Learning by Design

Peyton Young
University of Oxford

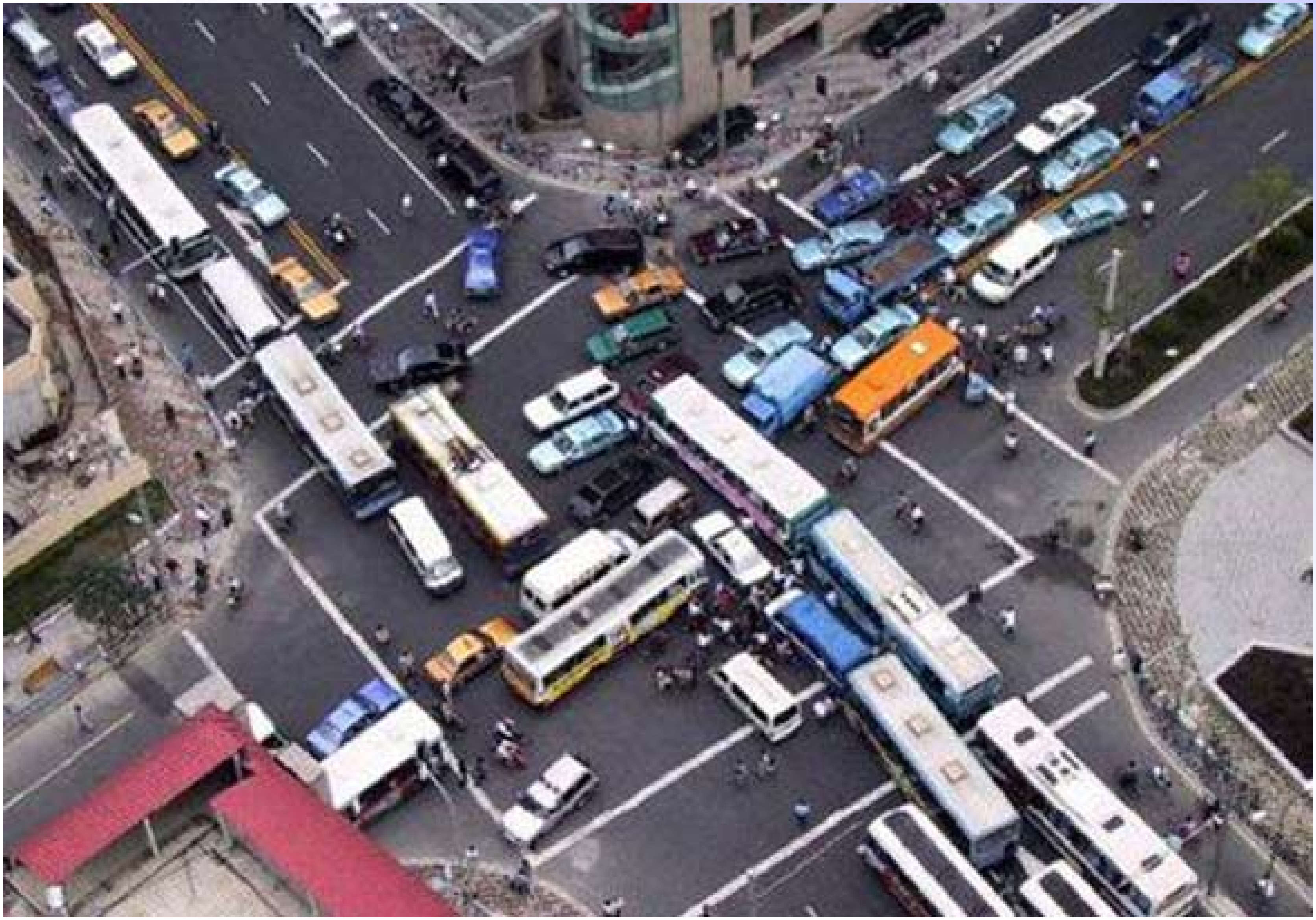
"One of the most important contributions of game theory is the game matrix"..Tom Schelling

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

But while good for pedagogical purposes, the game matrix can lead to delusional thinking by game theorists about how real games should be modeled

It tends to make us believe that:

- The number of players is small
- The rules are common knowledge
- The payoffs, or at least the distribution of possible payoffs, is common knowledge
- Everyone is rational



- *Game Engineering* focuses attention on dynamical systems with large numbers of interacting agents

- The game theorist gets to *choose* the agents' learning rules and even their utility functions, so agents have high or low rationality *by design*

- Key problem is to characterize the *dynamics of adjustment* given the agents' learning rules

We shall demonstrate a family of adaptive learning rules that are extremely simple to implement and result in a probability distribution over outcomes that place high probability on (some of) the Nash equilibria

Payoff-Based Learning with Experimentation

- Agents adjust their behavior only in response to their *own realized payoffs* (rules are *completely uncoupled)*
- Agents have *no knowledge* of the overall structure of the game
- They *do not observe* the actions or payoffs of most other players – perhaps *any* other players
- They occasionally *experiment* with new strategies

Applications

- Choosing traffic routes
- Routing packets between information processors in large networks
- Designing sensors for coordinated search

Notation

Players $i = 1, 2, \ldots, n$

Action spaces A_i

Joint action space $A = \prod_i A_i$

Utility functions $u_i : A \to R$

Rule I : Simple Experimentation

Time is discrete : $t = 1, 2, 3, ...$

At time t the state of player i is a pair

$$z_i(t) = (\overline{a}_i(t), \overline{u}_i(t))$$

$\overline{a}_i(t)$ is a benchmark action

$\overline{u}_i(t)$ is a benchmark payoff level (aspiration level)

At time $t+1$

Agent i experiments with probability $0 < \varepsilon < 1$

Not experiment $\Rightarrow$ play current benchmark

Experiment $\Rightarrow$ play $a_i \in A_i$ drawn uniformly at random

Realized payoff $> \bar{u}_i(t) \Rightarrow i$ adopts experimental action and corresponding payoff as new benchmarks

Otherwise i keeps the previous benchmarks

A game G on A is *weakly acyclic* if from every n-tuple of actions $\vec{a} \in A$ there exists a better reply path -- one player moving at a time -- to a pure Nash equilibrium of G

G is a *potential game* if there exists a function $\phi : A \rightarrow R$ such that for all i, a_i, a_i', a_{-i}

$$\phi(a_i', a_{-i}) - \phi(a_i, a_{-i}) = u_i(a_i', a_{-i}) - u_i(a_i, a_{-i})$$

Every potential game is weakly acyclic

- In weakly acyclic games, the better reply dynamic converges with probability one to a pure Nash equilibrium
- But this assumes that each player *knows* what actions the other players are taking
- Simple experimentation assumes only that players react to their *own* recent payoffs.

Theorem 1. *Let G be a finite n-person weakly acyclic game. When all players learn by simple experimentation and the rate of experimentation is sufficiently small, a Nash equilibrium is played a high proportion of the time.*

Marden, Young, Arslan, and Shamma, "Payoff-based dynamics for multi-player weakly acyclic games," SIAM Journal on Control and Optimization, 2009

Rule II: Regret Testing

- A player decides to start experimenting in any given period with small probability ε (provided he is not already experimenting)
- Experimentation involves computing the average payoff from a *random sample* of s recent plays and comparing it with the average payoff from the current mixed (or pure) strategy

- If the sample average is better by some small amount $\tau > 0$ (the *tolerance level*), the player switches to the better strategy with high probability, and switches to a new strategy with small probability, where the new strategy is drawn uniformly at random fr

- The new strategy is drawn uniformly at random from the space of discretized mixed strategies (probabilities are rounded to some finite number of decimal places d, and in particular contain all pure strategies)

Theorem 2. *Let G be an n-person normal form game with generic payoffs and let $\delta > 0$. If the tolerance τ and experimentation rate ε are sufficiently small, the sample size s is sufficiently large, and the discretization of the state space is sufficiently fine (large d), then a δ–equilibrium of G is played at least $1 - \delta$ of the time.*

Foster and Young, Theoretical Economics, 2006

Germano and Lugosi, GEB, 2007

Rule III: Interactive Trial and Error Learning

- Search behavior depends on a player's "mood"
- If *content*, experiment with probability $\varepsilon > 0$.
- If the experiment results in a higher payoff, adopt the new action and payoff as the new benchmarks; otherwise remain content with the previous benchmarks

- If *discontent*, flail around: choose an action at random
- Become spontaneously content again with a probability that increases in realized payoff
- Otherwise stay discontent and flail some around some more

Let Γ_A be the set of all n-person games on finite action space A that possess at least one pure Nash equilibrium.

Theorem 2. *For almost all games in* Γ_A *, if all players use interactive trial and error learning with sufficiently small experimentation rate* $\varepsilon > 0$*, then they play a Nash equilibrium an arbitrarily high proportion of the time.*

H.P. Young, Learning by Trial and Error, Games and Economic Behavior, vol. 65, 2009.

Open problems

- Can one design learning algorithms that get close to equilibrium quickly in important subclasses of games?
- Is Nash equilibrium the right concept for predicting long-run behavior in large systems of interacting agents?

  Perhaps correlated equilibrium or an even weaker notion of equilibrium is more useful as a solution concept for such games.

For a survey of some of the literature see *Strategic Learning and Its Limits*, Oxford University Press, 2004