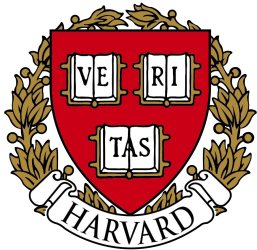


Finding Friend and Foe in Multi-agent Games

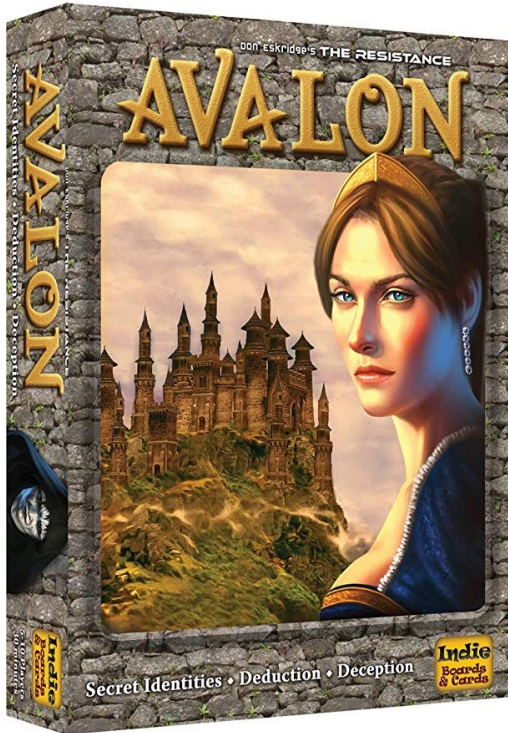
Jack Serrino*, Max Kleiman-Weiner*,
David Parkes, Josh Tenenbaum

Harvard, MIT, Diffeo

Poster #197



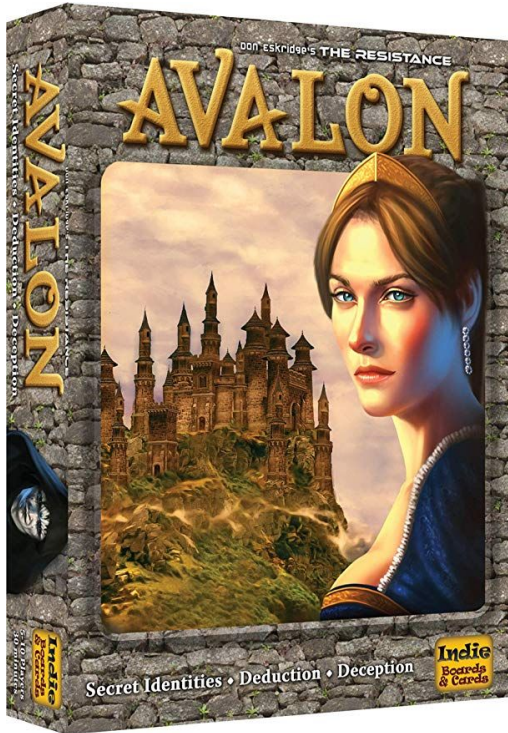
The Resistance: Avalon as a testbed for multi-agent learning and thinking



(Eskridge, 2012)

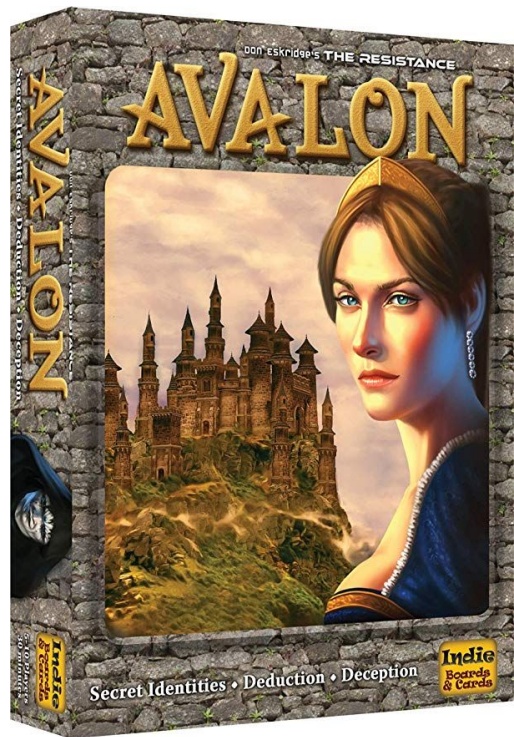
The Resistance: Avalon as a testbed for multi-agent learning and thinking

Recent progress limited to games where teams are known or play is fully adversarial (Dota, Go, Poker).



(Eskridge, 2012)

The Resistance: Avalon as a testbed for multi-agent learning and thinking

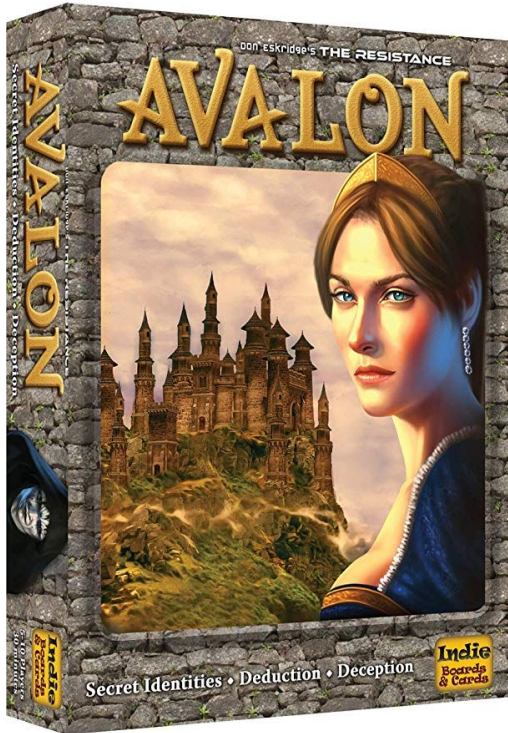


Recent progress limited to games where teams are known or play is fully adversarial (Dota, Go, Poker).

Avalon (5 Players)

- Two teams: “*Spy*” and “*Resistance*”
 - **Spies** know who is Spy and who is Resistance
 - Goal: *plan* to sabotage Resistance while hiding their own identity.
 - **Resistance** only know they are Resistance
 - Goal: *learn* who is a Spy & who is Resistance.

The Resistance: Avalon as a testbed for multi-agent learning and thinking



Recent progress limited to games where teams are known or play is fully adversarial (Dota, Go, Poker).

Avalon (5 Players)

- Two teams: “*Spy*” and “*Resistance*”
 - **Spies** know who is Spy and who is Resistance
 - Goal: *plan* to sabotage Resistance while hiding their own identity.
 - **Resistance** only know they are Resistance
 - Goal: *learn* who is a Spy & who is Resistance.

Information about intent is often noisy and ambiguous and adversaries may be intentionally acting to deceive.

Combining counterfactual regret minimization with deep value networks

- Approach follows DeepStack system developed for NL poker (Moravcik et al, 2017).

Proposal Regrets

Player 1

Proposal	R	M (2,3)	...	S (4)	A (5)
[4, 3]	.5	0		.3	0
[1, 5]	1.2	.1		0	4.3
...					
[2, 3]	2.1	3.2		0	1

Voting Regrets

All players

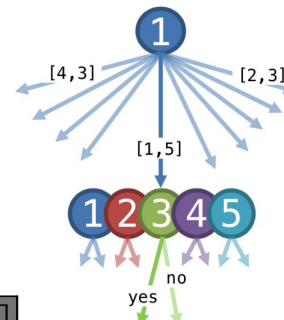
Vote	R	M (2,3)	...	S (4)	A (5)
Yes	.6	.3		.2	1
No	3.2	.1		.1	.3

Mission Regrets

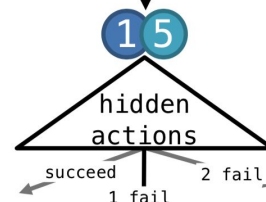
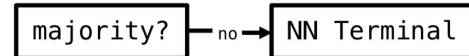
Players 1, 5

Mission	R	M (2,3)	...	S (4)	A (5)
Succeed	1.0	1.0		.1	.3
Fail	N/A	N/A		.1	.7

Proposal 1



Proposal 2



NN Terminal

Combining counterfactual regret minimization with deep value networks

- Approach follows DeepStack system developed for NL poker (Moravcik et al, 2017).

Main contributions:

- Actions themselves are only partially observed:
 - Deduction required in the loop of learning

Proposal Regrets

Player 1

Proposal	R	M (2,3)	...	S (4)	A (5)
[4, 3]	.5	0		.3	0
[1, 5]	1.2	.1		0	4.3
...					
[2, 3]	2.1	3.2		0	1

Voting Regrets

All players

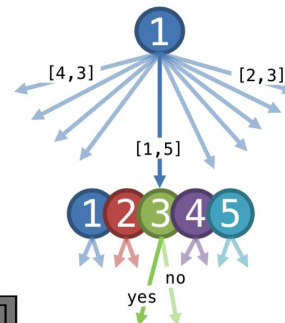
Vote	R	M (2,3)	...	S (4)	A (5)
Yes	.6	.3		.2	1
No	3.2	.1		.1	.3

Mission Regrets

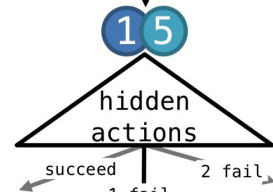
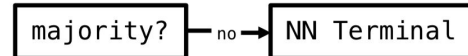
Players 1, 5

Mission	R	M (2,3)	...	S (4)	A (5)
Succeed	1.0	1.0		.1	.3
Fail	N/A	N/A		.1	.7

Proposal 1



Proposal 2



NN Terminal

Combining counterfactual regret minimization with deep value networks

- Approach follows DeepStack system developed for NL poker (Moravcik et al, 2017).

Main contributions:

- Actions themselves are only partially observed:
 - Deduction required in the loop of learning
- Unconstrained value networks are slower and less interpretable:
 - Develop an interpretable win-probability layer with better sample efficiency.

Proposal Regrets

Player 1

Proposal	R	M (2,3)	...	S (4)	A (5)
[4, 3]	.5	0		.3	0
[1, 5]	1.2	.1		0	4.3
...					
[2, 3]	2.1	3.2		0	1

Voting Regrets

All players

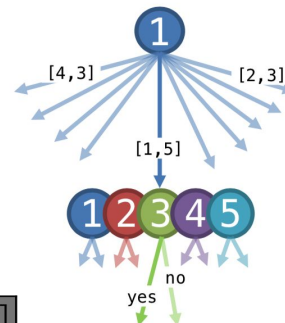
Vote	R	M (2,3)	...	S (4)	A (5)
Yes	.6	.3		.2	1
No	3.2	.1		.1	.3

Mission Regrets

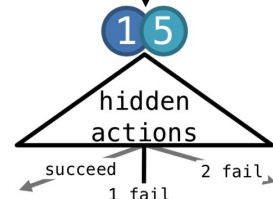
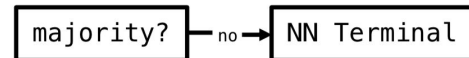
Players 1, 5

Mission	R	M (2,3)	...	S (4)	A (5)
Succeed	1.0	1.0		.1	.3
Fail	N/A	N/A		.1	.7

Proposal 1



Proposal 2



NN Terminal

Deductive reasoning enhances learning when actions are not fully public

procedure CALCTERMINALBELIEF($h, \mathbf{b}, \vec{\pi}_{1\dots p}$)

for $\rho \in \mathbf{b}$ **do**

1. $\mathbf{b}_{\text{term}}[\rho] \leftarrow \mathbf{b}[\rho] \prod_i \vec{\pi}_i(I_i(h, \rho))$

2. $\mathbf{b}_{\text{term}}[\rho] \leftarrow \mathbf{b}_{\text{term}}[\rho](1 - \mathbb{1}\{h \vdash \neg\rho\})$

▷ Zero beliefs that are logically inconsistent

end for

return \mathbf{b}_{term}

end procedure

1. Calculate joint probability of assignment given the public game history
2. Zero out assignments that are impossible given the history.

2) is not necessary in games like Poker, with fully observable actions!

The Win Layer

$$V(I, \pi^\sigma) \in \mathbb{R}^{n \times |P|}$$

$|P|$:= number of assignments to roles, ρ

n := number of players

Previous approaches:

$$NN(I, \pi^\sigma) \approx V(I, \pi^\sigma)$$

Our approach:

$$\vec{w}(I, \pi^\sigma) = \begin{bmatrix} P(\text{good win} | I, \pi^\sigma, \rho_1) \\ \vdots \\ P(\text{good win} | I, \pi^\sigma, \rho_{|P|}) \end{bmatrix} \in [0, 1]^{|P|}$$

$$V(I, \pi^\sigma) = f(\vec{w}(I, \pi^\sigma))$$

$$NN(I, \pi^\sigma) \approx \vec{w}(I, \pi^\sigma)$$

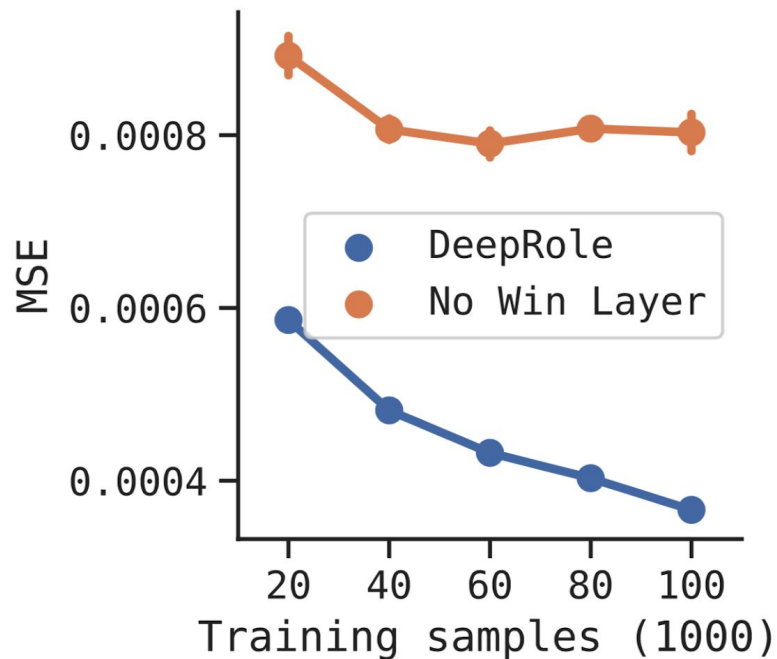
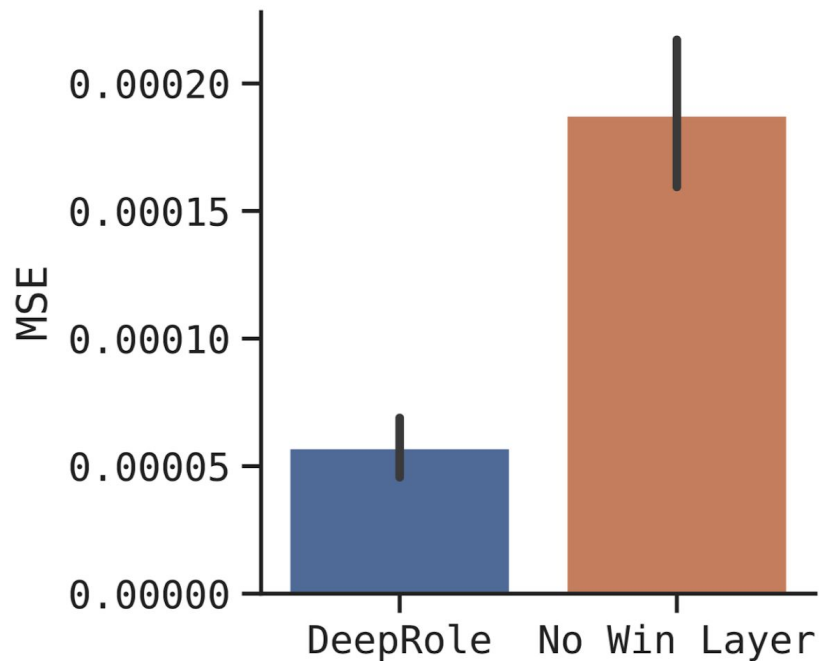
- In 5-player Avalon, **300 values to estimate!**

- Correlations are learned imperfectly.

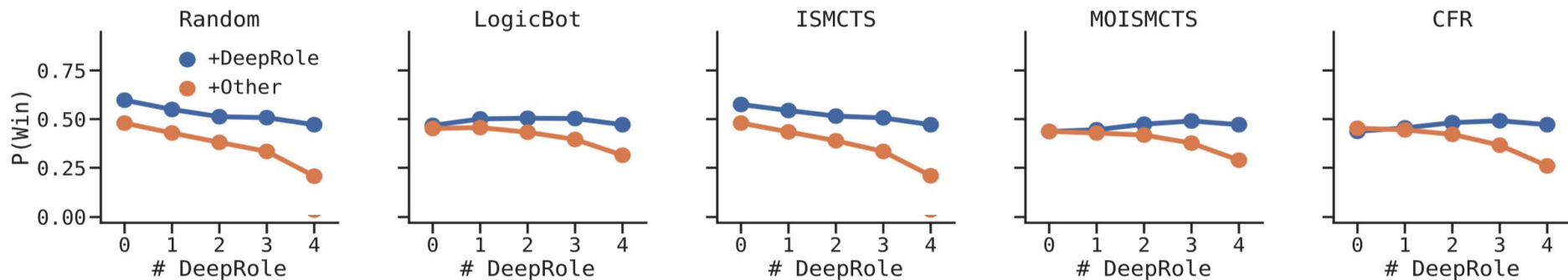
- 60 values to estimate (via sigmoid)

- Correlations are exact.

The Win Layer enables faster + better NN training



DeepRole wins at higher rates than: vanilla-CFR, MCTS, heuristic algorithms



DeepRole played online in mixed teams of human and bot players w/o communication (1,500+ games)

The screenshot displays a game interface for DeepRole. At the top, there are five player avatars: StarWarsXD (Merlin), DeepRole#49 (Resistance), DeepRole#67 (Spy), DeepRole#97 (Resistance), and DeepRole#15 (Assassin). The Spy player has a score of 2 and 3, while the Resistance players have scores of 3 and 3. A 'Claim' button is visible in the top right.

The main area shows a message: "Game has finished. The spies have won." Below this is a chat window with tabs for "All Chat", "Game Chat", "Vote History", and "Misc". The chat log contains the following messages:

- [00:46] Mission 4.1 was rejected.
- [00:46] DeepRole#67 has picked: DeepRole#67, DeepRole#15, StarWarsXD.
- [00:47] Mission 4.2 was rejected.
- [00:47] DeepRole#15 has picked: DeepRole#67, DeepRole#15, StarWarsXD.
- [00:47] Mission 4.3 was rejected.
- [00:47] DeepRole#97 has picked: DeepRole#67, DeepRole#97, DeepRole#49.
- [00:47] Mission 4.4 was rejected.
- [00:47] DeepRole#49 has picked: DeepRole#97, DeepRole#49, StarWarsXD.
- [00:47] Mission 4.5 was approved.
- [00:47] **Mission 4 succeeded.**
- [00:47] StarWarsXD has picked: StarWarsXD, DeepRole#49, DeepRole#97.
- [00:47] Mission 5.1 was approved.
- [00:47] **Mission 5 succeeded.**
- [00:47] **The assassin has shot StarWarsXD! They were correct!**
- [00:47] **The spies win!**

To the right of the chat is a mission table with columns for Mission 1 through Mission 5 and rows for each player. The table uses green for approved votes and red for rejected votes.

	Mission 1	Mission 2	Mission 3	Mission 4	Mission 5
DeepRole#67	✓	✓	✓	✓	✓
DeepRole#15	✓	✓	✓	✓	✓
DeepRole#97	✓	✓	✓	✓	✓
DeepRole#49	✓	✓	✓	✓	✓
StarWarsXD	✓	✓	✓	✓	✓

Below the mission table is a "Card history" section with a legend:

- Leader
- On mission team
- Investigator
- Excalibur
- Excalibured
- Approved vote
- Rejected vote

DeepRole outperformed humans playing online as both a collaborator and competitor

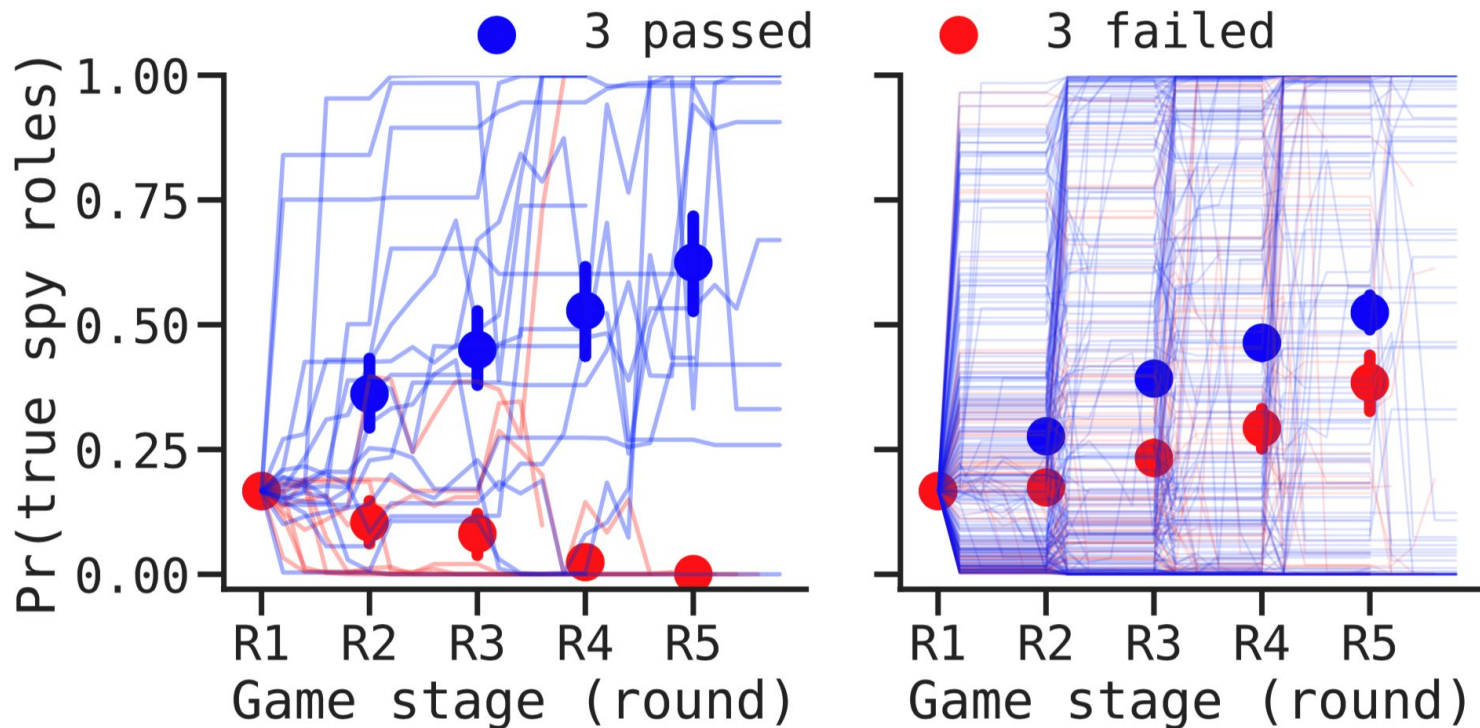
Adding DeepRole or a Human
to 4 DeepRole

	+DeepRole		+Human	
	Win Rate (%)	(N)	Win Rate (%)	(N)
Overall	46.9 ± 0.6	(7500)	38.8 ± 1.3	(1451)
Resistance	34.4 ± 0.7	(4500)	25.6 ± 1.5	(856)
Spy	65.6 ± 0.9	(3000)	57.8 ± 2.0	(595)

DeepRole outperformed humans playing online as both a collaborator and competitor

	Adding DeepRole or a Human							
	to 4 DeepRole				to 4 Human			
	+DeepRole		+Human		+DeepRole		+Human	
	Win Rate (%)	(N)	Win Rate (%)	(N)	Win Rate (%)	(N)	Win Rate (%)	(N)
Overall	46.9 ± 0.6	(7500)	38.8 ± 1.3	(1451)	60.0 ± 5.5	(80)	48.1 ± 1.2	(1675)
Resistance	34.4 ± 0.7	(4500)	25.6 ± 1.5	(856)	51.4 ± 8.2	(37)	40.3 ± 1.5	(1005)
Spy	65.6 ± 0.9	(3000)	57.8 ± 2.0	(595)	67.4 ± 7.1	(43)	59.7 ± 1.9	(670)

DeepRole make rapid accurate inferences about human roles during play and observation



Finding Friend and Foe in Multi-agent Games

Jack Serrino*, Max Kleiman-Weiner*,
David Parkes, Josh Tenenbaum

Harvard, MIT, Diffeo

Poster #197

Play online: ProAvalon.com

