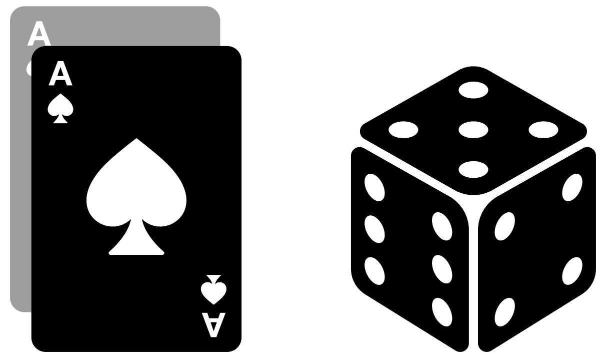
Equilibria in Games



"Tony" Runzhe Yang

Seung Lab Meeting

07/07/2023

Why games matter

Axios

Two new AI systems beat humans at complex games

Two new papers from AI powerhouses DeepMind and Meta describe how
are notching wins against human players in complex games...

Dec 1, 2022

Forbes

Artificial Intelligence Masters The Game of Poker – What Does That Mean For Humans?

When artificial intelligence (AI) system Pluribus successfully beat professional poker players in six-player Texas Hold 'em,...

Sep 13, 2019

ZME Science

A human just defeated an AI in Go. Here's why that matters

In 2016, the news was that AI beat humans at Go. Fast for
news is that humans beat AI at Go.

Feb 24, 2023

Forbes

Meta's AI Gamer Beat Humans In Diplomacy, Using Strategy And Negotiation

An AI agent created by Meta counted more than double the average human player in a competition for the online game Diplomacy,...

Nov 22, 2022

Nature

DeepMind AI topples experts at complex game Stratego

Game-playing AIs that interact with humans are laying important groundwork for real-world applications.

Dec 1, 2022

Elements of a game: Players, Actions, Payoffs

On a date night ...

2 players: Boy and Girl

Deciding either going to a football game or going to an opera

Boy prefers football game

Girl prefers opera

Elements of a game: Players, Actions, Payoffs

	Football	Opera
Football	2 , 1	-1 , -1
Opera	-1 , -1	1 , 2

(boy's payoff, girl's payoff) when the action profile is (Football, Opera)

GAME 1 : the battle of the sexes

Assumptions of players

1. <u>Complete information</u>: each player knows the **payoffs** and possible **actions** of all players.

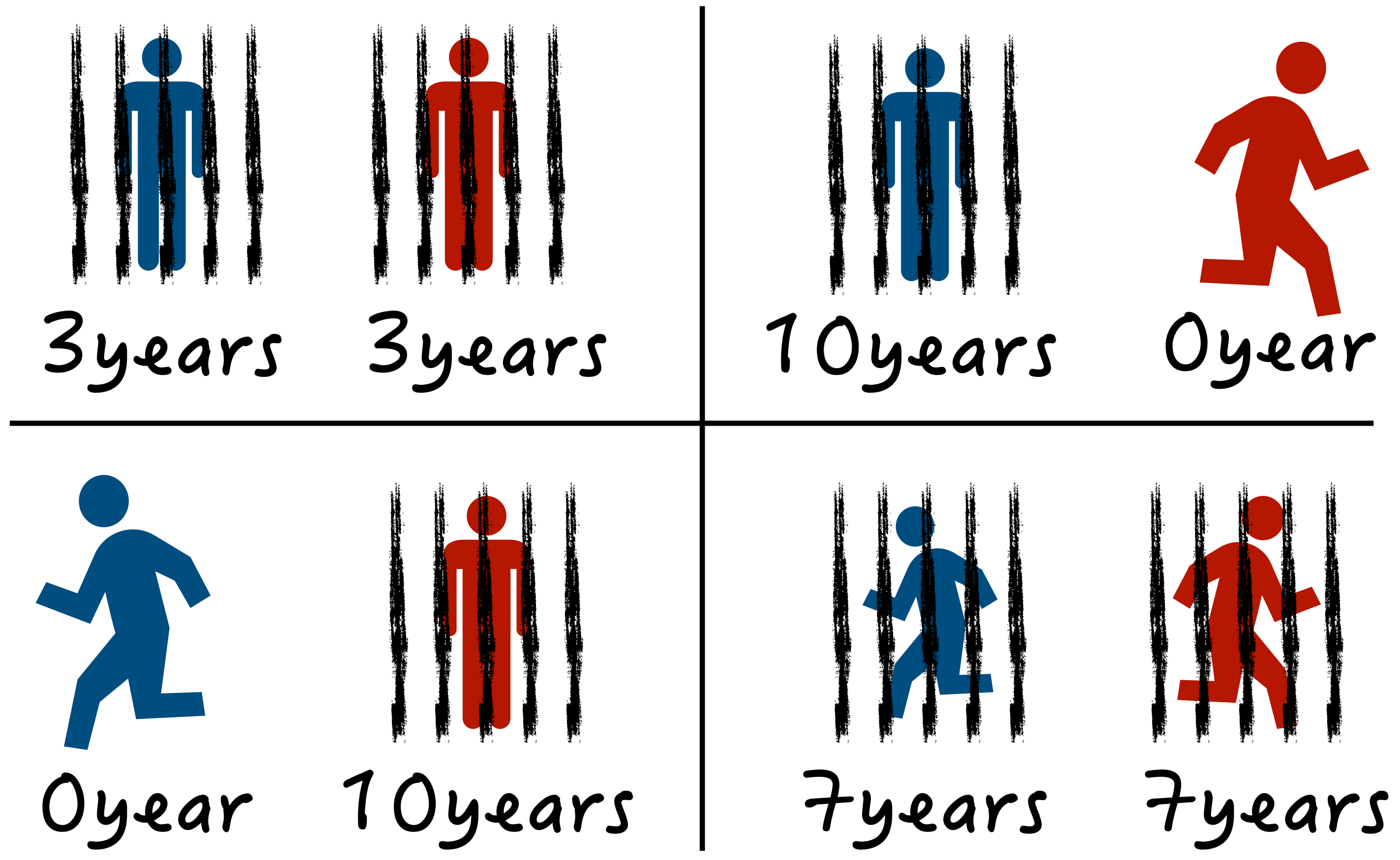
2. <u>Rational</u>: each player is interested to maximize his/ her payoff.

3. <u>Self-interest</u>: each player does NOT consider **the effect of actions on the others**, but only on his/her own.

Main Question: How would rational, self-interested players behave in a game?

The famous prisoner's dilemma

	Stay silent	Betray
Stay silent	-3, -3	-10, 0
Betray	0, -10	-7, -7



GAME II : prisoner's dilemma

Staying silent is "unstable"

	Stay silent	Betray
Stay silent	-3, -3	-10, 0
	↓ +3	↓ +3
Betray	0, -10	-7, -7
	→ +3	

Both players have an incentive to betray,
no matter what the other player does.

Betray-betray is the choice of rational, selfish players

	Stay silent	Betray
Stay silent	-3, -3	-10, 0
Betray	0, -10	-7, -7

Both players have NO incentive to deviate from (betray, betray).

Betray-betray is a "Nash Equilibrium"

Def. A strategy profile (a collection of strategies played by all players) is called a "**Nash Equilibrium**", <u>if NO player can gain more by changing only its own strategy.</u>

"Pure" strategy vs "mixed" strategies

	Paper	Rock	Scissors
Paper	0 , 0	1 , -1	-1 , 1
Rock	-1 , 1	0 , 0	1 , -1
Scissors	1 , -1	-1 , 1	0 , 0

GAME III : paper-rock-scissors

Mixed strategy Nash equilibrium in PRS

Opponent's strategy:

Paper	Rock	Scissors
1/3	1/3	1/3

My choice:

My choice	× Paper		× Rock		× Scissors		My expected payoff
Paper	0		1		-1		0
Rock	-1	+	0	+	1	=	0
Scissors	1		-1		0		0

More formal description of Nash equilibria

In a game w/ n players:

A := action profiles = $A_1 \times A_2 \times \cdots \times A_n$

u := joint payoff function = $(u_1, u_2, \cdots, u_n)$ $\quad u_i$: $A \rightarrow$ real number

a strategy profile $\sigma := (\sigma_1, \sigma_2, \cdots \sigma_n)$ where $\sigma_i(a_i)$ is the probability player i choosing action $a_i \in A_i$, and $\sigma(a) := \sigma_1(a_1) \times \sigma_2(a_2) \times \cdots \times \sigma_n(a_n)$

σ is a Nash equilibrium if, for all player i and all strategy σ'_i

$$\sum_{\text{all action profile } a} \sigma(a) \cdot u_i(a) - \sum_{\text{all action profile } a} \sigma'_i(a_i)\, \sigma_{-i}(a_{-i}) \cdot u_i(a) \geq 0$$

Pure and mixed NEs in the battle of the sexes

	Football	Opera
Football	2 , 1 (Pure NE)	-1 , -1
Opera	-1 , -1	1 , 2 (Pure NE)

Pure NEs in GAME 1 : the battle of the sexes

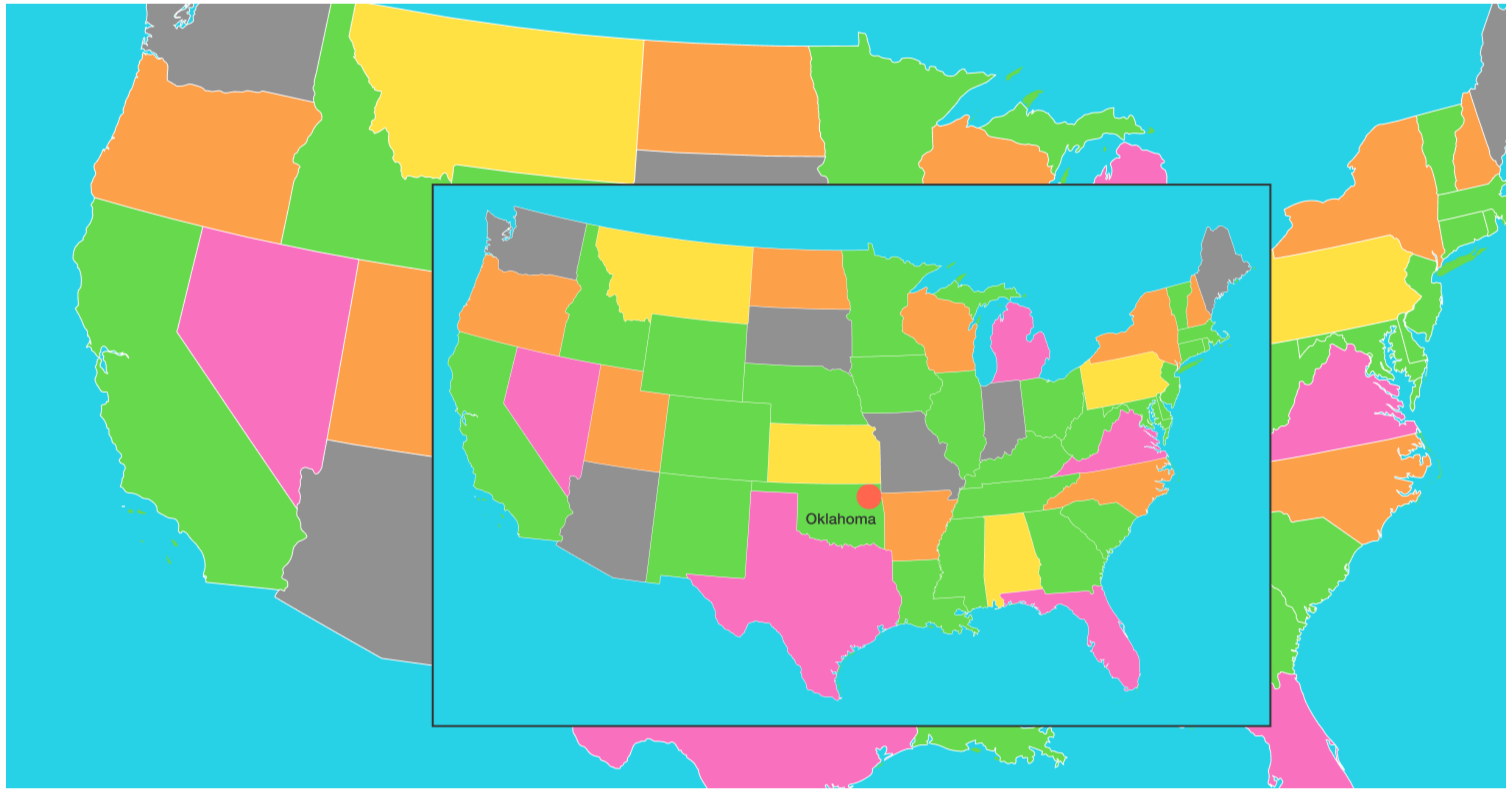
Pure and mixed NEs in the battle of the sexes

		40% Football	60% Opera
60%	Football	2 , 1 24%	-1 , -1 36%
40%	Opera	-1 , -1 16%	1 , 2 24%

expected payoff: (0.2, 0.2)

Mixed NEs in GAME 1 : the battle of the sexes

Nash equilibrium always exists



Every finite game has at least one Nash equilibrium.

(John Nash proved it with a fixed-point theorem in 1950)

Sub question #1:

Is playing a Nash equilibrium always good strategy for a rational player against other rational players?

Observation 1: Some NEs look dumb

		40% Football	60% Opera
60%	Football	2 , 1 (24%)	-1 , -1 (36%)
40%	Opera	-1 , -1 (16%)	1 , 2 (24%)

expected payoff: (0.2, 0.2)

Will you propose this mixed NE strategy to your boyfriend/girlfriend?

2-player coin matching game

	Head	Tail
Head	1 , 1	0 , 0
Tail	0 , 0	1 , 1

GAME IV : 2-player coin matching

3 NEs in total:

(Head, Head) — payoff=1

(Tail, Tail) — payoff=1

(50%Head+50%Tail, 50%Head+50%Tail) — payoff=0

Refinement of Nash equilibrium

Rational players will avoid dumb NEs
via pre-play communication, if allowed.

Pareto-Optimal NEs

Proposal #1: When existing multiple NEs, rational players will pick ones that are not Pareto dominated by others (at least one player has better payoff).

Strong NEs (Aumann, 1959)

Proposal #2: When playing the equilibrium strategies, NO (sub-)group of players can collectively change their strategies to improve the payoff for every one in the group.

Pareto-Optimal NEs vs Strong NEs

(1) Strong NEs $\subseteq$ Pareto-Optimal NEs $\subseteq$ NEs

Proof. select the group of players = all players. ∎

(2) A Pareto-Optimal NE always exists.

(3) A Strong NE might not exist.

No strong NE in Prisoner's Dilemma

	Stay silent	Betray
Stay silent	-3, -3	-10, 0
Betray	0, -10	-7, -7

Dominate (arrow from -3, -3 to -7, -7)

Pareto-Optimal NE (-7, -7 boxed)

Neither strong NE nor Pareto-optima NE is satisfactory

	Head (p3): Head	Head (p3): Tail	Tail (p3): Head	Tail (p3): Tail
Head	1, 1, -5	-5, -5, 0	-1, -1, 5	-5, -5, 0
Tail	-5, -5, 0	0, 0, 10	-5, -5, 0	-2, -2, 0

GAME V : 3-player coin matching

p1, p2: try to match each other, and prefer p3 to play Head.

p3: try to be different from p1, p2 when they match.

Strong NE is too strong: no strong NE in the game

	Head		Tail	
	Head	Tail	Head	Tail
Head	1, 1, -5	-5, -5, 0	|-1, -1, 5|	-5, -5, 0
Tail	-5, -5, 0	|0, 0, 10|	-5, -5, 0	-2, -2, 0

p1, p2 want to change

1 Pure NE

2 Pure NE

3 Mixed NE: (5/11Head+6/11Tail, 5/11Head+6/11Tail, Head)
w/ expected payoffs (-25/11, -25/11, 235/121)

Pareto-Optimal NE is too weak when more then two players

	Head: Head	Head: Tail	Tail: Head	Tail: Tail
Head	1, 1, -5	-5, -5, 0	-1, -1, 5	-5, -5, 0
Tail	-5, -5, 0	0, 0, 10	-5, -5, 0	-2, -2, 0

p1, p2 want to change

Pareto-Optimal NE

But p3 won't pick such an NE

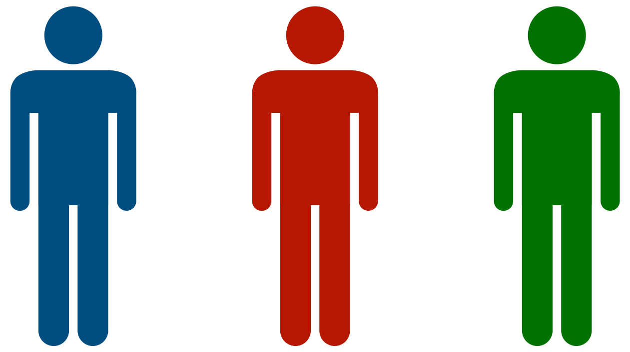
Coalition-proof Nash equilibrium (Bernheim, Peleg, and Whinston, 1987)

	Head		Tail	
	Head	Tail	Head	Tail
Head	1, 1, -5	-5, -5, 0	-1, -1, 5	-5, -5, 0
Tail	-5, -5, 0	0, 0, 10	-5, -5, 0	-2, -2, 0

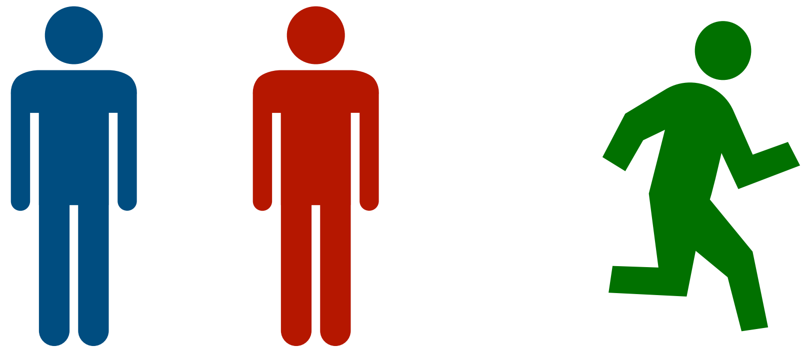
CPNE: rational players will pick it

Even though p1, p2, p3 can collectively change to make them all better,
p3 won't do that because (Tail, Tail, Head) is **NOT self-enforcing** for p1, p2.

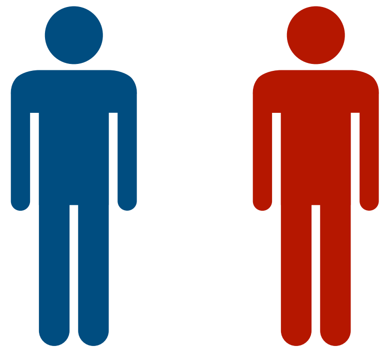
Intuition of the coalition-proof Nash equilibrium



all players meet in a room to <u>find an agreement</u> of their strategies



anyone can announce its strategy and proposal, and leave the room at anytime



but the remaining players may take its strategy as fixed, and reach a new agreement

No player wants to be the first to exit the room, **unless the agreement it wants to achieve remains no matter who leaves the room first.**

Recursive Definition of CPNE

Def. CPNEs are strategy profiles that are self-enforcing and not dominated by other self-enforcing strategy profiles.

When $n=2$, Pareto-Optimal NEs are CPNE.

When $n>2$, assume CPNE is defined for game with fewer than n players:

A strategy profile is self-enforcing, if for all subgroup of players, their strategy profile is a CPNE in the game with other players' strategy profile fixed.

Pareto-Optimal NEs vs Strong NEs vs CPNE

$$\text{Strong NEs} \subseteq \begin{matrix}\text{Pareto-Optimal NEs} \\ \text{Coalition-Proof NEs}\end{matrix} \subseteq \text{NEs}$$

PONE $\neq$ CPNE when $n>2$

CPNE is weaker than Strong NE because it only requires equilibrium strategies to be not dominated by **other self-enforcing strategies.**

CPNE may not always exist.

No coalition-proof NE in this game

	Head		Tail	
	Head	Tail	Head	Tail
Head	1, 1, -2	-1, -1, 2	-1, -1, 2	-1, -1, 2
Tail	-1, -1, 2	-1, -1, 2	-1, -1, 2	1, 1, -2

GAME VI : 3-player coin matching, version2

p1, p2: try to match each other, and mismatch p3.

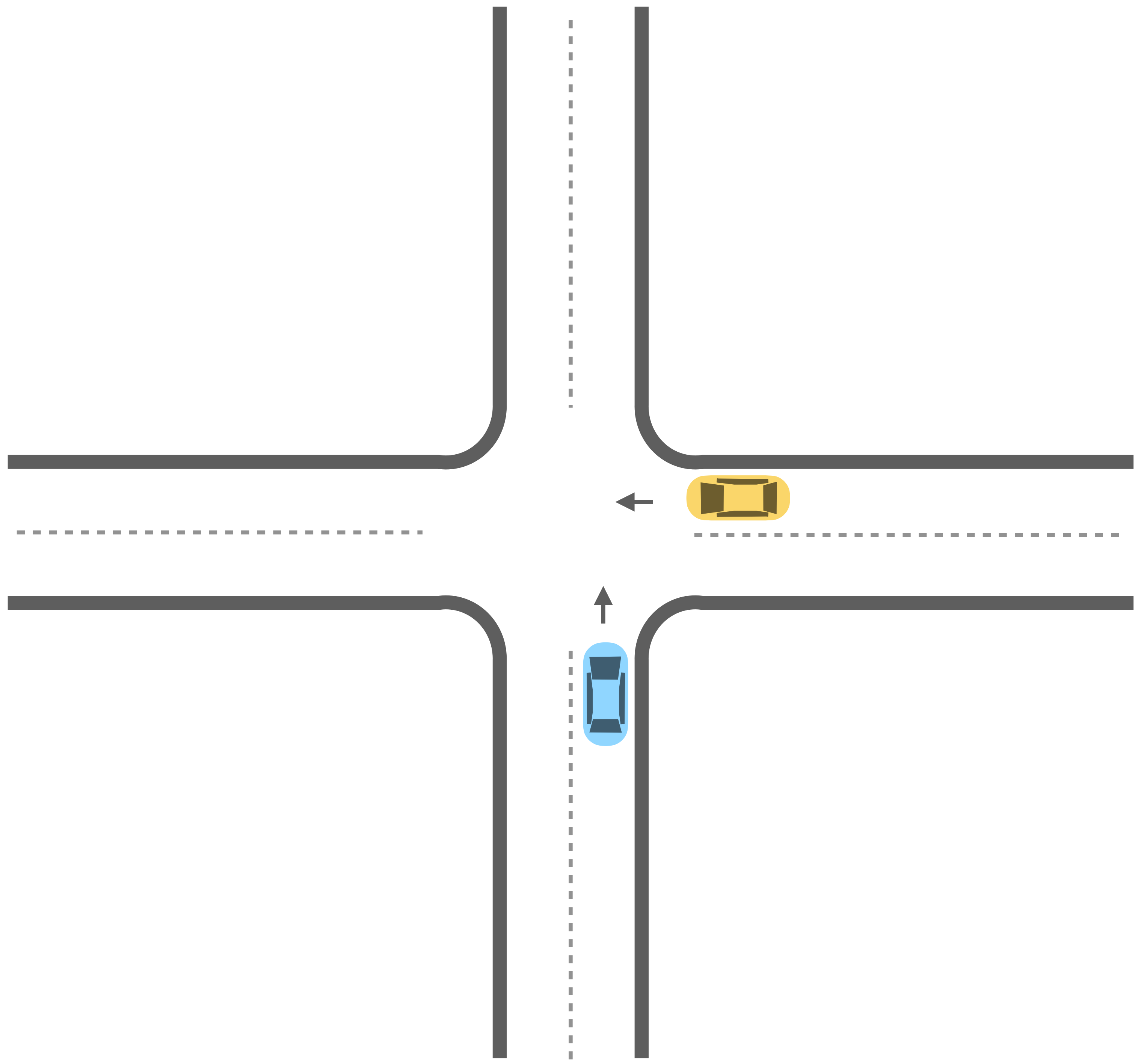
p3: try to be different from p1, p2.

Sub question #2:

Will rational players play strategies that are better than any Nash equilibria?

Observation 2: Players might not move independently

GAME VII : crossroad

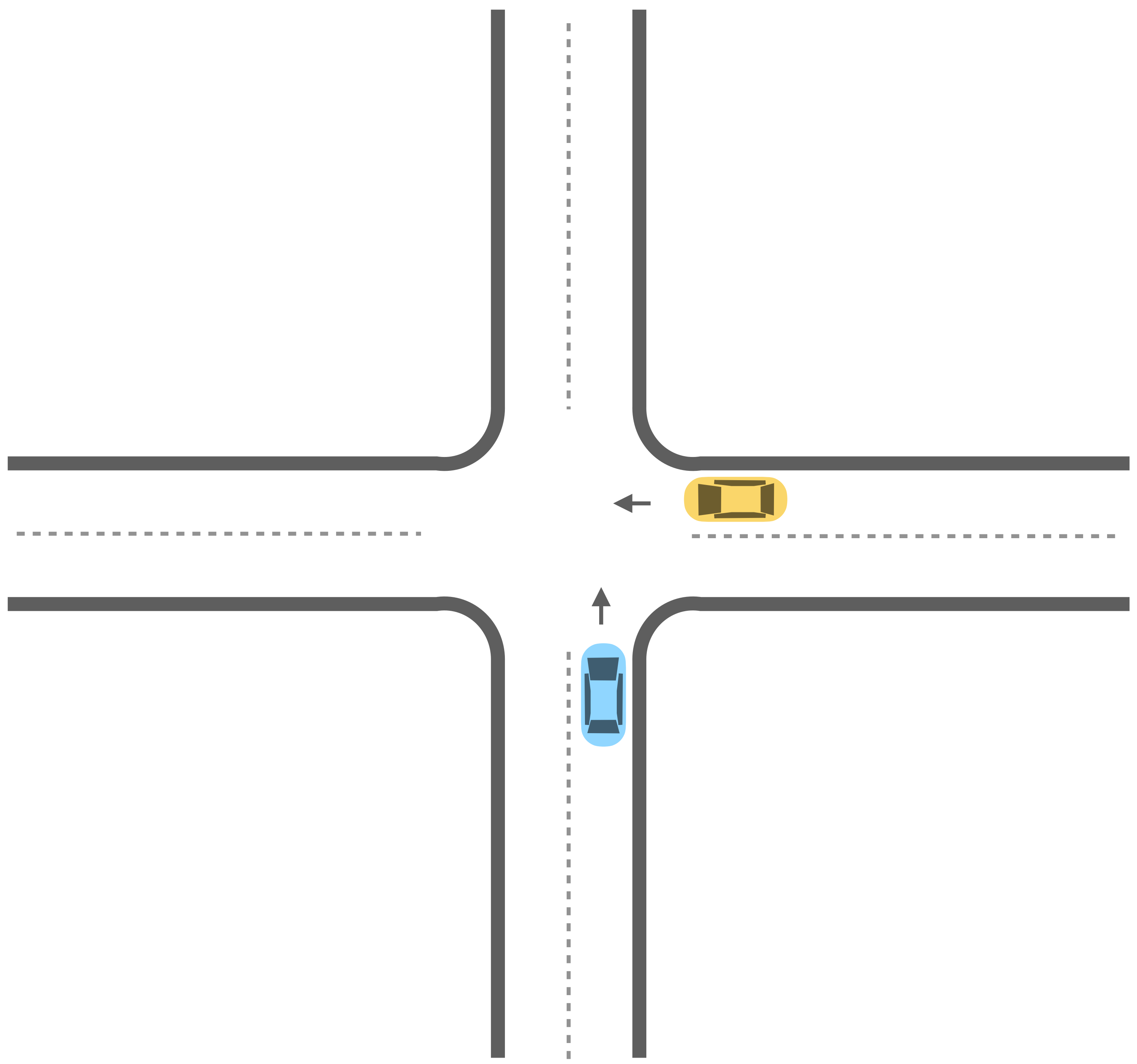


	Go	Stop
Go	-6 , -6	4 , -4
Stop	-4 , 4	2 , 2

Two bad-tempered drivers both want to pass the crossroad.
They'll be very unhappy if the other passes first.
They'll be happy if they both stop.
Very dangerous if they simultaneous choose Go.

Observation 2: Players might not move independently

GAME VII : crossroad



All NEs are Pareto Optimal

	Go	Stop
Go	-6 , -6	4 , -4
Stop	-4 , 4	2 , 2

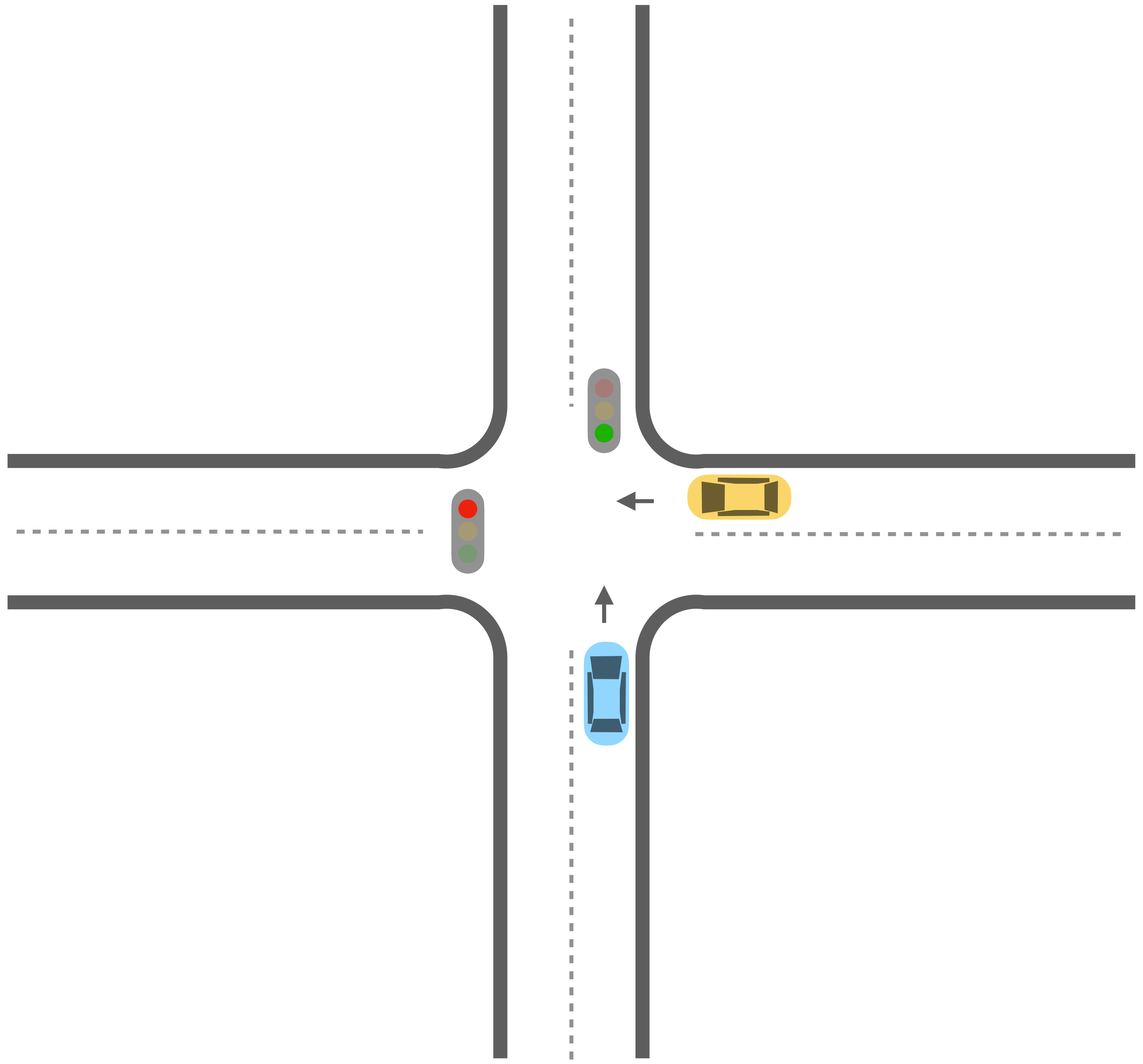
1 Pure NE: (Go, Stop) → 4 , -4

2 Pure NE: (Stop, Go) → -4 , 4

3 Mixed NE: (50%Go+50%Wait, 50%Go+50%Wait)
w/ expected payoffs (-1, -1)

Observation 2: Players might not move independently

GAME VII : crossroad



After communication, drivers setup signal lights.

If they follow the fair signal lights

	Go	Stop
Go	-6 , -6 0%	4 , -4 50%
Stop	-4 , 4 50%	2 , 2 0%

the expected payoffs are (0, 0)
which is better than mixed NE

Correlated equilibrium (Aumann, 1974)

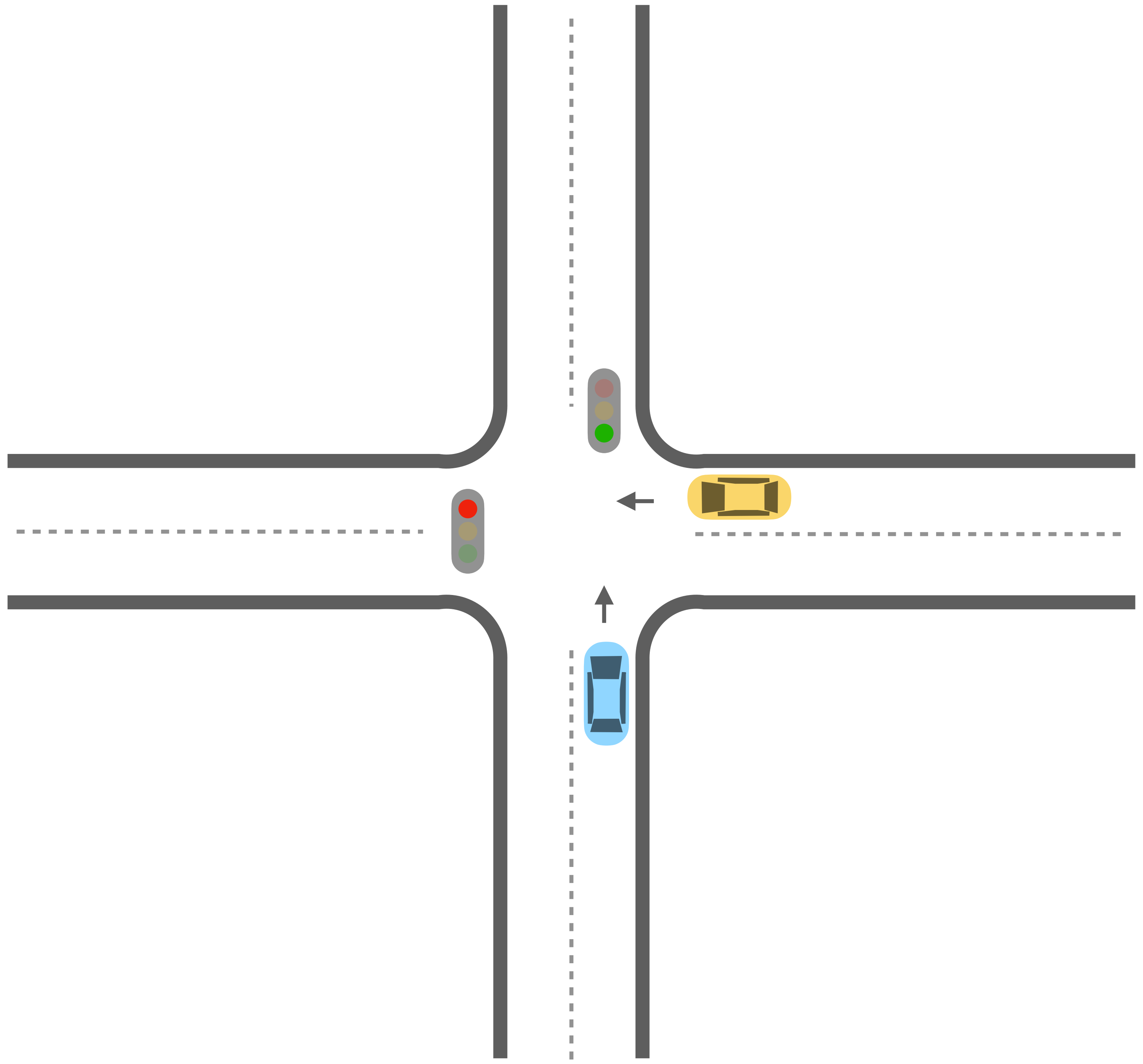
Rational players may used some third-party mediator (e.g., a coin, a signal light) to correlate their strategies.

How the mediator sends strategy suggestion to each player is known to all players — but each player can only receive its own suggestion.

Def. A strategy profile is called a "**Correlated Equilibrium**", if NO player can gain more by deviating from the received strategy suggestion.

A safer design of signal light

GAME VII : crossroad



If they follow the fair signal lights

	Go	Stop
Go	-6 , -6 0%	4 , -4 33%
Stop	-4 , 4 33%	2 , 2 33%

the expected payoffs are (2/3, 2/3)

which is higher than the fair signal light case

There are usually infinite CEs

For all player i, for all action pairs a_i and b_i,
a strategy profile σ is a CE if:

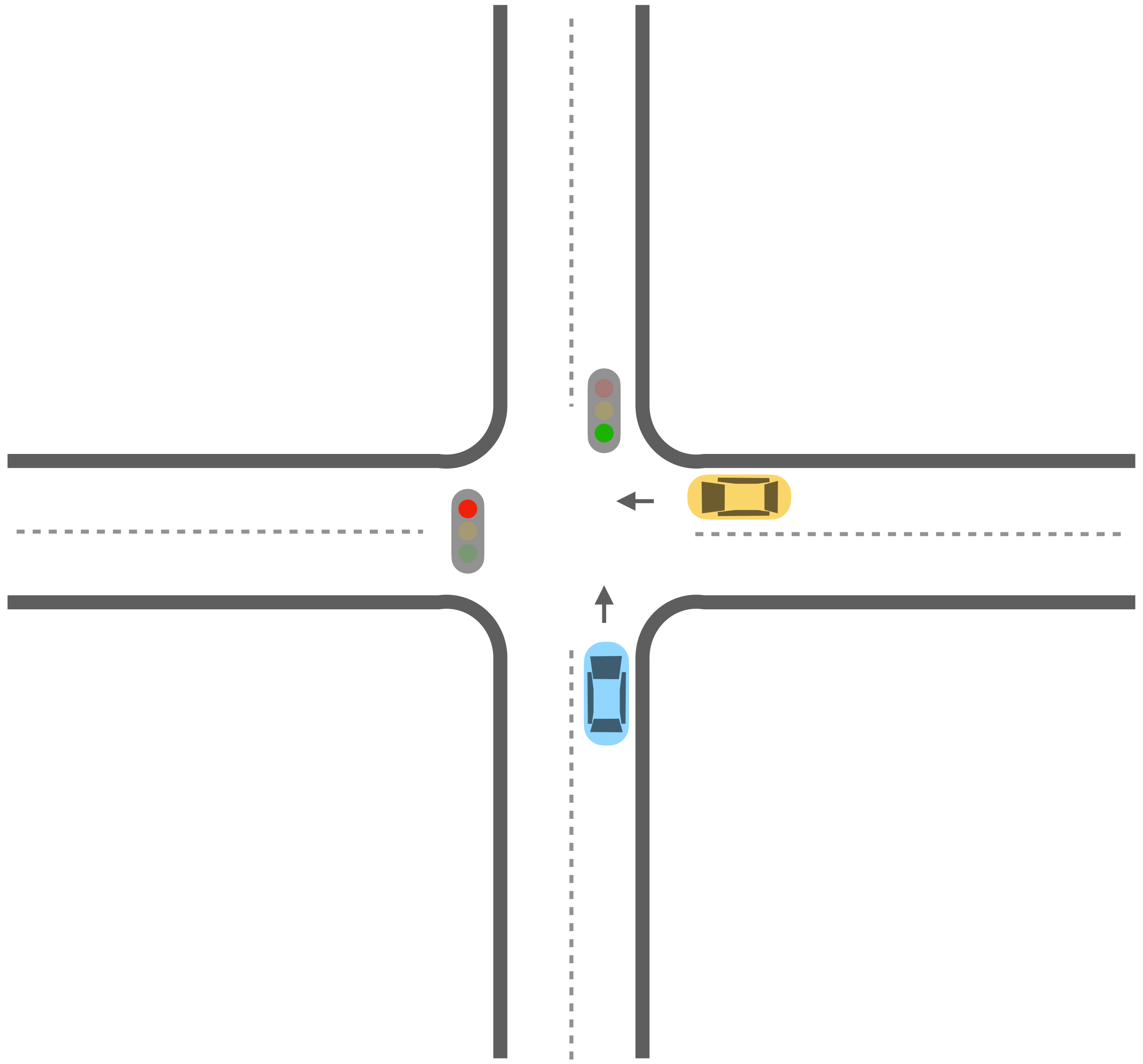
$$\sum_{\substack{\text{all action profile } a \\ \text{w/ } a_i \in a}} \sigma(a) \cdot [\underbrace{u_i(a_i, a_{-i})}_{\text{original payoff}} - \underbrace{u_i(b_i, a_{-i})}_{\text{payoff when playing otherwise}}] \geq 0$$

$$\sigma(a) \geq 1 \text{ and } \Sigma_a \, \sigma(a) = 1$$

CE is a convex polytope in an at most dim(A)-1 dimensional space
⇔ Any convex combination of CEs is a CE.

NEs ⊆ CEs

GAME VII : crossroad

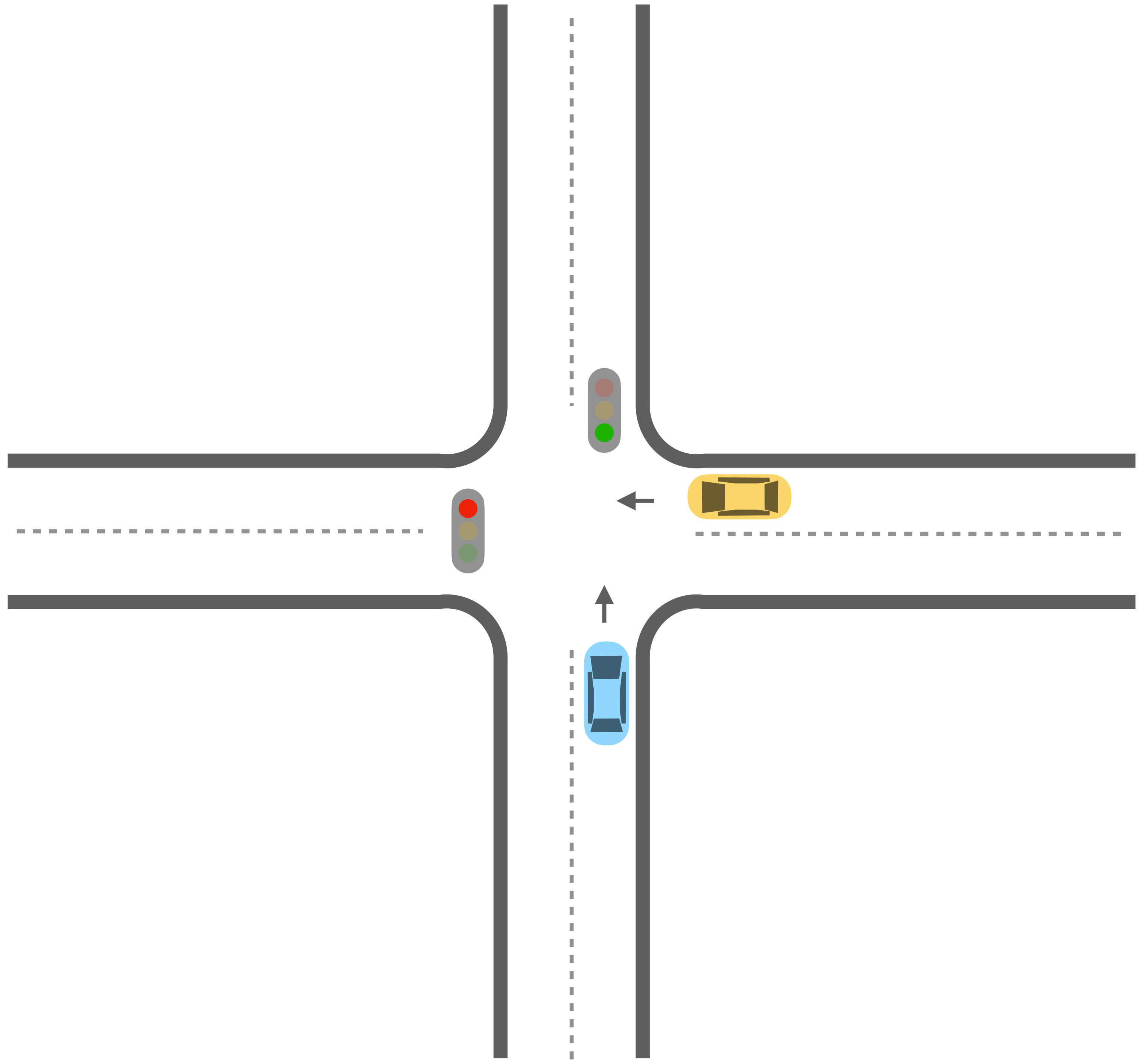


	Go	Stop
Go	-6 , -6 25%	4 , -4 25%
Stop	-4 , 4 25%	2 , 2 25%

CE is a NE when the recommendation matches an NE strategy profile.

All basis CEs in crossroad game

GAME VII : crossroad



1

	Go	Stop
Go	0%	100%
Stop	0%	0%

2

	Go	Stop
Go	0%	0%
Stop	100%	0%

3

	Go	Stop
Go	25%	25%
Stop	25%	25%

4

	Go	Stop
Go	0%	33%
Stop	33%	33%

5

	Go	Stop
Go	33%	33%
Stop	33%	0%

We cannot design a signal light with more than 33% times showing red lights on both sides. Drivers will have incentive to deviate even though they are happy to be safe.

Coalition-Proof CE (Moreno & Wooders, 1995)

Pre-play communication happens **before** the design of mediator.

No subgroup of players have an incentive to deviate from the mediator design to another self-enforcing design.

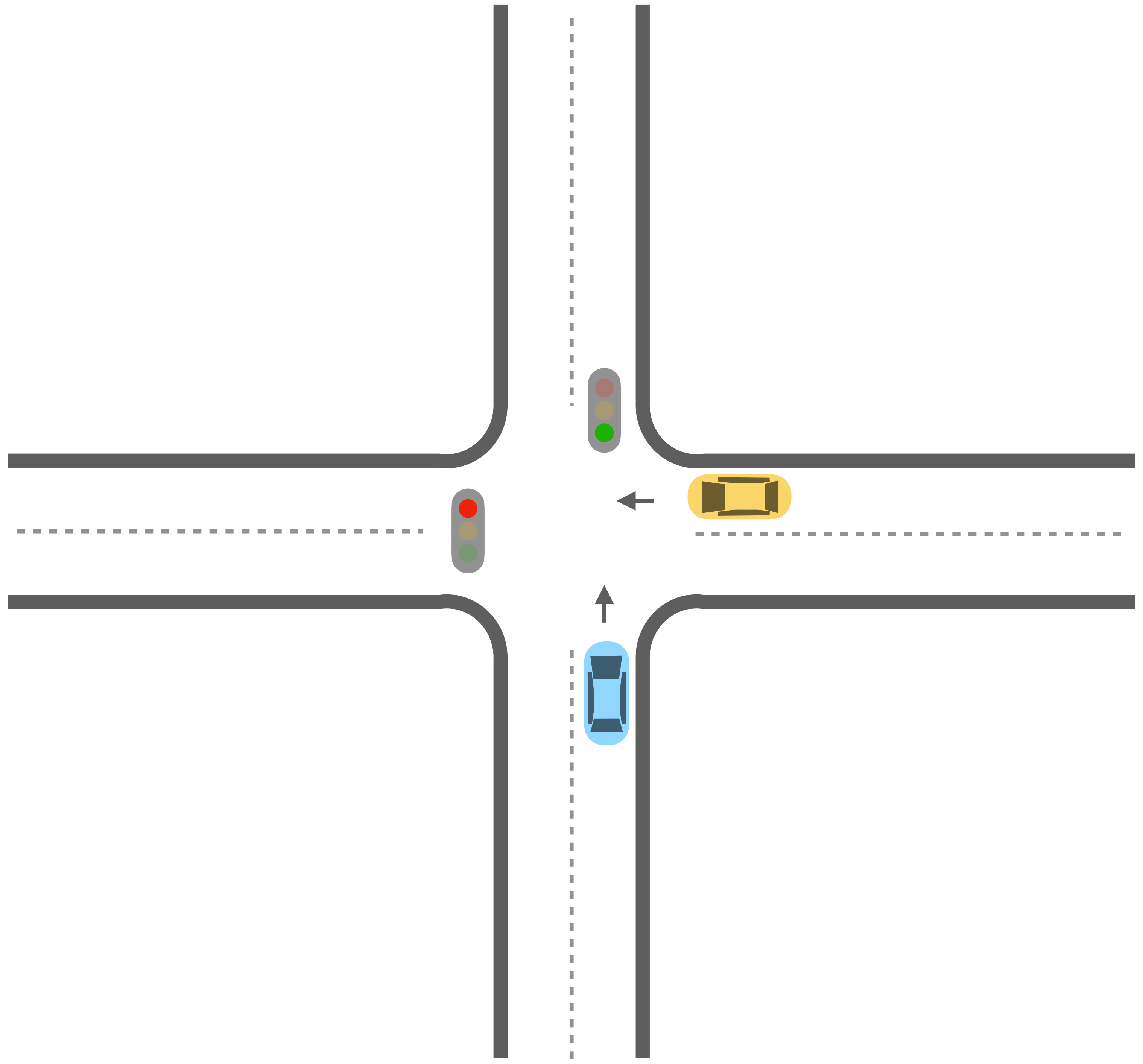
CPNE $\not\subseteq$ CPCE:

Coalition-proofness is sensitive to strategy correlation

A necessary and sufficient condition for existence is still open.

Coalition Proof CEs in crossroad game

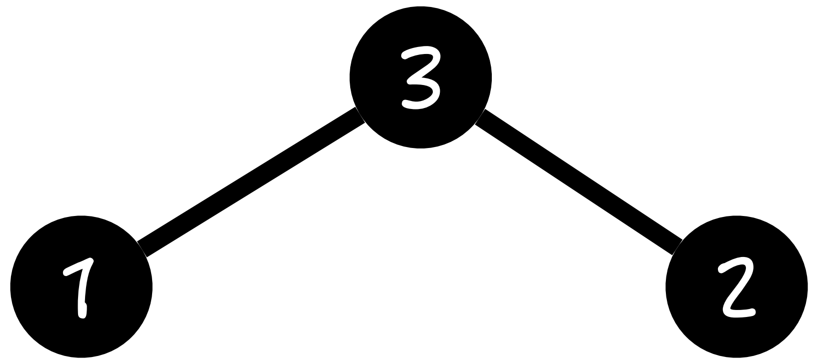
GAME VII : crossroad



1	Go	Stop
Go	0%	100%
Stop	0%	0%

2	Go	Stop
Go	0%	0%
Stop	100%	0%

3	Go	Stop
Go	0%	33%
Stop	33%	33%



Every CE lying at the boundary
1-3-2 is coalition proof

Sub question #3:
Which solution concept does AI algorithms converge to?

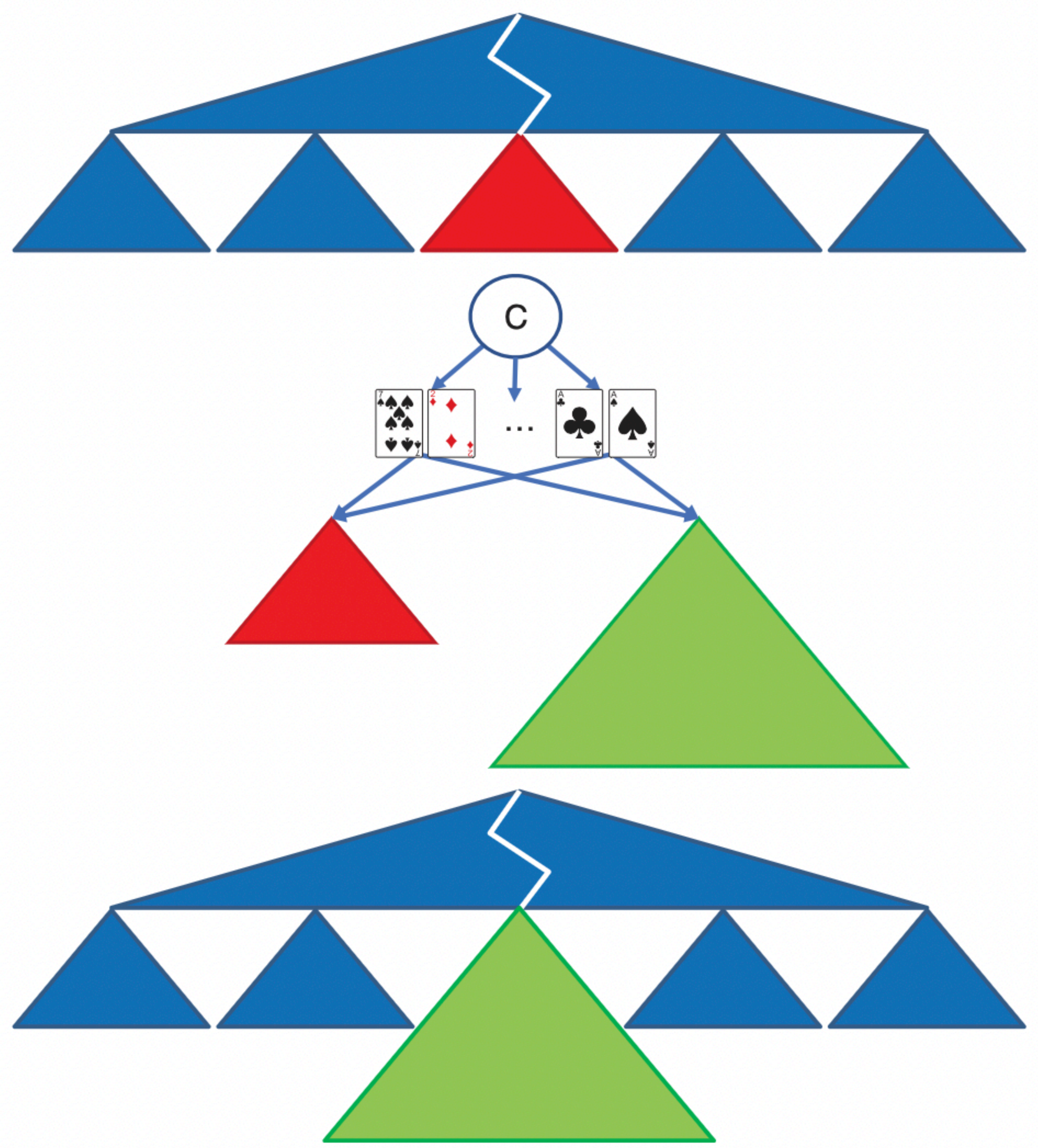
Mainly focusing on 2-player, zero-sum, symmetric games

COMPUTER SCIENCE

Superhuman AI for heads-up no-limit poker: Libratus beats top professionals

Noam Brown and Tuomas Sandholm*

No-limit Texas hold'em is the most popular form of poker. Despite artificial intelligence (AI) successes in perfect-information games, the private information and massive game tree have made no-limit poker difficult to tackle. We present Libratus, an AI that, in a 120,000-hand competition, defeated four top human specialist professionals in heads-up no-limit Texas hold'em, the leading benchmark and long-standing challenge problem in imperfect-information game solving. Our game-theoretic approach features application-independent techniques: an algorithm for computing a blueprint for the overall strategy, an algorithm that fleshes out the details of the strategy for subgames that are reached during play, and a self-improver algorithm that fixes potential weaknesses that opponents have identified in the blueprint strategy.



Libratus (2017): 2-player poker

Monte Carlo + Regret Minimization

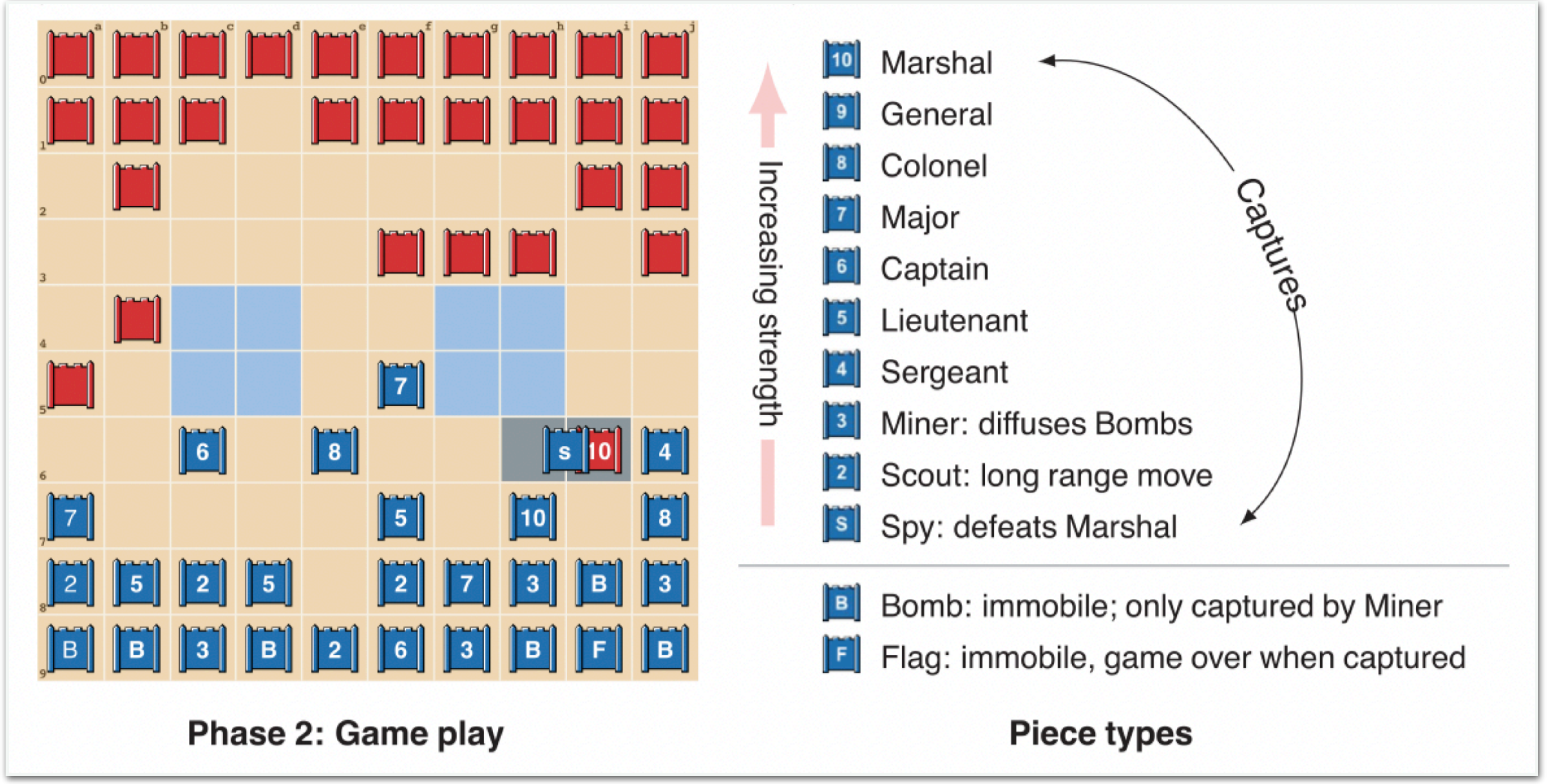
Mainly focusing on 2-player, zero-sum, symmetric games

MACHINE LEARNING

Mastering the game of Stratego with model-free multiagent reinforcement learning

Julien Perolat*†, Bart De Vylder*†, Daniel Hennes, Eugene Tarassov, Florian Strub, Vincent de Boer‡, Paul Muller, Jerome T. Connor, Neil Burch, Thomas Anthony, Stephen McAleer, Romuald Elie, Sarah H. Cen, Zhe Wang, Audrunas Gruslys, Aleksandra Malysheva, Mina Khan, Sherjil Ozair, Finbarr Timbers, Toby Pohlen, Tom Eccles, Mark Rowland, Marc Lanctot, Jean-Baptiste Lespiau, Bilal Piot, Shayegan Omidshafiei, Edward Lockhart, Laurent Sifre, Nathalie Beauguerlange, Remi Munos, David Silver, Satinder Singh, Demis Hassabis, Karl Tuyls*†

We introduce DeepNash, an autonomous agent that plays the imperfect information game Stratego at a human expert level. Stratego is one of the few iconic board games that artificial intelligence (AI) has not yet mastered. It is a game characterized by a twin challenge: It requires long-term strategic thinking as in chess, but it also requires dealing with imperfect information as in poker. The technique underpinning DeepNash uses a game-theoretic, model-free deep reinforcement learning method, without search, that learns to master Stratego through self-play from scratch. DeepNash beat existing state-of-the-art AI methods in Stratego and achieved a year-to-date (2022) and all-time top-three ranking on the Gravon games platform, competing with human expert players.



DeepNash (2022): Stratego

Mainly focusing on 2-player, zero-sum, symmetric games

A

		Player 2	
		Head: H	Tail: T
Player 1	Head: H	1	−1
	Tail: T	−1	1

B

R-NaD Iteration

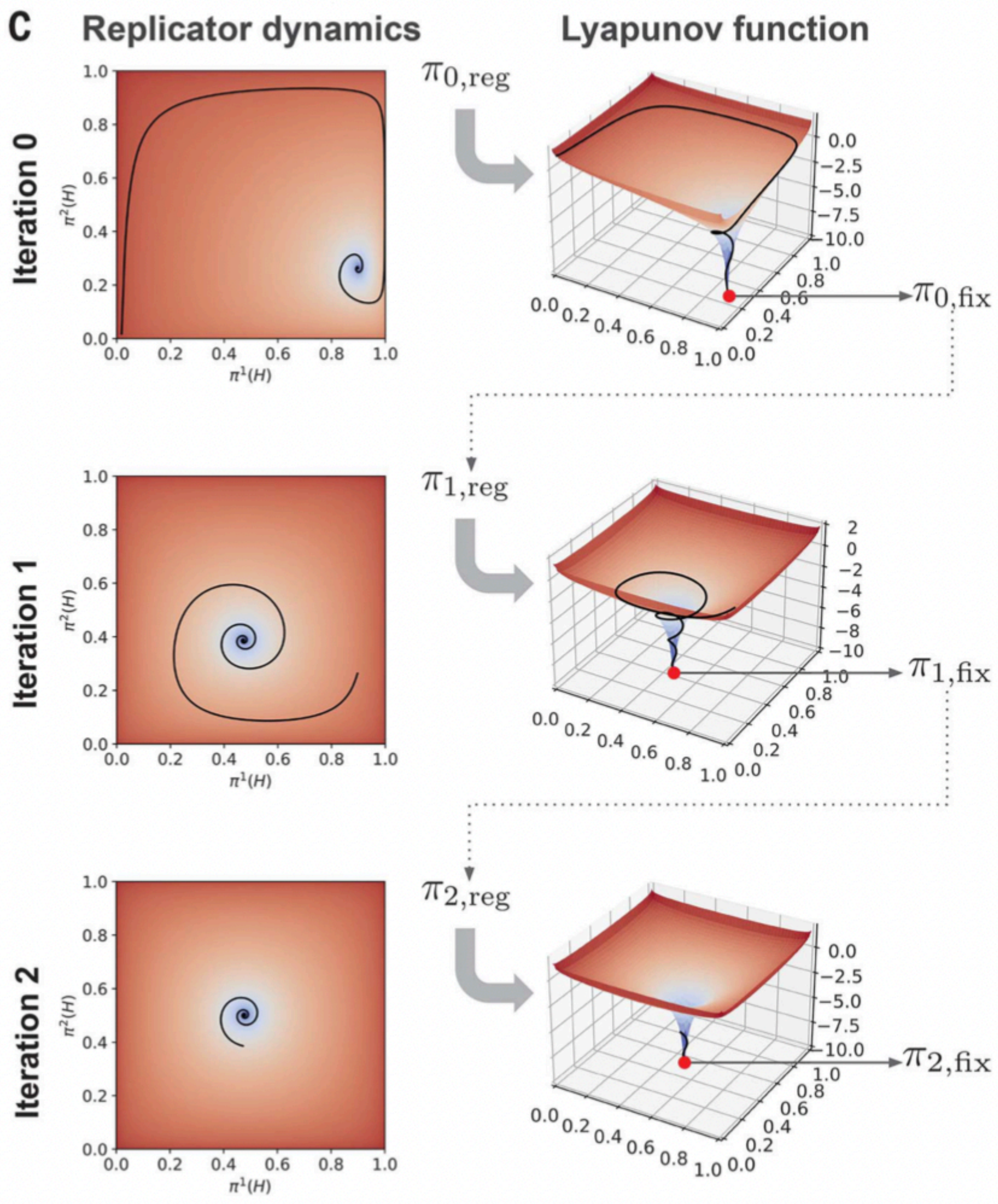
Start with an arbitrary regularization policy: $\pi_{0,\text{reg}}$

1. Reward transformation: Construct the transformed game with: $\pi_{m,\text{reg}}$
2. Dynamics: Run the replicator dynamics until convergence to: $\pi_{m,\text{fix}}$
3. Update: Set the regularization policy:

$$\pi_{m+1,\text{reg}} = \pi_{m,\text{fix}}$$

Repeat stages until convergence

C



Self-play + Reward transformation

Converging to which NE? It's not a problem···

Suppose $(\sigma_{1*}, \sigma_{2*})$ is an NE, and in zero-sum symmetric game we have:

$$u_1(\sigma_{1*}, \sigma_{2*}) = -u_2(\sigma_{1*}, \sigma_{2*}) \text{ and}$$

$$u_1(\sigma, \sigma) = u_2(\sigma, \sigma) = 0 \text{ for all } \sigma.$$

We argue $(\sigma_{1*}, \sigma_{1*})$ must be another NE, since

$$0 = u_1(\sigma_{2*}, \sigma_{2*}) \leq u_1(\sigma_{1*}, \sigma_{2*}) = -u_2(\sigma_{1*}, \sigma_{2*}) \leq -u_2(\sigma_{1*}, \sigma_{1*}) = 0$$

$\Rightarrow$ All NEs in the game have a payoff (0,0)

Converging to which NE? It's not a problem···

⇔ Even if two player plays strategies in different NEs, they must get zero reward when matched up against each other.

$$(\sigma_1^*, \sigma_2^*), (\sigma_1', \sigma_2') \in \text{NEs}$$

$$\Rightarrow (\sigma_1^*, \sigma_1^*), (\sigma_2', \sigma_2') \in \text{NEs}$$

$$0 = u_1(\sigma_2', \sigma_2') \leq u_1(\sigma_1^*, \sigma_2') = -u_2(\sigma_1^*, \sigma_2') \leq -u_2(\sigma_1^*, \sigma_1^*) = 0$$

As long as AI plays a Nash equilibrium solution, it's fine.

More than 2 players?

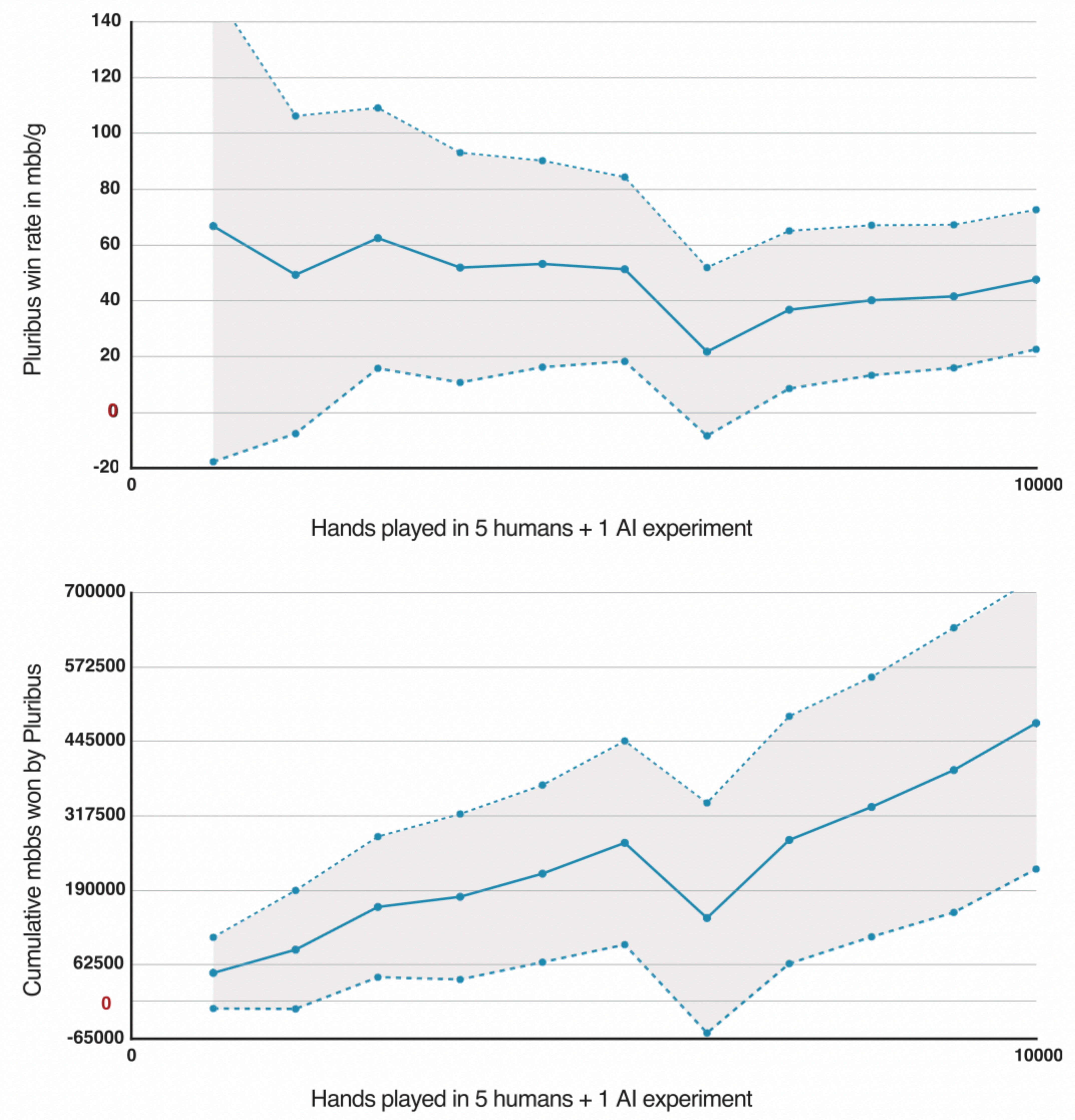
RESEARCH ARTICLE

COMPUTER SCIENCE

Superhuman AI for multiplayer poker

Noam Brown[1,2]* and Tuomas Sandholm[1,3,4,5]*

In recent years there have been great strides in artificial intelligence (AI), with games often serving as challenge problems, benchmarks, and milestones for progress. Poker has served for decades as such a challenge problem. Past successes in such benchmarks, including poker, have been limited to two-player games. However, poker in particular is traditionally played with more than two players. Multiplayer games present fundamental additional issues beyond those in two-player games, and multiplayer poker is a recognized AI milestone. In this paper we present Pluribus, an AI that we show is stronger than top human professionals in six-player no-limit Texas hold'em poker, the most popular form of poker played by humans.



"Selecting the right equilibrium is hard so let's just give up."

self-play + regret minimization

no communication

More than 2 players? With communication

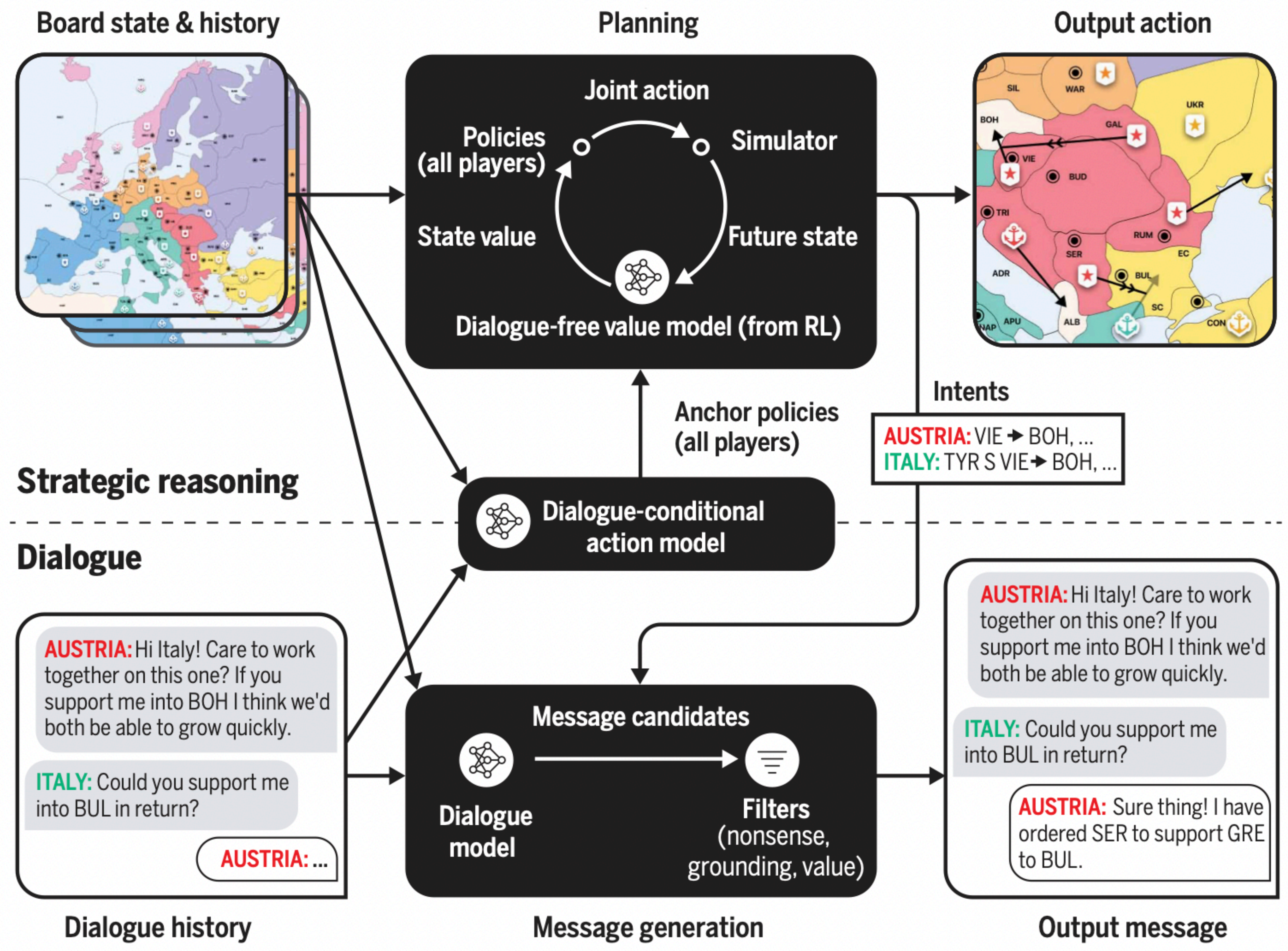
RESEARCH ARTICLE

COMPUTER SCIENCE

Human-level play in the game of *Diplomacy* by combining language models with strategic reasoning

Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin[1]‡, Noam Brown[1]*‡, Emily Dinan[1]*‡, Gabriele Farina[1], Colin Flaherty[1]‡, Daniel Fried[1,2], Andrew Goff[1], Jonathan Gray[1]‡, Hengyuan Hu[1,3]‡, Athul Paul Jacob[1,4]‡, Mojtaba Komeili[1], Karthik Konath[1], Minae Kwon[1,3], Adam Lerer[1]*‡, Mike Lewis[1]*‡, Alexander H. Miller[1]‡, Sasha Mitts[1], Adithya Renduchintala[1]‡, Stephen Roller[1], Dirk Rowe[1], Weiyan Shi[1,5]‡, Joe Spisak[1], Alexander Wei[1,6], David Wu[1]‡, Hugh Zhang[1,7]‡, Markus Zijlstra[1]

Despite much progress in training artificial intelligence (AI) systems to imitate human language, building agents that use language to communicate intentionally with humans in interactive environments remains a major challenge. We introduce Cicero, the first AI agent to achieve human-level performance in *Diplomacy*, a strategy game involving both cooperation and competition that emphasizes natural language negotiation and tactical coordination between seven players. Cicero integrates a language model with planning and reinforcement learning algorithms by inferring players' beliefs and intentions from its conversations and generating dialogue in pursuit of its plans. Across 40 games of an anonymous online *Diplomacy* league, Cicero achieved more than double the average score of the human players and ranked in the top 10% of participants who played more than one game.



Try to converge to some correlated equilibrium (no control of which) + regularization to mimic human behavior

Summary

Main Question: How would rational, self-interested players behave in a game? (still open)

How would we characterize AI's be behavior and enforce alignment to human value?

Can we build an AI beats all other AIs?

...