Learning How to Behave in an Unfamiliar Society

Dana Nau

Laboratory for Computational Cultural Dynamics

University of Maryland

College Park, MD, USA

Behavior in an Unfamiliar Society

- Suppose you enter an environment that's inhabited by agents who are unfamiliar to you
- You know what actions are *possible*
- But you don't know
  - » What behaviors and outcomes the agents prefer, and why
  - » How they're likely to react to your actions
  - » What collection of behaviors is most likely to elicit the responses *you* prefer
- How can you decide how to behave?
- **First, a simple example**
  - » The game we asked you to play when you registered today

ICCD 2008

Please help us by playing a game:

- Choose a number in the range from 0 to 100, and write it in the space below.
- We'll take the average of all of the numbers. The winner(s) will be whoever chose a number that's closest to 2/3 of the average.
- Dana Nau will announce the results during his talk this afternoon.

Your number: ________________________

Your name (optional): ________________________

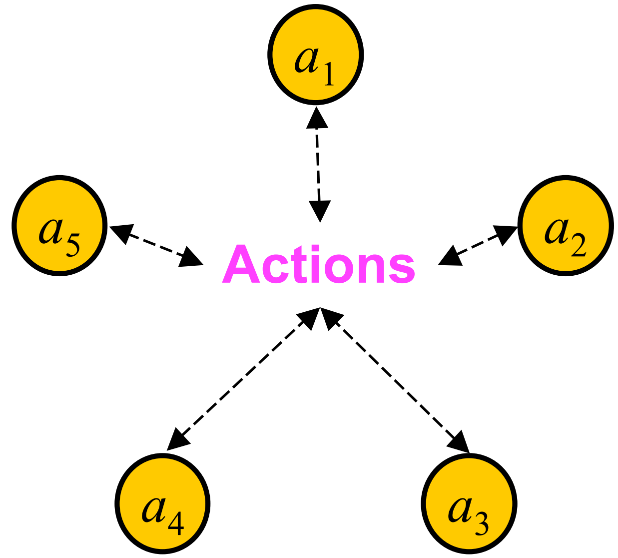
ICCCD 2008

Please help us by playing a game:

- Choose a number in the range from 0 to 100, and write it in the space below.
- We'll take the average of all of the numbers. The winner(s) will be whoever chose a number that's closest to 2/3 of the average.

- This game is famous among economists and game theorists
  - » It's called the ***p*-beauty contest** (I used $p = 2/3$)
- What does game theory tell us about it?
  - First, a very brief review of some game-theoretic concepts

Classical Game Theory



- Consider a game played by a set of agents $A = \{a_1, a_2, \ldots, a_n\}$
- An agent's *strategy*: description of what it will do in every possible situation
  - » May be deterministic or probabilistic
- Let $S = \{s_1, s_2, \ldots, s_n\}$ be the strategies used by $\{a_1, a_2, \ldots, a_n\}$, respectively
  - » Then a_i's *expected utility* is a_i's average payoff given S
- S is a *Nash equilibrium* if *no* agent can get a higher expected utility by *unilaterally* switching to a different strategy
  - » I.e., each agent is doing the best that it can do, *given what the other agents are doing*
- An agent is *rational* if it makes choices that optimize its expected utility
  - » Hence a set of rational agents should gravitate toward a Nash equilibrium

Nash Equilibrium for the *p*-Beauty Contest

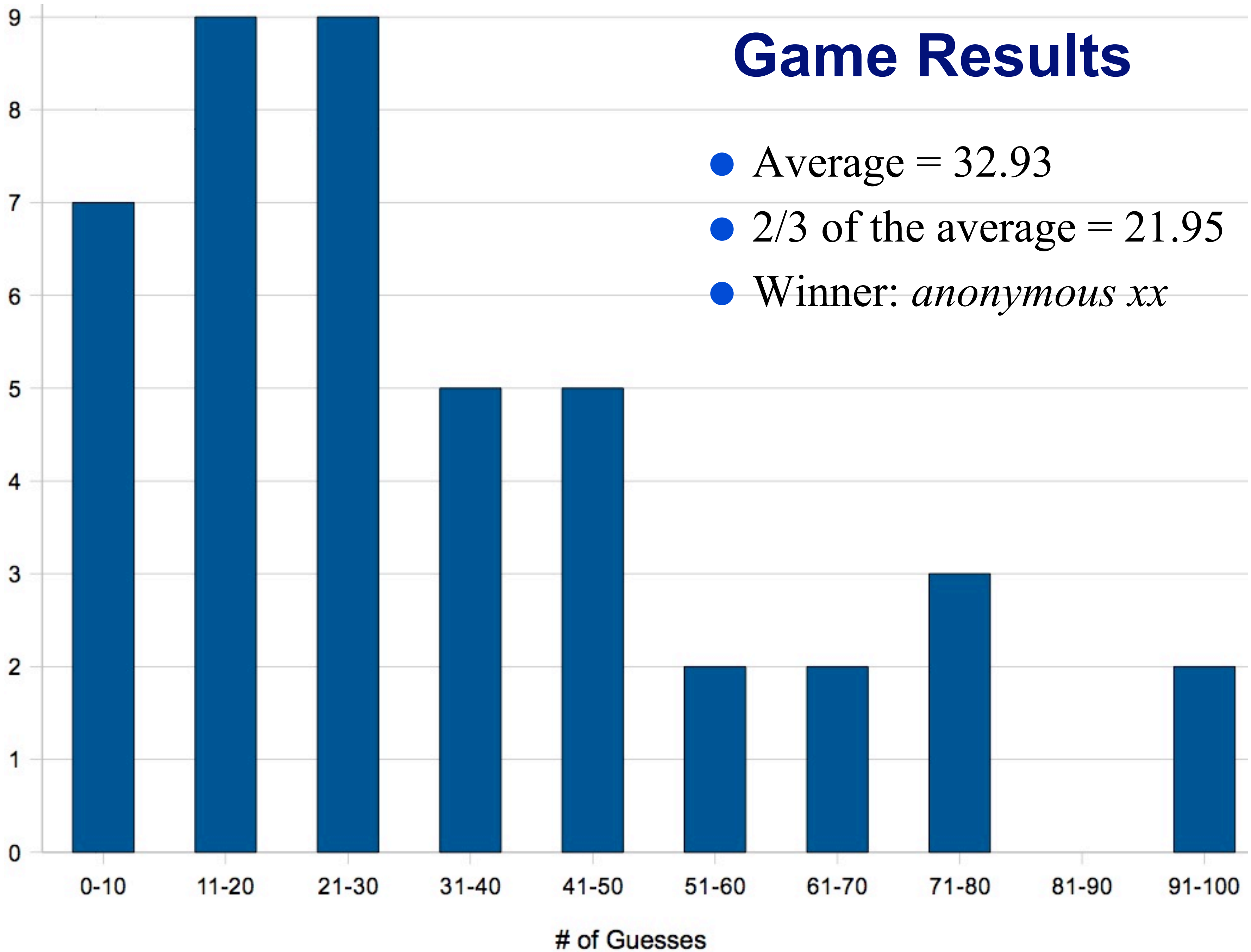
- We can find a Nash equilibrium for the *p*-beauty contest by doing *backward induction*
  - All of the numbers are ≤ 100
    - average ≤ 100 => 2/3 of the average < 67
  - If everyone figures this out, they'll choose 67 or less
    - average ≤ 67 => 2/3 of the average < 45
  - If everyone figures *this* out, then they'll choose 45 or less
    - average < 45 => 2/3 of the average < 30
  - …
- Nash equilibrium strategy: everybody chooses 0
- For those of you who are familiar with evolutionary game theory, this strategy is evolutionarily stable

We aren't game-theoretic "rational" agents

- Huge literature on *behavioral economics* going back to about 1979
  - Many cases where humans (or aggregations of humans) tend to make different decisions than the game-theoretically optimal ones
  - Daniel Kahneman received the 2002 Nobel Prize in Economics for his work on that topic

Game Results

- Average = 32.93
- 2/3 of the average = 21.95
- Winner: *anonymous xx*



Choosing "Irrational" Strategies

- Why did ***you*** choose a non-equilibrium strategy?
  - Limitations in reasoning ability
  - Hidden payoffs
  - Opponent modeling

Limitations in Reasoning Ability

- Maybe you didn't calculate the Nash equilibrium correctly, or you didn't know how to calculate it, or you didn't even know the concept

- R. Nagel (1995) "Unravelling in Guessing Games: An Experimental Study." *American Economic Review* 85, 1313–1326
  - Empirical results compatible with the assertion that
    - 13% of subjects used no backward induction
    - 44% used one level of backward induction
    - 37% used two levels
    - 4% used more than two levels

- Some games are so complicated that even though an optimal strategy exists, it's not feasible to figure out what it is
  - In chess, the number of possible moves is larger than the number of atoms in the universe
    - The number of possible strategies is even larger

Hidden Payoffs

- "Hidden payoffs" are payoffs not included in the game model
  - The game model assumed your objective was to *win* the game
- Maybe you participated in the game for a different reason:
  - Because you thought it would be fun
  - Because you were curious what would happen
  - Because you thought it might help me or help the conference
  - Because the people at the registration desk asked you to
  - Because you wanted to create mischief
- In these cases, there wouldn't necessarily be any reason for you to choose the Nash equilibrium strategy

Opponent Modeling

- Maybe you predicted that the other players' likely moves made it unlikely that the Nash equilibrium strategy would win

- More generally,
  - A Nash equilibrium strategy is best for you *if the other agents also use their Nash equilibrium strategies*
  - In many cases, the other agents *won't* be using Nash equilibrium strategies
  - In such cases, if you can predict the other agents' likely actions, you may be able to do much better than the Nash equilibrium strategy

- I'll give you several examples …

Roshambo (Rock-Paper-Scissors)

A_2 \ A_1	Rock	Paper	Scissors
Rock	0, 0	–1, 1	1, –1
Paper	1, –1	0, 0	–1, 1
Scissors	–1, 1	1, –1	0, 0

- Rock beats scissors
- Scissors beats paper
- Paper beats rock

- Nash equilibrium strategy:
  - Choose randomly, probability 1/3 for each move
  - Expected utility = 0

Roshambo (Rock-Paper-Scissors)

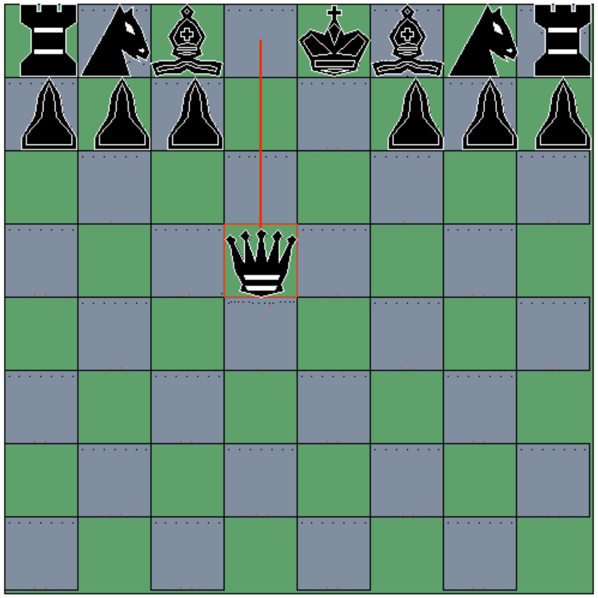
- International roshambo programming competitions
  - » 1999 and 2000, Darse Billings, U. of Alberta
  - » http://www.cs.unimaas.nl/ICGA/games/roshambo
- The 2000 competition
  - » First phase: round-robin
    - 64 programs competed
    - For each program, 1000 iterations against each of the programs (including itself)
    - Hence 64000 points possible per program
  - » Results averaged over 100 trials
    - Highest score: 9268
    - Lowest score: –52074

Poker

- Sources of uncertainty
  - The card distribution
  - The opponents' betting styles
- Lots of recent AI work on the most popular variant of poker
  - Texas Hold 'Em
- The best AI programs are starting to approach the level of human experts
  - Construct a statistical model of the opponent
    - What kinds of bets the opponent is likely to make under what kinds of circumstances
  - Combine with game-theoretic reasoning
    - Billings, Davidson, Schaeffer, and Szafron (2002). "The challenge of poker." *Artificial Intelligence* (Vol. 134, Issue 1-2)

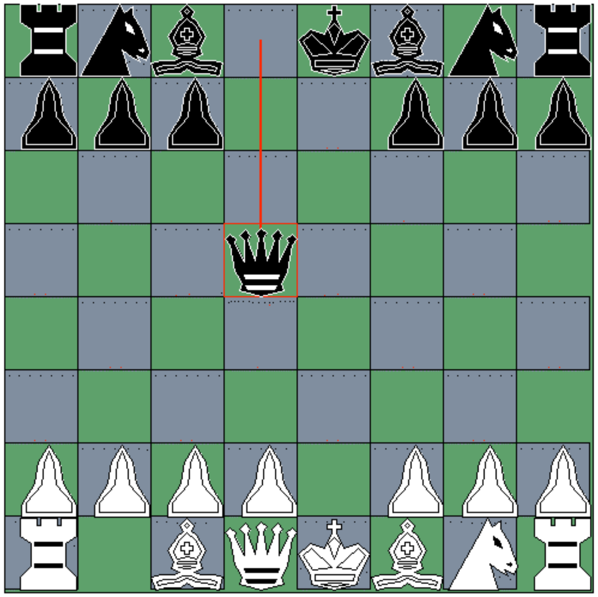
Kriegspiel Chess

- **Kriegspiel**: an imperfect-information variant of chess
  - Developed by a Prussian military officer in 1824
  - Became popular as a military training exercise
  - Progenitor of modern military wargaming
- Like chess, but
  - You don't know where your opponent's pieces are, because you can't observe most of their moves
- Only ways to observe:
  - You take a piece, they take a piece, they put your king in check, you make an illegal move
- Size of belief state (set of all states you *might* be in):
  - Texas hold'em: 10^3 (one thousand)
  - bridge: 10^7 (ten million)
  - kriegspiel: 10^{14} (ten trillion)



Monte-Carlo Information-Set Search

- Recursive formulas for computing expected utilities of belief states
  - Explicitly incorporates an opponent model
- Infeasible computation, due to belief-space size
- Monte Carlo approximations of the belief states
  - Reduces the computation to sort-of feasible
- Results:
  - One of the world's best kriegspiel programs
  - The *minimax* opponent model
    (Nash equilibrium in ordinary chess)
    is *not* the best opponent model for kriegspiel
  - A better model is an "overconfident" one
    that assumes the opponent won't play very well



- Parker, Nau, and Subrahmanian (2006). Overconfidence or paranoia? search in imperfect-information games. *Proc. National Conf. on Artificial Intelligence (AAAI)*
  - http://www.cs.umd.edu/~nau/papers/parker06overconfidence.pdf

Behavior in an Unfamiliar Society

- Suppose you enter an environment that's inhabited by agents who are unfamiliar to you
- You know what actions are *possible*
- But you don't know
  - What behaviors and outcomes the agents prefer, and why
  - How they're likely to react to your actions
  - What collection of behaviors is most likely to elicit the responses *you* prefer
- How can you decide how to behave?

Behavior in an Unfamiliar Society

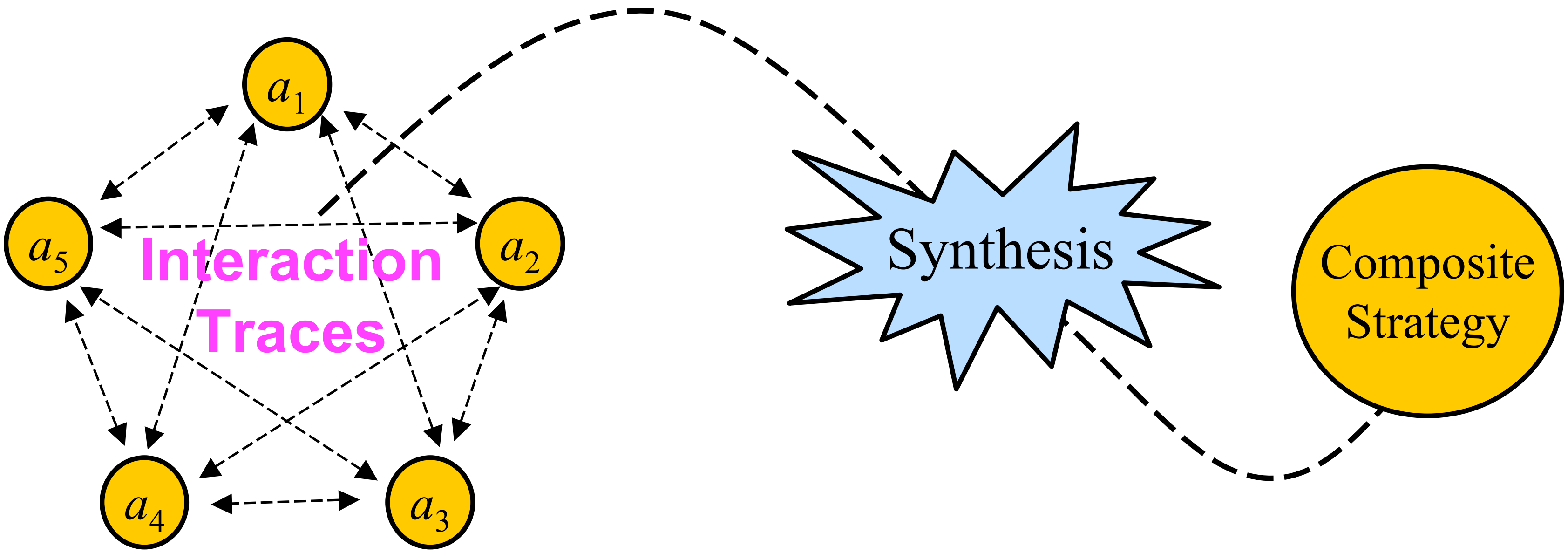
- Suppose you enter an environment that's inhabited by agents who are unfamiliar to you
- You know what actions are *possible*
- But you don't know
  - » What behaviors and outcomes the agents prefer, and why ——— *their payoffs*
  - » How they're likely to react to your actions ——— *their strategies*
  - » What collection of behaviors is most likely to elicit the responses *you* prefer ——— *your best strategy*
- How can you decide how to behave?

Technical Approach

- Work by my recent PhD graduate, Tsz-Chiu Au
    - Finished his PhD this August
    - Now at University of Texas

- For technical details, see
    - » T.-C. Au, D. Nau, and S. Kraus. Synthesis of strategies from interaction traces. *International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2008
    - » http://www.cs.umd.edu/~nau/papers/au08synthesis.pdf

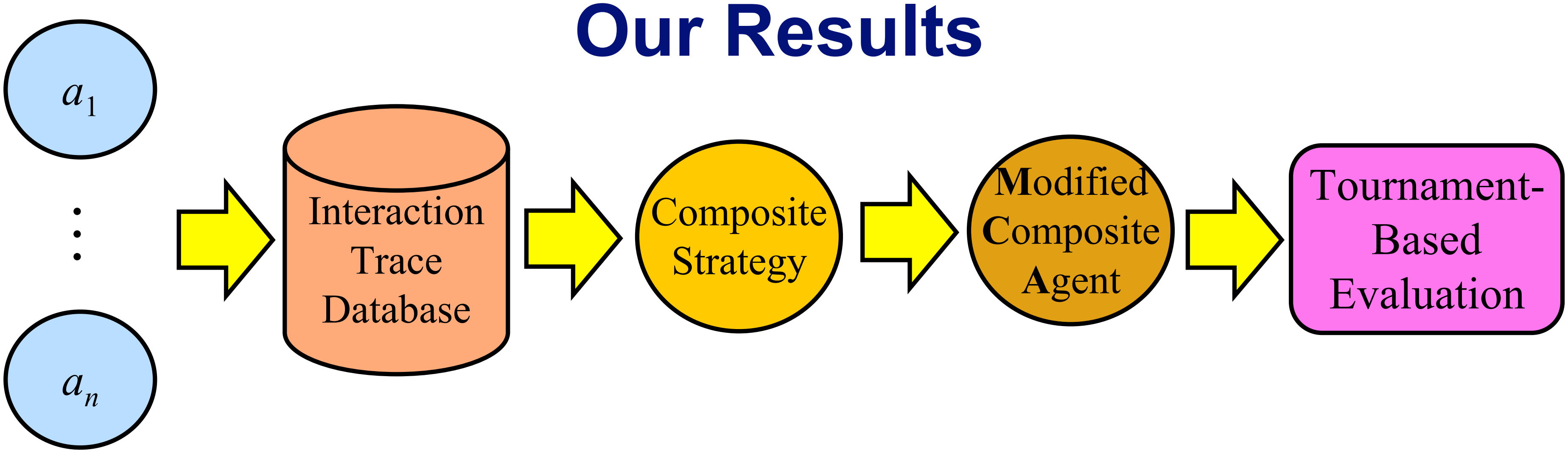
Learning from Interaction Traces



- Suppose the other agents are competent members of their society
  - Even without knowing their payoffs, we can guess that many of their interactions produce payoffs that at least are acceptable to them
- So let's see if we can use those interactions ourselves
  - » Observe agents' interactions, collect *interaction traces*
  - » Look at what interaction traces produce outcomes that *we* prefer

    i.e., high payoff for us if we interact with those agents
  - » Synthesize a ***composite strategy*** that combines those traces

Our Results



- **Necessary and sufficient conditions**
  - » for combining interaction traces into a composite strategy
- **The CIT algorithm**
  - » selects the best set (i.e., highest expected utility) of combinable interaction traces, and combines them
- **Modified composite agent**
  - » augments an agent to use the composite strategy to enhance its performance
- **Cross-validated experimental results**

Repeated Games

- Used by game theorists, economists, social and behavioral scientists as simplified models of various real-world situations

- Some well-known examples
  - » Roshambo
  - » Iterated Prisoner's Dilemma
  - » Iterated Battle of the Sexes
  - » Iterated Chicken Game
  - » Repeated Stag Hunt
  - » Repeated Ultimatum Game
  - » Repeated Matching Pennies

- I'll describe three of them

Iterated Prisoner's Dilemma

- **Prisoner's Dilemma**
  - » Each prisoner can *cooperate* with the other or *defect* (incriminate them)
- **Iterated** Prisoner's Dilemma (IPD)
  - » Iterations => incentive to cooperate
- Widely used to study emergence of cooperative behavior among agents
- IPD tournaments [Axelrod, *The Evolution of Cooperation*, 1984]
  - » ***Tit-for-Tat (TFT)***
    - On 1st move, cooperate. On nth move, repeat the other player's (n–1)-th move
  - » Could establish and maintain advantageous cooperations with many other players
  - » Could prevent malicious players from taking advantage of it

Prisoner's Dilemma:

a_1 \ a_2	Cooperate	Defect
Cooperate	3, 3	0, 5
Defect	5, 0	1, 1

Nash equilibrium

If I defect now, he might punish me by defecting next time

Iterated Chicken Game

- **Chicken Game**:
  - » Made famous in *Rebel Without a Cause*
  - » Two people drive toward a cliff
  - » The first one to turn away loses face
  - » If neither turns away, both will be killed
- Example
  - » Two groups need to divide a piece of land between them
  - » If they can't agree how to divide it, they'll fight
- Nash equilibria (with no iteration):
  - » Do the opposite of what the other agent does

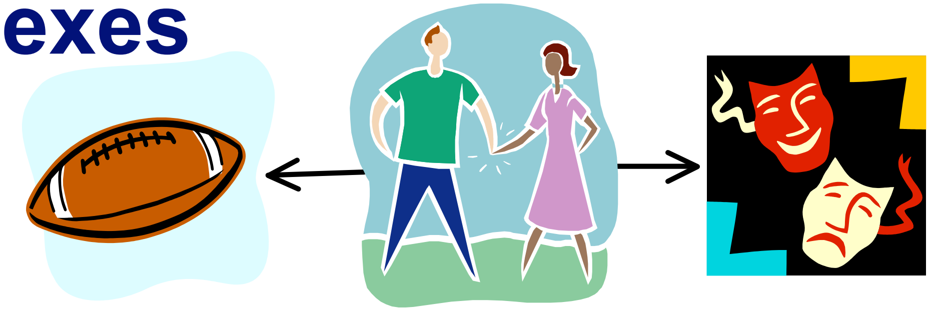
Chicken game:

a_1 \ a_2	Cooperate	Defect
Cooperate	4, 4	3, 5
Defect	5, 3	0, 0

Nash equilibria

- **Iterated** Chicken Game (ICG)
  - » Mutual cooperation does not emerge
    - Each player wants to establish him/herself as the defector

Iterated Battle of the Sexes



- **Battle of the Sexes**:
  - Two players need to coordinate their actions to achieve some goal
  - They each prefer different actions
- Original scenario: husband prefers football, wife prefers opera
- Another scenario:
  - Two nations must act together to deal with an international crisis
  - They prefer different solutions

Nash equilibria

Battle of the Sexes:

Husband / *Wife*	Opera (G)	Football (T)
Opera (T)	2, 1	0, 0
Football (G)	0, 0	1, 2

T (take): choose your preferred activity
G (give): choose the other's preferred activity

- **Iterated** Battle of the Sexes (IBS)
  - Two players repeatedly play the Battle of the Sexes
  - Not very much is known about what strategies work well in this game

Example: Two Interaction Traces

- Suppose an observer sees the following two interaction traces
- The game is the IBS, but the observer doesn't necessarily know that

Someone interacting with an agent who is using a "fair" strategy:

- » Behavior: alternately *Take* & *Give*

Someone interacting with an agent who is using a "selfish" strategy:

- » Behavior: always *Take*

	Trace 1		Trace 2	
	a_1	a_2 (fair)	a_3	a_4 (selfish)
Round 1:	G	T	G	T
Round 2:	T	G	T	T
Round 3:	G	T	T	T
Round 4:	T	G	G	T
Round 5:	G	T	G	T
Round 6:	T	G	G	T
…	…	…	…	…

How Should We Behave?

- In the games that I described, prejudice is impossible
- The agents don't have *tags*
  - » a_2 doesn't know a_1's name, reputation, ethnicity, gender, social status, …
  - » a_2's only information about a_1 is how a_1 behaves toward a_2
  - » if we interact with a_2 and we behave like a_1, then a_2 should behave like it did with a_1

All I know of you is your actions

a_1	a_2
G	T
T	G
G	T
T	G
G	T
T	G
…	…

us	a_2
G	T
T	G
G	T
T	G
G	T
T	G
…	…

- This also could happen in a game where prejudice is possible, if a_2 had the same prejudice toward both a_1 and us

How Should We Behave?

- If a_1 and a_3 are both competent members of their society, we might want to
  - Emulate a_1's behavior if we're playing with a_2
  - Emulate a_3's behavior if we're playing with a_4
- Problem: how do we know whether the other agent is a_2 or a_4?
  - Recall that the agents don't have tags
  - The only way we can find out is by observing how the agent acts
- Solution:
  - Combine the behaviors of a_1 and a_3 to get a strategy that tells how to act with both of them
  - We can do this because the interaction traces have a property called *compatibility*

Trace 1		Trace 2	
a_1	a_2 (fair)	a_3	a_4 (selfish)
G	T	G	T
T	G	T	T
G	T	T	T
T	G	G	T
G	T	G	T
T	G	G	T
…	…	…	…

Compatible Interaction Traces Example

Trace 1		Trace 2	
a_1	a_2	a_3	a_4
G	T	G	T
T	G	T	T
G	T	T	T
T	G	G	T
G	T	G	T
T	G	G	T
…	…	…	…

- These two traces are *compatible*

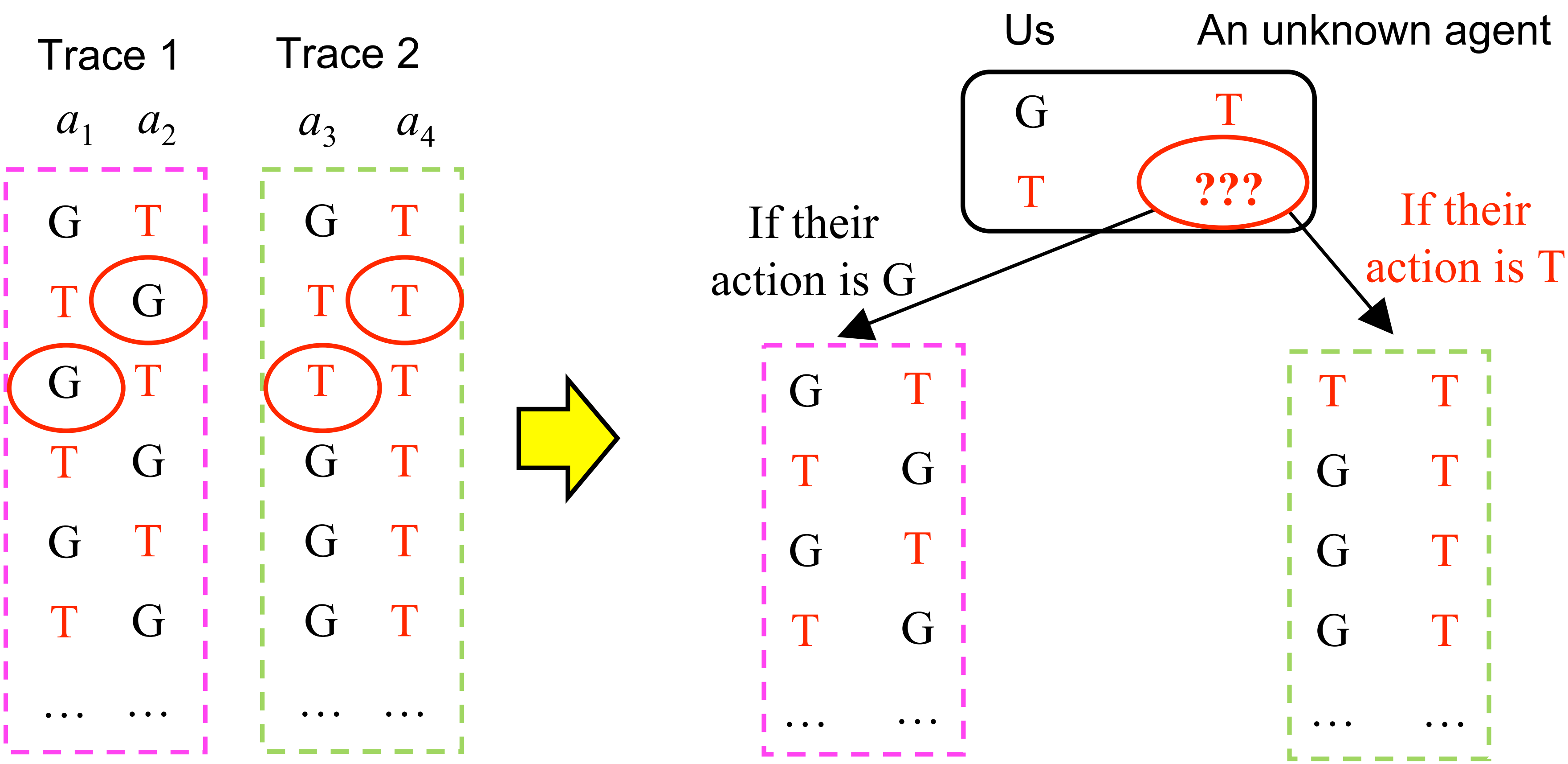
  because

  ← The point where the other agents' actions differ

  ← is *before* the point where *our* actions must differ

- We can identify the other agent's behavior soon enough to decide what to do

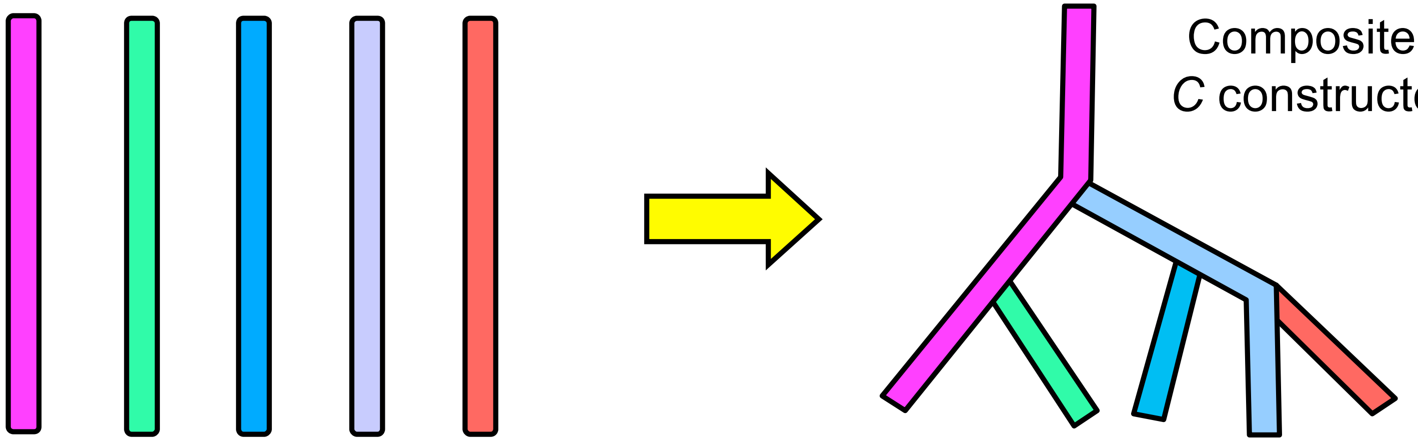
Compatible Interaction Traces Example



- Combine the behaviors of a_1 and a_3 into a *composite strategy*
  - The decision tree shown here

Compatible Interaction Traces General Case

T = **compatible** set of interaction traces



Theorem: If a set of interaction traces *T* is compatible, then we can combine *T* into a composite strategy *C*

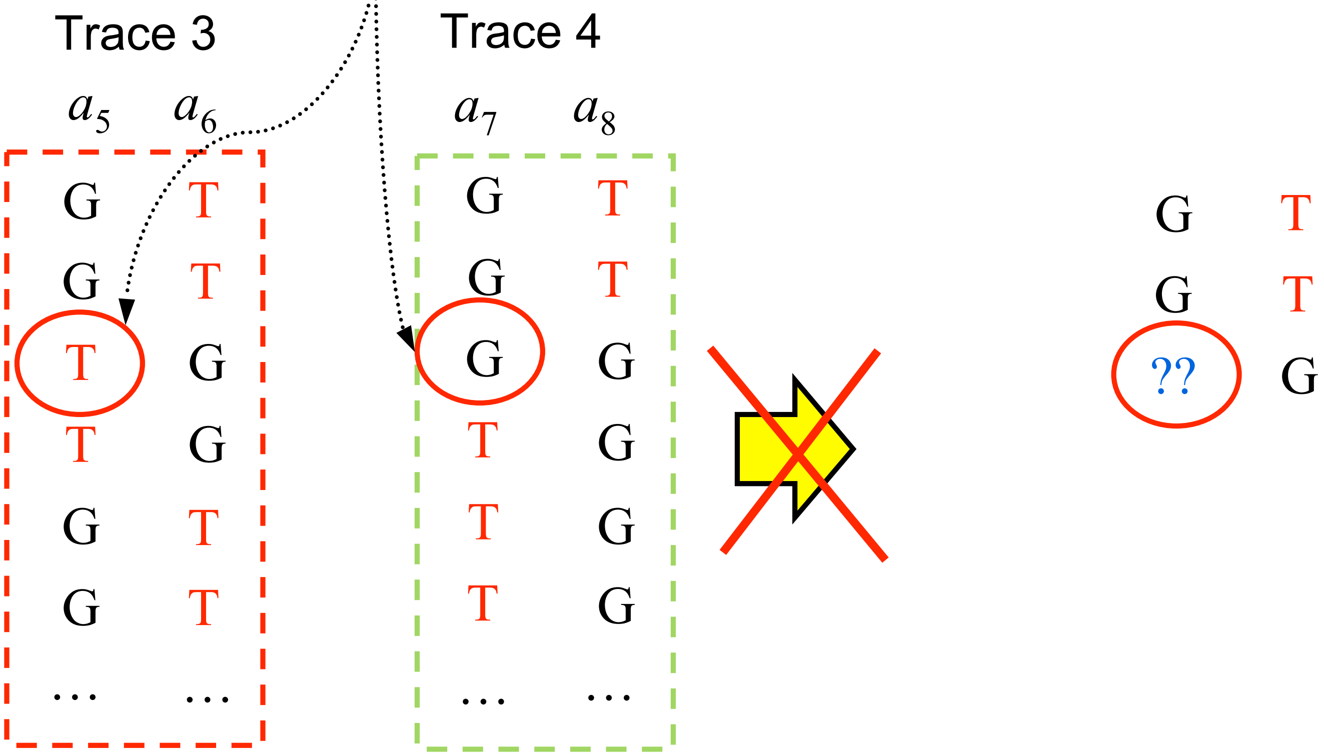
- If *T* includes an interaction trace for every pure strategy, then *C* will be a *total* strategy
  - I.e., it will specify an action for every situation we might encounter
- Otherwise *C* will be a *partial* strategy
  - In some cases it won't tell us what action to perform

Incompatible Interaction Traces Example

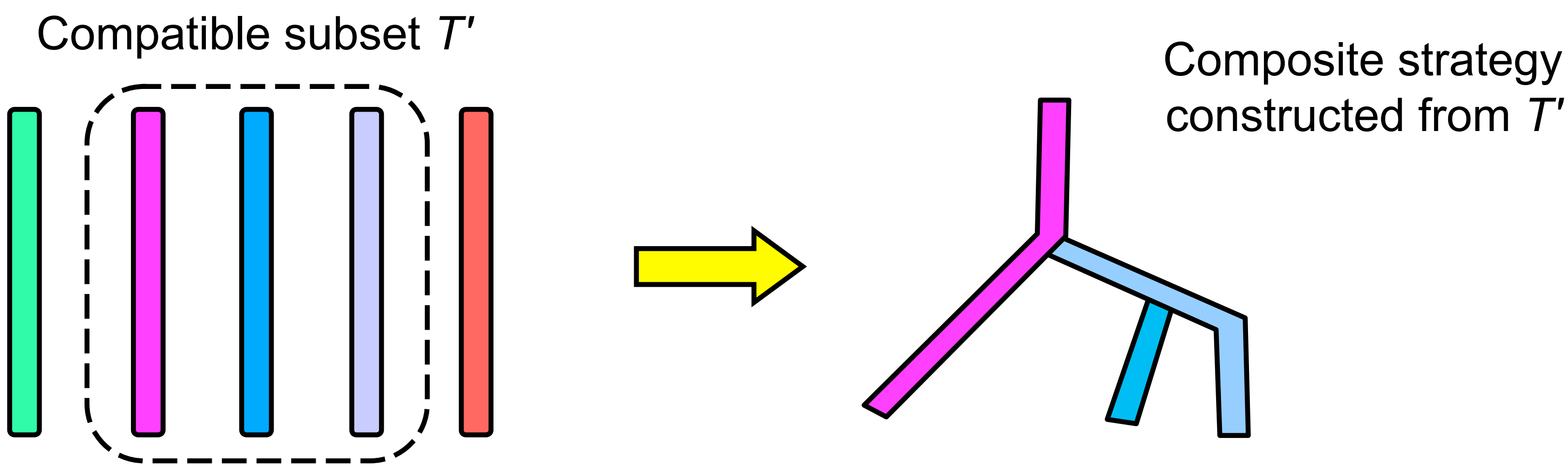
- Must choose between two incompatible moves *before* we have enough information to see how the opponent is going to behave

  - Which move to choose?

Can't make both of these moves



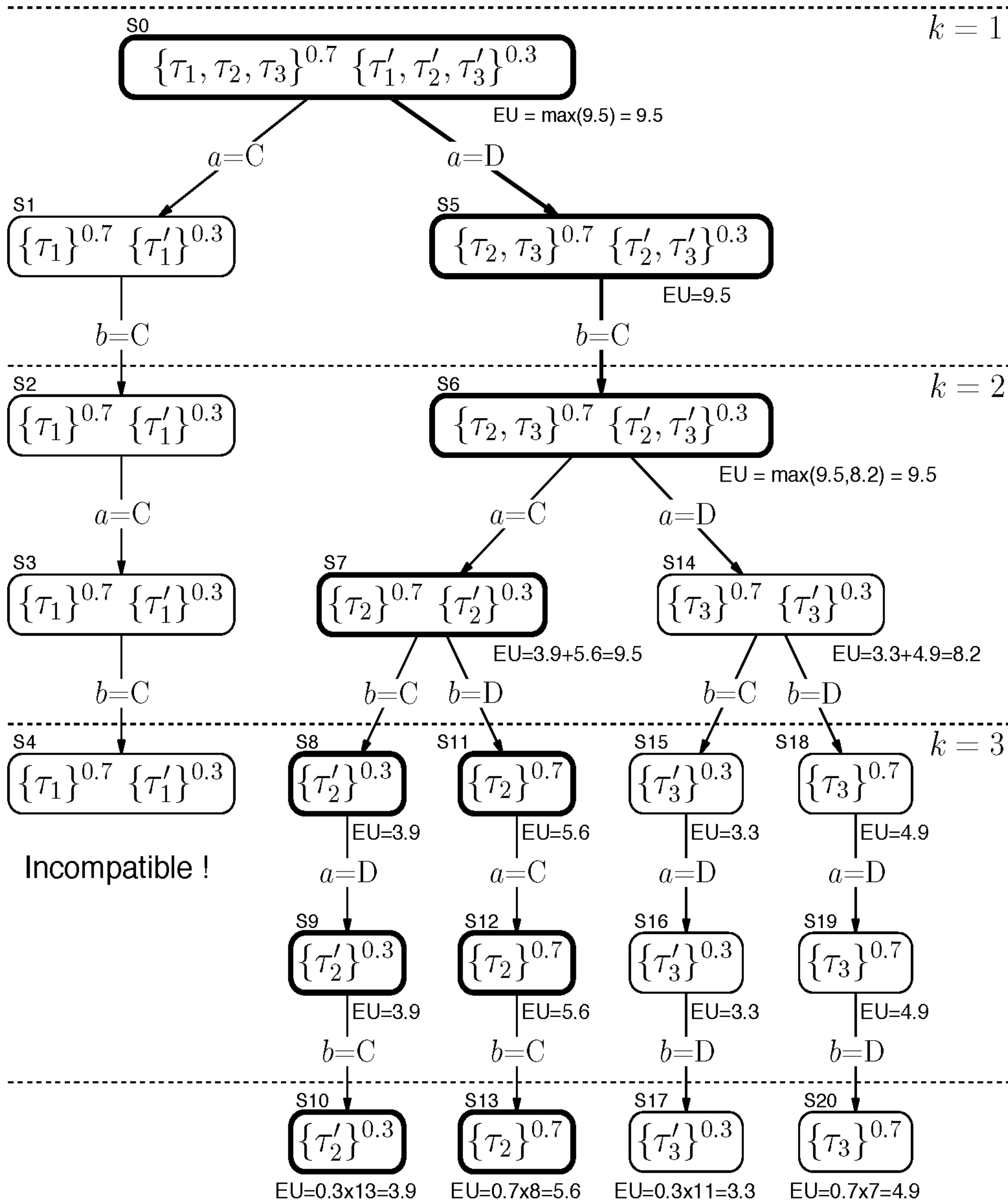
Incompatible Interaction Traces
General Case



- If T isn't compatible, we can find a compatible subset T' of T and construct a composite strategy C' from T'
  - Want to choose T' so as to maximize our expected utility
  - Don't need to know other agents' payoffs and utilities, just ours
- Problem 1: exponentially many possibilities for T'
  - Reduce to polynomial using divide-and-conquer
- Problem 2: to compute expected utility, need probability of each trace
  - Get this from how many times the trace occurs in T

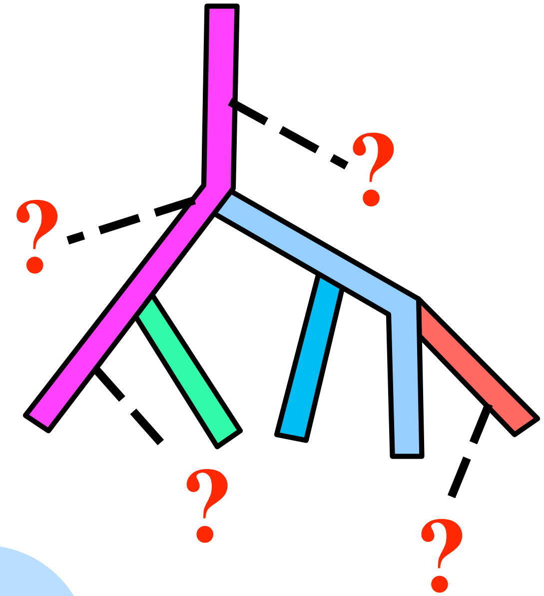
The CIT Algorithm

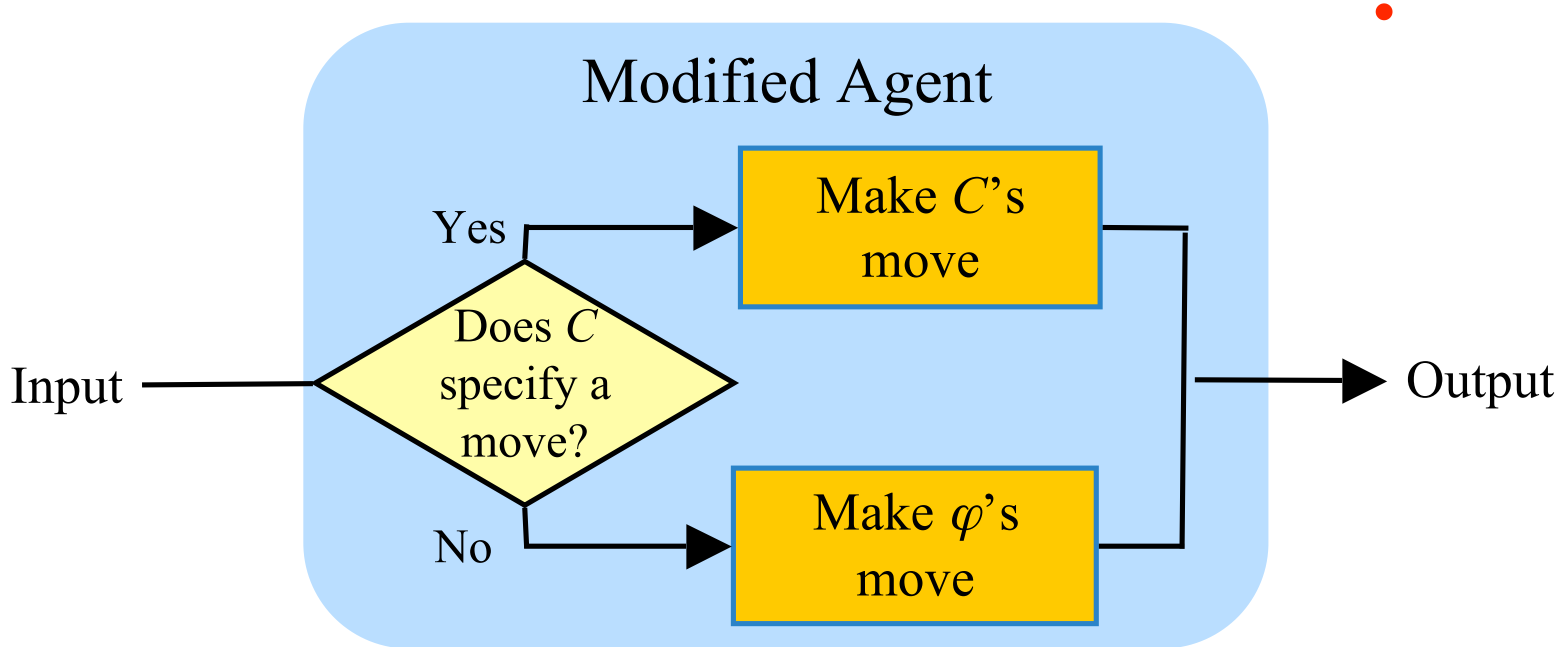
- Let T be a set of interaction traces
- Let **C** = {all composite strategies that can be formed from compatible subsets of T}
- **The CIT algorithm**
  - » Finds an optimal composite strategy (highest expected utility of any strategy in **C**)
  - » Divide-and-conquer algorithm
  - » Running time
    = $O(|T| \times$ length of game)



Using a Composite Strategy to Enhance an Agent's Performance

- Given an agent φ and a composite strategy C,
- *Modified Composite Agent* MCA(φ,C)
  - » If C specifies a move for the current situation, then make that move
  - » Otherwise make whatever move φ would normally make





Experimental Evaluation

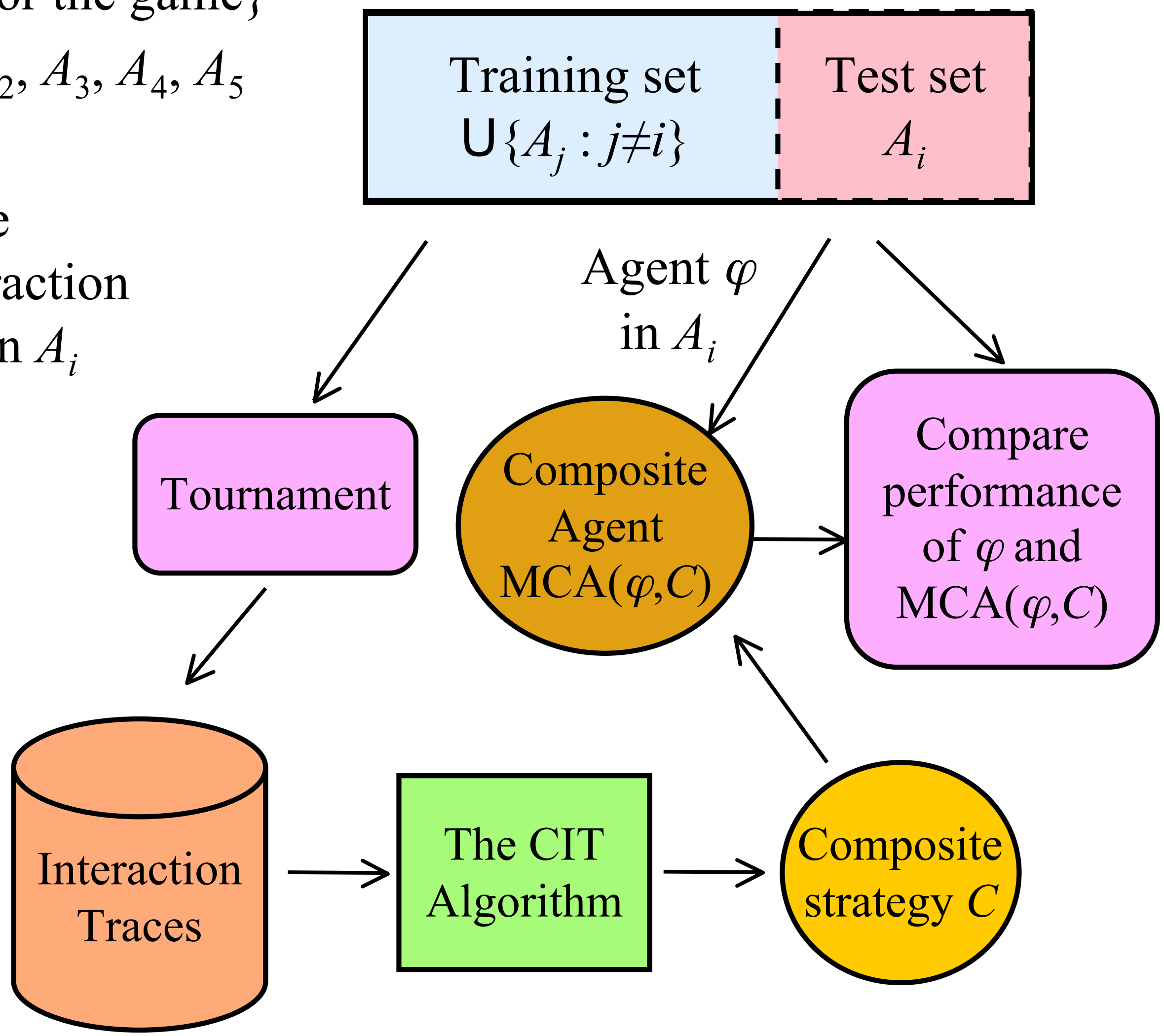
- Evaluated in three games:
  - Iterated Prisoner's Dilemma (IPD)
  - Iterated Chicken Game (ICG)
  - Iterated Battle of the Sexes (IBS)
- Sources of agents:
  - Asked students in advanced AI classes to contribute agents to play in tournaments
  - Also added the usual "standard" agents:
    - ALLC, ALLD, GRIM, NEG, PAVLOV, RAND, STFT, TFT, TFTT

	IPD	**ICG**	**IBS**
No. of deterministic agents	34	22	17
No. of probabilistic agents	18	24	29
Total no. of agents	52	46	46

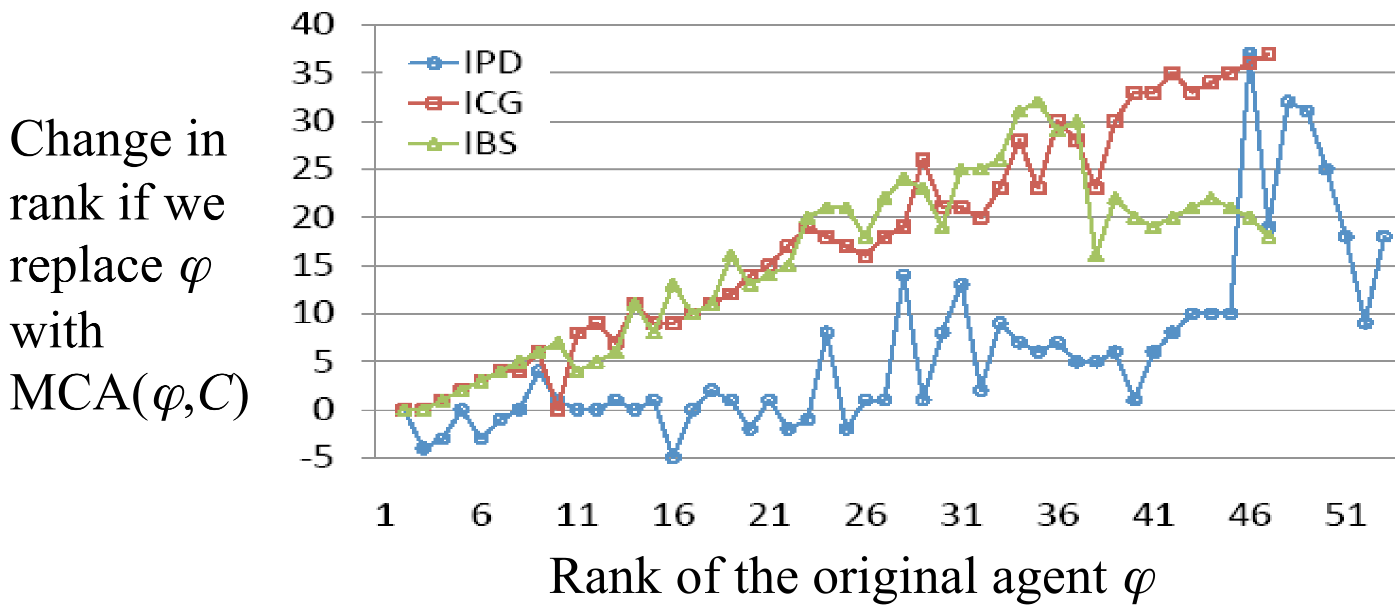
Experimental Setup

For each game, a five-fold cross-validation experiment:

- A = {all the agents we have for the game}
- Divide A into 5 subsets A_1, A_2, A_3, A_4, A_5
- For $i = 1, 2, 3, 4, 5$
  - Create optimal composite strategy C from the interaction traces of the agents *not* in A_i
  - For each agent φ in A_i, play φ and MCA(φ,C) against the agents in A_i
    - Which does better: MCA(φ,C) or φ?
    - How often does MCA(φ,C) use C?

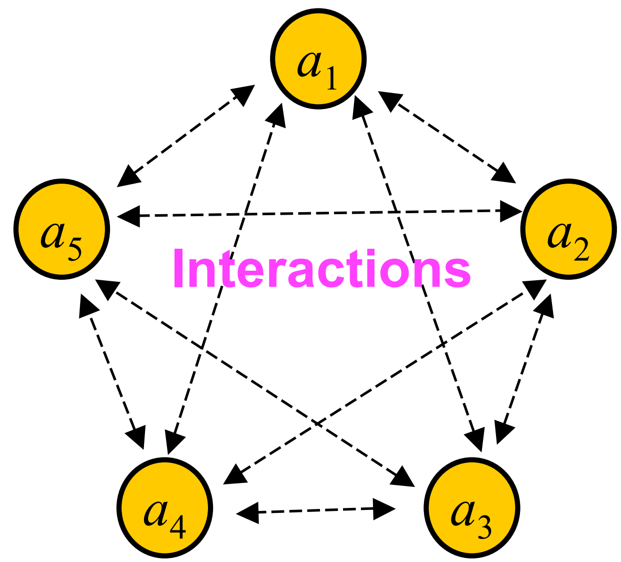


Experimental Results



- In nearly every case MCA(φ,C)'s score and rank were much higher than φ's
  - Especially in cases where the strategy φ was weak
- Smallest improvement in the IPD
  - Reason: most IPD agents already are very good to begin with
- **Using composite strategies can greatly enhance an agent's performance**

Is This Work Generalizable?



- Limitations of this work
  - Repeated games
  - All interactions were among pairs of agents
  - In each interaction, an agent had only two available actions
    - cooperate or defect, give or take
- What about
  - Sequential games
  - Interactions among *n* agents at once
  - More than two available actions
- The number of possible strategies could increase exponentially
  - Would the same approach have any realistic chance of working?
- I think that in many cases, the answer may be yes
  - We only needed to use a very small fraction of the possible strategies

How Often the Composite Strategies Were Used

- Each game lasted 200 iterations,
  - Hence, 2^{200} possible behaviors (sequences of actions) the other agent might use
- On average, C contains only about 200 traces
  - C only tells us how to behave against 200 of the 2^{200} possible behaviors
- If we play MCA(φ,C) against a random agent a
  - Pr[C gives us a behavior to use with a] $\approx 200/2^{200} \approx 10^{-58}$
- In our experiments, MCA(φ,C) used C ***much*** more frequently than that
  - in the IPD, 84% of the time
  - in the IBS, 48% of the time
  - in the ICG, 45% of the time

- In other words …

A Small Set of "Conventional" Behaviors

- One reason why our approach worked well
  - There was a set of "conventional behaviors" that was extremely small compared to the set of all possible behaviors
  - By observing and analyzing the interactions, we synthesized a strategy that utilized these behaviors
  - The strategy worked successfully in many or most circumstances

- We believe that in more complex environments, there again will be a small set of conventional behaviors
  - This provides a reason to believe that learning from observed interactions can be useful there too

Summary

- How to construct a strategy for a new, unfamiliar environment?
- Approach:
  - » Observe the interactions among agents who are familiar with the environment
  - » Combine interaction traces into a *composite strategy*
- Results:
  - » Necessary and sufficient conditions for combining a set of traces
  - » The CIT algorithm: finds an optimal composite strategy
  - » Modified composite agent: uses the composite strategy to enhance the performance of an existing agent
  - » Experimental results in three different non-zero-sum games
    - Our approach greatly enhanced the performance of most agents
  - » Small set of conventional behaviors
    - Observing and learning them gives you a large fraction of what you need in order to do well

Future Work on This Topic

- How to synthesize strategies from interaction traces for *noisy* environments?
  - » Environments in which accidents and miscommunications can happen
  - » Can cause big problems for some strategies
- Game can go on longer than the interaction traces do
  - » Machine-learning techniques such as Q-learning and TD(λ) can handle infinite horizons
  - » Can we do the same?
- Other kinds of environments
  - » Zero-sum games
  - » POMDPs
  - » Other multi-agent environments

Ongoing Work on Related Topics

- V. S. Subrahmanian and I co-direct the Lab for Computational Cultural Dynamics
- Highly cross-disciplinary partnership
  - Computer Science
  - Political science
  - Psychology
  - Criminology
  - Linguistics
  - Public Policy
  - Business
  - Systems Engineering

- LCCD'S active partners include
- Univ. of Pennsylvania (sub)
- Nat. Consortium for the Study of Terrorism and Responses to Terrorism (START)
- Specific regional experts
  - Minister of S&T, Rwanda
  - Former Afghan Deputy Minister of the Interior
  - Former State Dept. officials who served in Pakistan-Afghanistan
  - Former general involved in busting the Tupac Amaro and Shining Path in Peru

LCCD Laboratory for **Computational** Cultural Dynamics

SOMA Rules

- SOMA Rules: predict an agent's behavior under a set of arbitrary conditions
  - » group *g* will take action *a* with probability *x*% to *y*% when condition *c* holds
  - » Techniques for extracting such rules automatically from databases (e.g., Minorities at Risk) and from electronic text
  - » Algorithms for using SOMA rules to create forecasts
    - E.g., Aaron Mannes's talk this morning

- V. Subrahmanian, M. Albanese, M. V. Martinez, D. Nau, D. Reforgiato, G. I. Simari, A. Sliva, O. Udrea, and J. Wilkenfeld. CARA: A cultural-reasoning architecture. *IEEE Intelligent Systems*, Mar./Apr. 2007

Social Learning Strategies Tournament

- International competition among computer programs in a evolutionary cultural adaptation environment
  - Consortium of researchers funded by the European Union
  - €10,000 prize
  - http://www.intercult.su.se/cultaptation/tournament.php

- Several of our students have entered programs
- In addition, they have successfully analyzed several simplified versions of the game, to find provably optimal strategies
  - Carr, Raboin, Parker, and Nau: “When Innovation Matters: An Analysis of Innovation in a Social Learning Game.” *ICCCD-2008*
    - Presentation tomorrow
  - Carr, Raboin, Parker, and Nau: “Balancing innovation and exploitation in a social learning game.” *AAAI Fall Symposium on Adaptive Agents in Cultural Contexts*. To appear.

International Planning Competition

- RFF - algorithm for generating plans under uncertainty
  - Guillaume Infantès (ONERA, visitor to our lab)
  - Florent Teichteil-Königsbuch (ONERA, visitor to our lab)
  - Ugur Kuter (research scientist in our lab)
- I just received word today that RFF has won the probabilistic planning track of the 2008 International Planning Competition

Postdoctoral Research Opportunity

- I want to hire a postdoctoral researcher to do research involving game theory
- If you know of anyone who might be suitable, please let me know

  - Dana S. Nau
  - nau@cs.umd.edu
  - 301-405-2684