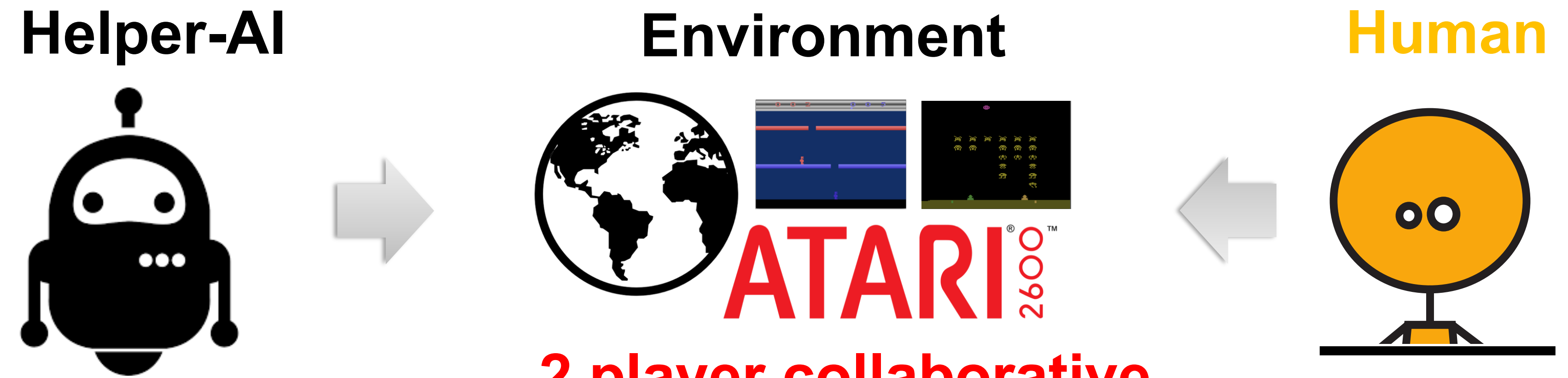
Learning Robust Helpful Behaviors in Two-Player Cooperative Atari Environments

Introduction.

- We study the problem of learning helpful behavior: **learning to cooperate with differently-skilled** and diverse **partners** in the context of two-player, cooperative **Atari games**.
- We show **robust** performance of these *Helper-AIs* when paired **with different kinds of partners** (both **human** and **artificial** agents), including partners that they have not previously encountered during training.

Helper-AI for Cooperative Atari 2600.

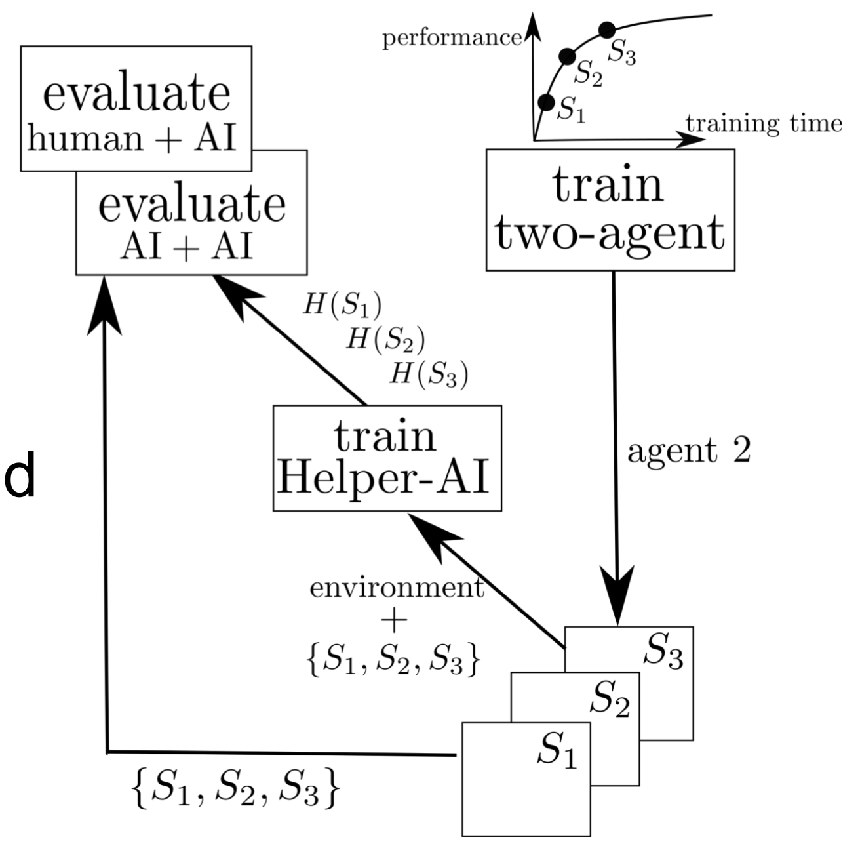


2 player collaborative

Agents

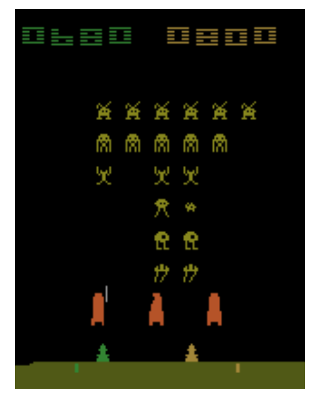
We use the **ACKTR** algorithm for reinforcement learning.

- S_1, …, S_4 - agents trained in self-play
- $H(S_i)$ - Helper-AI targeted for S_i
- $bH(S_i)$ - *Bounded* Helper-AI with limited number of training steps
- $rH(S_i)$ - Helper-AI with randomized starting positions in training
- *Intervention-AI* – AI that can override an action of its partner at some cost

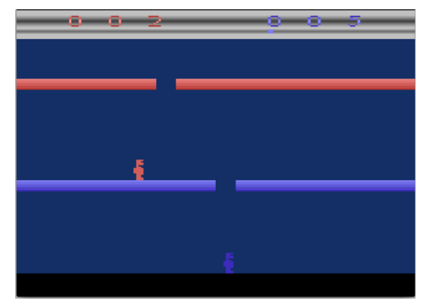


Workflow

Environments



Space Invaders



Fall Down

Intervention-AIs with AIs.

Two player Space Invaders	**Partner AI**		
	S_1	S_2	S_3
with self	878	1,134	2,141
with S_4	694	963	1,826
with Helper-AI	1,701	2,434	3,844
with Intervention-AI (0.05 cost)	1,772	2,534	3,985
with Intervention-AI (0.025 cost)	1,927	3,234	4,367

Two player Fall Down	**Partner AI**		
	S_1	S_2	S_3
with self	46.0	77.4	120.2
with S_4	32.7	44.3	79.1
with Helper-AI	63.8	93.7	151.9
with Intervention-AI (0.05 cost)	91.5	104.1	182.9
with Intervention-AI (0.02 cost)	111.8	151.6	203.8

Evaluating Intervention-AIs in *Cooperative Space Invaders* and *Cooperative Fall Down:*

- Game score, averaged over 100 games, of pairing a partner (columns) with different agents (rows): whether paired with self, a higher-skilled agent, an *on-target* Helper-AI, or an *on-target* Intervention-AI.
- We see a further **advantage from Intervention-AIs over Helper-AIs**, and **even though interventions incur a per-action cost**. For Fall Down, especially, the Intervention-AI provides a large boost in performance.

Robust Helper-AI Behavior.

	The Behavior of the Partner Agent				
	S_1	S_2	S_2-close	S_2-distant	S_3
Performance with self	878	1,134	1,111	1,141	2,141
… with expert-skill agent	694	963	457	711	1,826
with Helper-AI trained for different target behaviors					
$H(S_1)$	$\boxed{1,701}$	2,294	1,185	1,449	3,538
$H(S_2)$	1,587	$\boxed{2,434}$	1,227	1,548	3,792
$H(S_2\text{-close})$	1,254	1,836	$\boxed{1,932}$	1,405	2,733
$H(S_2\text{-distant})$	1,414	2,197	1,210	$\boxed{2,375}$	3,838
$H(S_3)$	1,282	2,204	1,220	1,670	$\boxed{3,844}$
$bH(S_2)$ (a bounded helper)	1,337	2,148	1,193	1,550	3,009

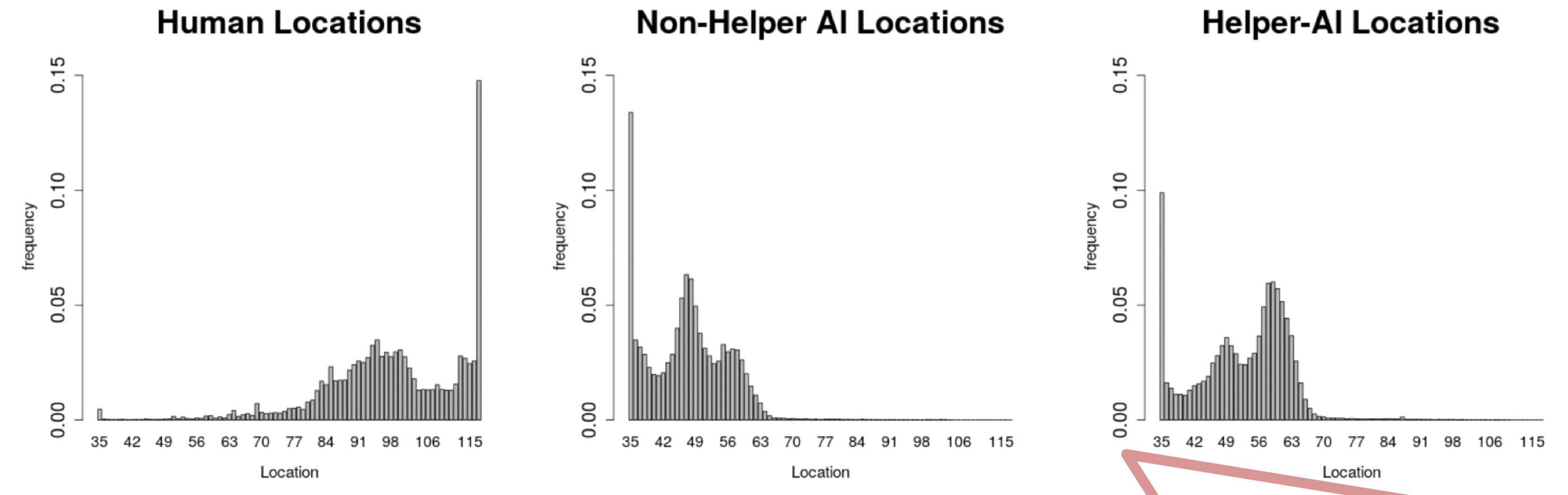
Results for ***Cooperative Space Invaders***. We modify the standard version of the game to make it cooperative.

1. **Helpful behavior vs. expert behavior**:
   - Pairing an agent with an **expert-skill agent consistently reduces performance** relative to self-pairing.
   - There is decisive and consistent performance **improvement** from pairing **an AI with its *on-target* Helper-AI**.
2. **Robust helpful behavior**:
   - There is a consistent **improvement** in performance when pairing an AI with an ***off-target* Helper- AI** than compared to the performance from self-pairing.
3. **Robust helpful behavior, bounded helpers**:
   - The ***bounded*-Helper-AI**, $bH(S_2)$, provides a consistent **improvement** in performance for partner agents relative to self-pairing.

Results for ***Cooperative Fall Down***. The standard version of the game is modified to incentivize cooperative play.

	Partner AI		
	S_1	S_2	S_3
with self	46.0	77.4	120.2
with S_4	32.7	44.3	79.1
with Helper-AI	63.8	93.7	151.9

Understanding Helper-AI Behavior.



Reason for Episode Ending	Player 1	Player 2	Observed Probability
Player 1 Hit	S_2	S_2	40%
Player 2 Hit	S_2	S_2	38%
Both Players Hit	S_2	S_2	4%
Aliens Land	S_2	S_2	18%
Player 1 Hit	S_4	S_2	33%
Player 2 Hit	S_4	S_2	42%
Both Players Hit	S_4	S_2	0%
Aliens Land	S_4	S_2	25%
Player 1 Hit	$H(S_2)$	S_2	19%
Player 2 Hit	$H(S_2)$	S_2	60%
Both Players Hit	$H(S_2)$	S_2	6%
Aliens Land	$H(S_2)$	S_2	15%

- Locations in two-player, *Cooperative Space Invaders*. Human-subjects start at location 117 (at the right) and the AIs start at location 35 (at the left).
- **Helper-AIs** tend to **spend less time at their initial location and play more in the center of the screen.**

- Reasons for episode termination in two-player, *Cooperative Space Invaders* over 100 games, with partner AI S_2, and varying the agent used in the role of Player 1.
- When Player 1 is replaced with Helper-AI $H(S_2)$, **overall miscoordination goes does down** to 15%.

Helper-AI Transfer to Human Partners.

AI Agent	Paired with S_2	Paired with Humans
S_2	1,134	704
S_4	963	545
$H(S_2)$	2,434	1,547
$bH(S_2)$	2,148	1,083
$H(S_2)$	-	950 (shock environment)
$rH(R_2)$	-	1,260 (shock environment)

- Comparative performance in *Cooperative Space Invaders* when **pairing AIs** with S_2 (another AI) or ten different **human subjects**.
- The decisive **performance advantage of the Helper- AIs**, compared with pairing with either S_2 or S_4, **holds up in transferring to this human environment**.
- The bottom half of the table reports results for $H(S_2)$ and the randomized-start position Helper-AI, $rH(S_2)$, in a setting where the human subjects are sometimes randomly teleported to different positions and sometimes asked to do something unexpected for a period of time.

Paul Tylkin ptylkin@g.harvard.edu
Goran Radanovic gradanovic@mpi-sws.org
David C. Parkes parkes@eecs.harvard.edu

HARVARD
John A. Paulson
School of Engineering
and Applied Sciences

MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

References:

- *Dimitrakakis et al., Multi-View Decision Processes: The Helper-AI Problem, NIPS'17.*
- *Wu et al., Scalable Trust-Region Method for Deep Reinforcement Learning using Kronecker-Factored Approximation, NIPS'17.*