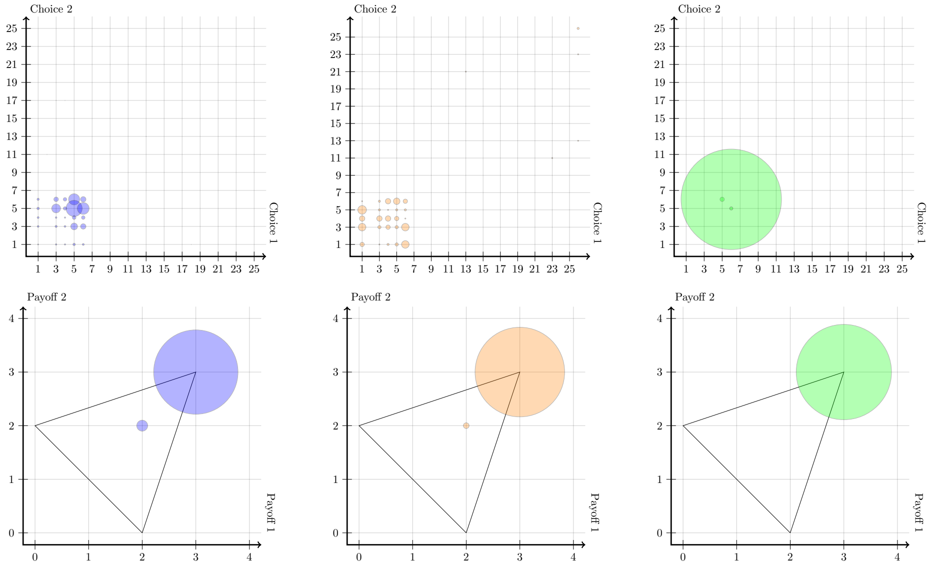


Battle of the Sexes

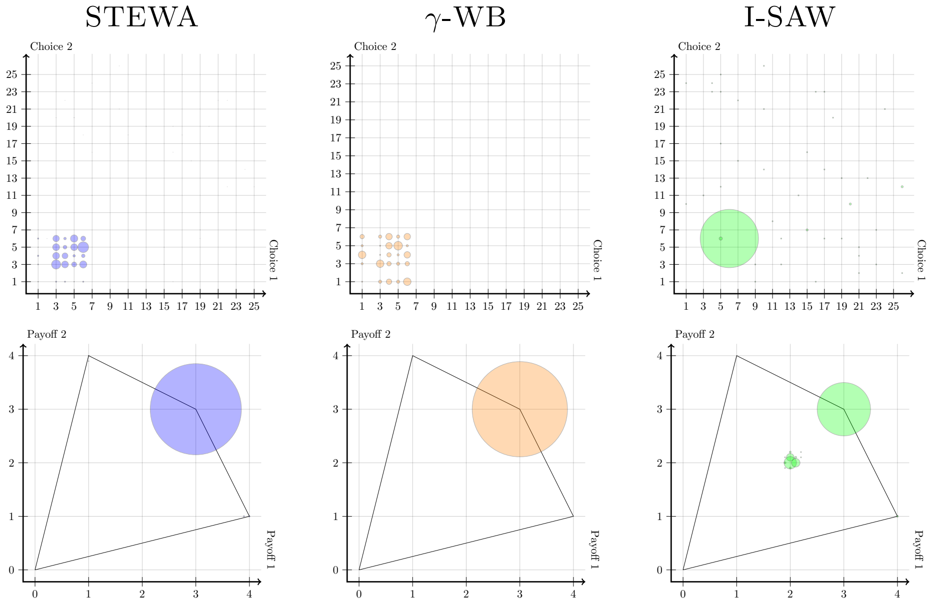


Stag-Hunt

STEWA γ-WB I-SAW



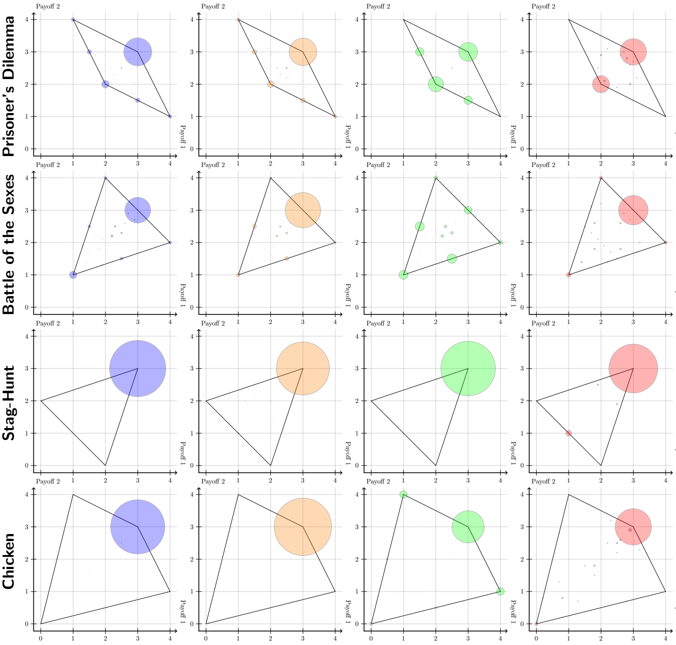
Chicken



Phase II

Strategy Learning

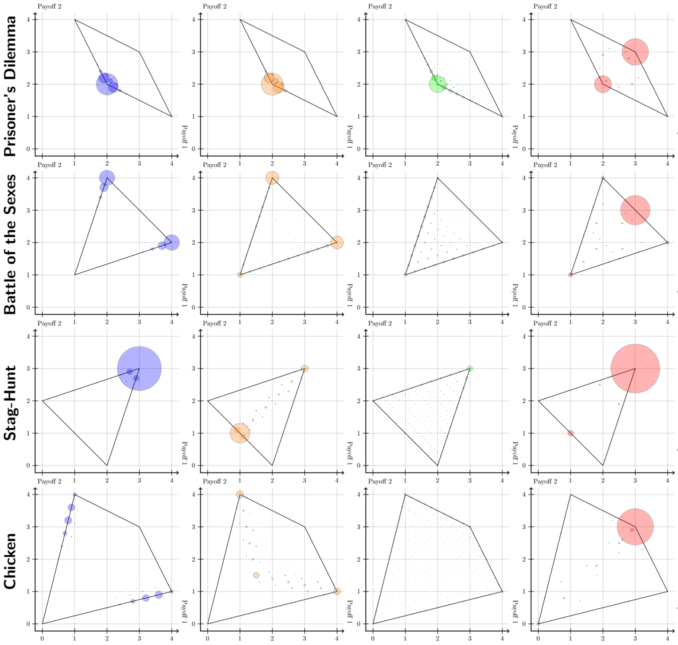
- STEWA
- γ-WB
- I-SAW
- Human Subjects



Phase II

Action Learning

- STEWA
- γ-WB
- I-SAW
- Human Subjects



Christos A. Ioannou

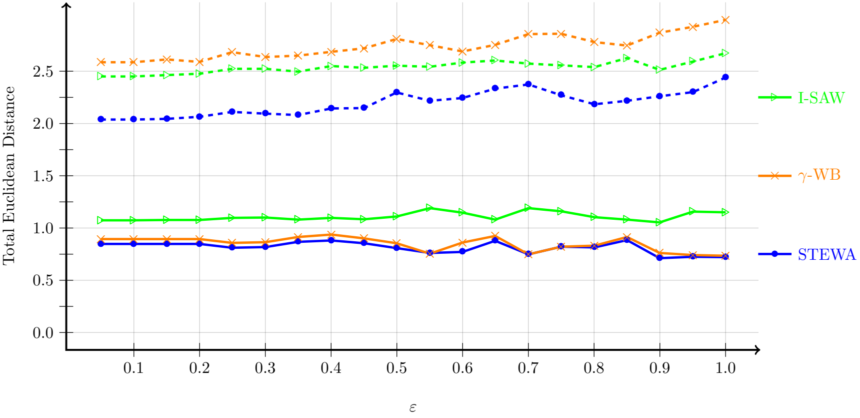
Which is Better? Formal Evidence

$$D(\pi) = \varepsilon \left\lfloor \frac{\pi}{\varepsilon} \right\rceil$$

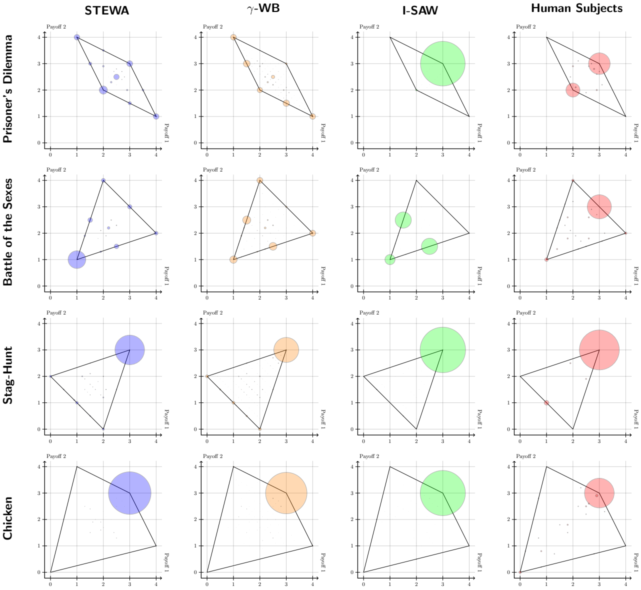
- π is the payoff,
- ε is the accuracy of the discretization,
- $D(\pi)$ denotes the transformed payoff, and
- $\lfloor \cdot \rceil$ rounds the fraction to the nearest integer.

	STEWA		γ-WB		I-SAW	
Game	**Action**	**Strategy**	**Action**	**Strategy**	**Action**	**Strategy**
PD	0.532	0.228	0.535	0.237	0.486	0.257
BoS	0.642	0.126	0.580	0.153	0.531	0.455
SH	0.191	0.165	0.782	0.122	0.614	0.140
CH	0.698	0.326	0.679	0.381	0.773	0.221
Total	2.064	0.846	2.575	0.893	2.404	1.073

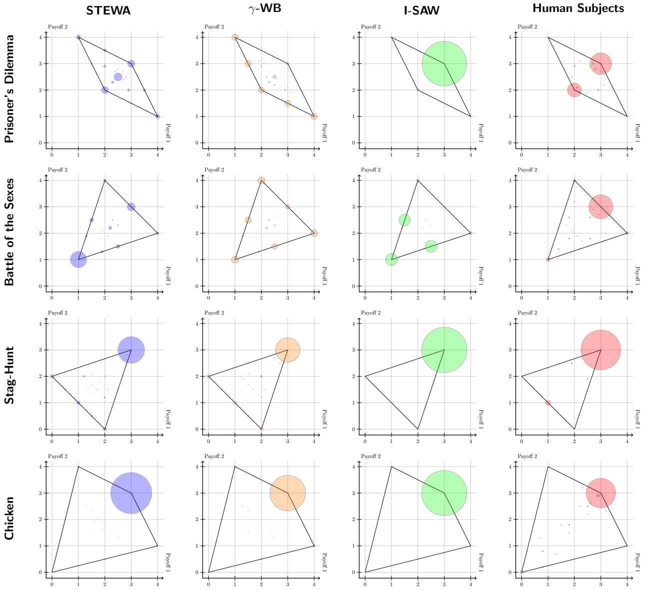
Robustness Check



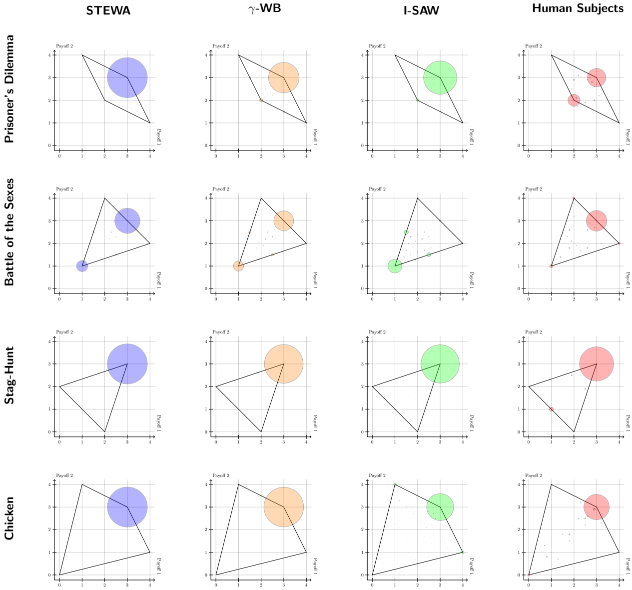
Fitness Function 1



Fitness Function 2

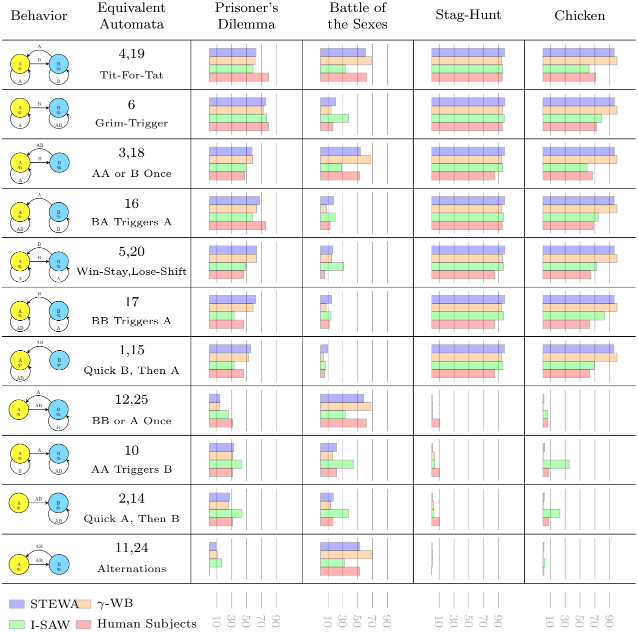


Synchronous Updating of Strategies



Christos A. Ioannou

Inferred Rules of Behavior



Contribution

Our suggested modeling framework:

1. accommodates a richer specification of strategies,
2. is consistent with belief-learning,
3. allows asynchronous updating of strategies,
4. is flexible enough to incorporate larger strategy sets, and
5. makes no a priori assumptions on social preferences.

Future Work

- Incomplete information
- Random matching protocols
- Cut-off strategies
- Asymmetric payoffs
- Small amounts of errors