Uncertainty-Driven Adaptive Exploration

Towards Enabling Agent Self-Evolution & Organizational Knowledge via Individual Experiences

Leonidas Bakopoulos & Georgios Chalkiadakis

MAs-ML

Technical University of Crete

Download published proceedings paper version at:
https://www.ifaamas.org/Proceedings/aamas2026/pdfs/HQII4250.pdf

TECHNICAL UNIVERSITY OF CRETE

AAMAS 2026 PAPHOS • CYPRUS 25 YEARS

[Safe] Uncertainty-Driven Adaptive Exploration

Towards Enabling Agent Self-Evolution & Organizational Knowledge via Individual Experiences

Leonidas Bakopoulos & Georgios Chalkiadakis

MAs-ML

Technical University of Crete

Download published proceedings paper version at:
https://www.ifaamas.org/Proceedings/aamas2026/pdfs/HQII4250.pdf

Motivation:

An agent should remember its past, to extend its knowledge

WHAT THE AGENT REMEMBERS

The memory of an agent is stored in the weights of its neural network (NN).

By using its NN, the agent can either:

EXPLOIT Recall its memory

EXPLORE Create a new memory

Too little exploration
Knowledge never expands.

Excessive exploration
Catastrophic forgetting of prior learning.

Motivation ++



In long horizon tasks, efficient and effective exploration is the key.

A common approach: push the agent to explore

A better approach: Decide ...

- When to explore ?
- When to trust your memory / gained knowledge?

Further Motivation / Multiagent Extension: Towards Organizational Knowledge

Starting from individual exploration and uncertainty...

Create a shared agents' memory, also encompassing their uncertainty

Effective and Efficient Multiagent Exploration
Avoiding redundant exploration

Our Contributions

The ADEU Framework

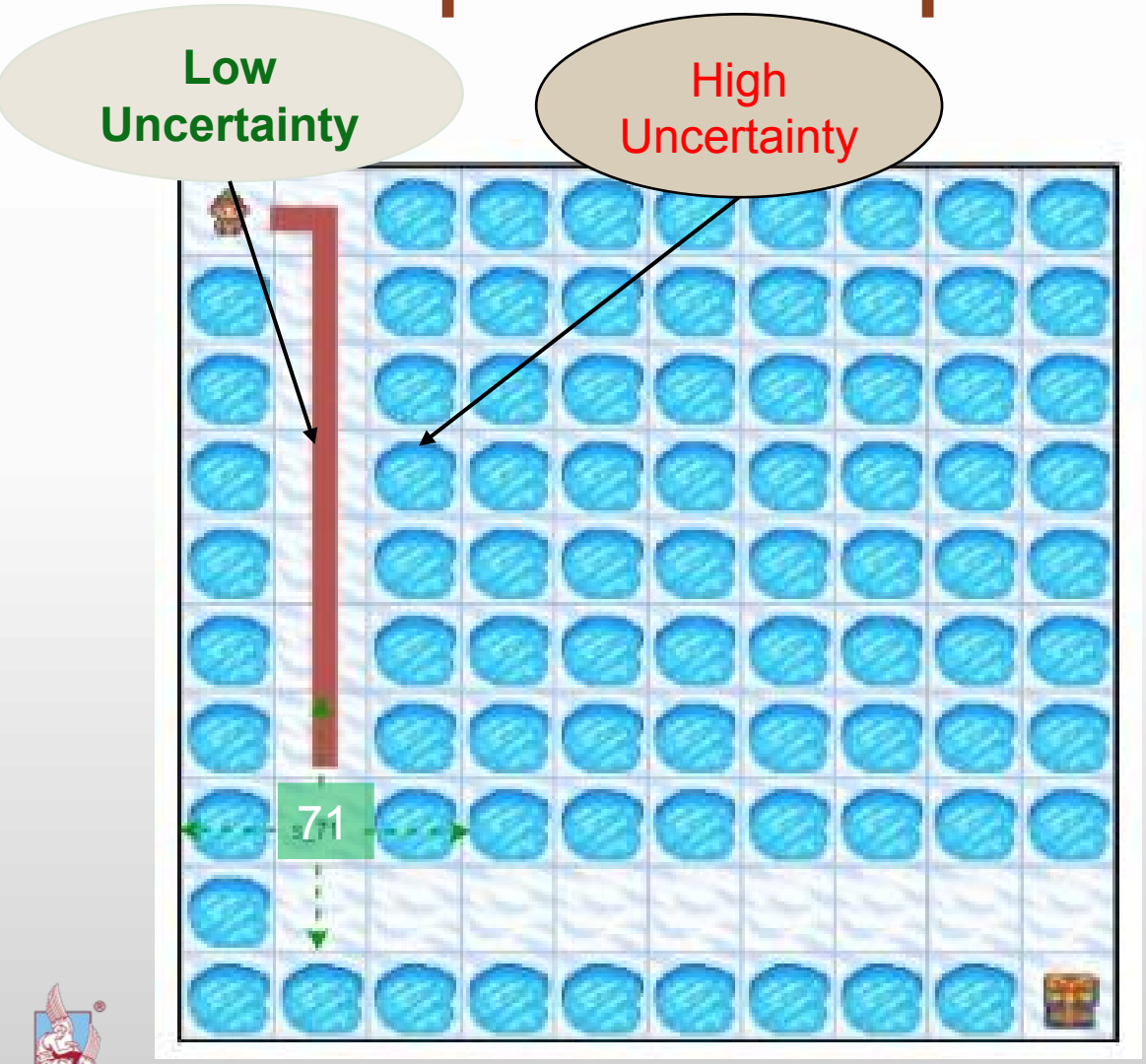
A generic ADaptive Exploration framework

- **Balances** exploration and exploitation in a **principled** way; i.e.,
- Exploration-exploitation alternation leveraged via **Uncertainty**

- ➢ Incorporates **any** uncertainty measuring mechanism
- ➢ Incorporates **all existing** adaptive exploration approaches (as special cases)
- ➢ Gives rise to **new** approaches
- ➢ …e.g., **safe exploration** ones

- ❖ Optimality guarantees in a general class of games
- ❖ Improved performance compared to standard methods

Adaptive Exploration in a nutshell



Standard current practices:

explore at a constant rate, or
select action to maximise uncertainty

✕ **Drown**

Adaptive Exploration:

Low uncertainty → Exploit
High uncertainty → Explore

√ **Reach 71**

Key Challenge:

*Determine **when** to alternate;*
*i.e., **when** uncertainty is low/high*

The ADEU Framework **AD**aptive **E**xploration via **U**ncertainty

action distribution

mechanism to measure uncertainty in state **s**

$$a \sim \mathcal{D}\left(\pi(s), g(f(s))\right)$$

(Eq. 1)

Normalization function

(current) policy at state **s**

Key intuition: distribution's spread ⇔ agent's uncertainty regarding $\pi(s)$

Generality of ADEU

***f(s)* candidate uncertainty-measuring mechanisms**

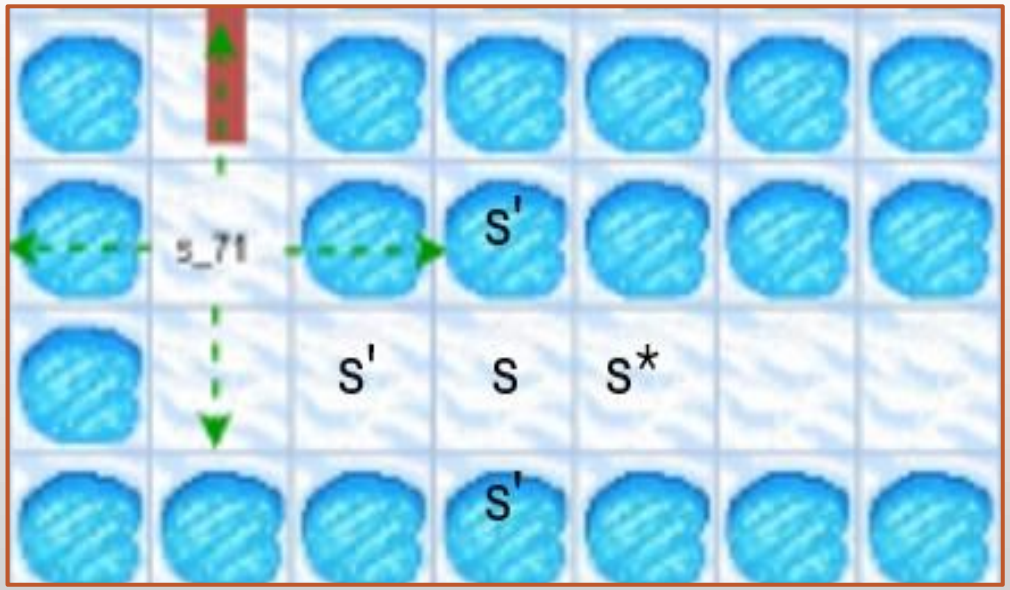
- *quantify novelty* in Intrinsic Motivation-based exploration
- *measure confidence in Q* in Uncertainty-Driven exploration
- can *emulate **all** known* adaptive exploration methods!
  - with sophisticated *f(s,…)* ; i.e., *f(s)* makes complex uncertainty measurements as required

ADEU utilizes these mechanisms to alternate between exploration and exploitation within the same episode

Theoretical Guarantees

Definition [Increasing-Reward Single-Agent Games (IR-SAGs] :

- In each state s, there is a unique future state s' = s* giving rise to a unique (s* , a*) state-action pair with which the agent achieves higher reward than the maximum reward in current state s.
- For *all other future states* (s' ≠ s*) , the maximum reward max(s' , a') is *lower* than the maximum reward of the current state
- For the current state *s*, there is a unique action b that maximizes the reward function in that state.
- When b is executed, the agent will be transferred to state s*



Increasing Reward Single Agent Games (IR-SAGs)

- A Generic Class of Games
  - Sorting
  - Ordering
  - Searching for a sole trajectory
- Reward motivates the agent to visit the true high value states

Theoretical Guarantees: In IR-SAGs…

1. An agent acts optimally *if and only if* it selects action *b*.
   - No added exploration is required

2. An ADEU agent that has executed action *b* in a state *s*, will reduce the degree of exploration in that state *s* proportionally to the effectiveness of *f(s)*

Further Guarantees



Assuming an efficient *f(s)*,

we guarantee efficient exploration

In **sparse** IR-SAG

- via defining hyper states and hyper actions

In **stochastic** versions of IR-SAG

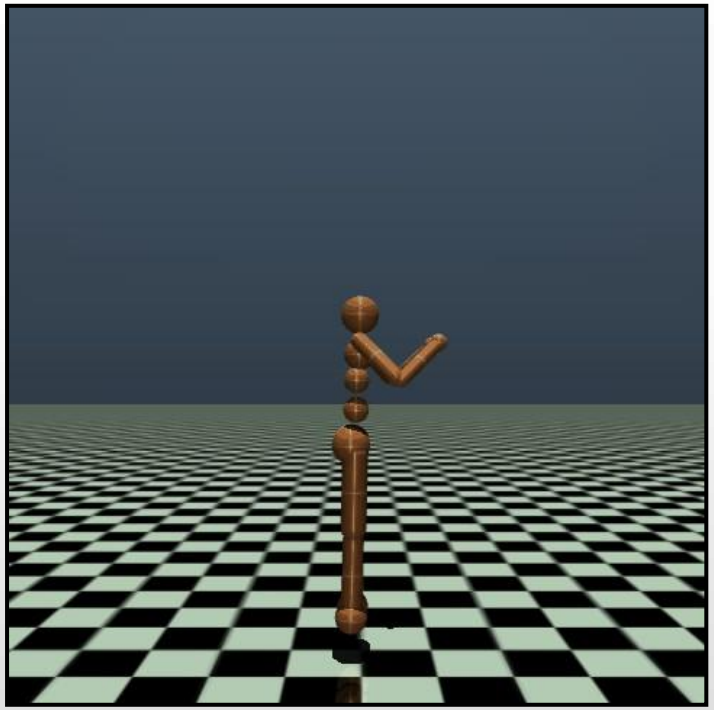


Assuming a specific *f(s)*, we calculate probabilistic bounds in tabular domains

- DeepSea
- Frozen Lake

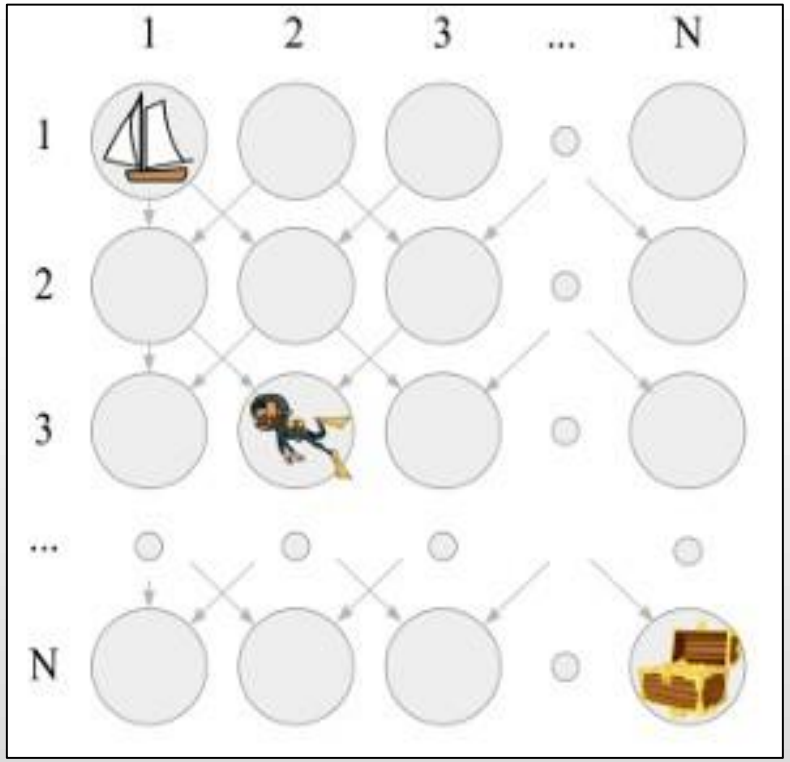
Experimental Evaluation: MuJoCo

- Demanding Control Tasks (MuJoCo)
- Test two effective ADEU instances
  - *f(s)*: the well-known RND mechanism used for Intrinsic Motivation
  - *f(s)*: the UCB criterion used to guide agents in highly uncertain states



Experimental Evaluation: Deep Sea

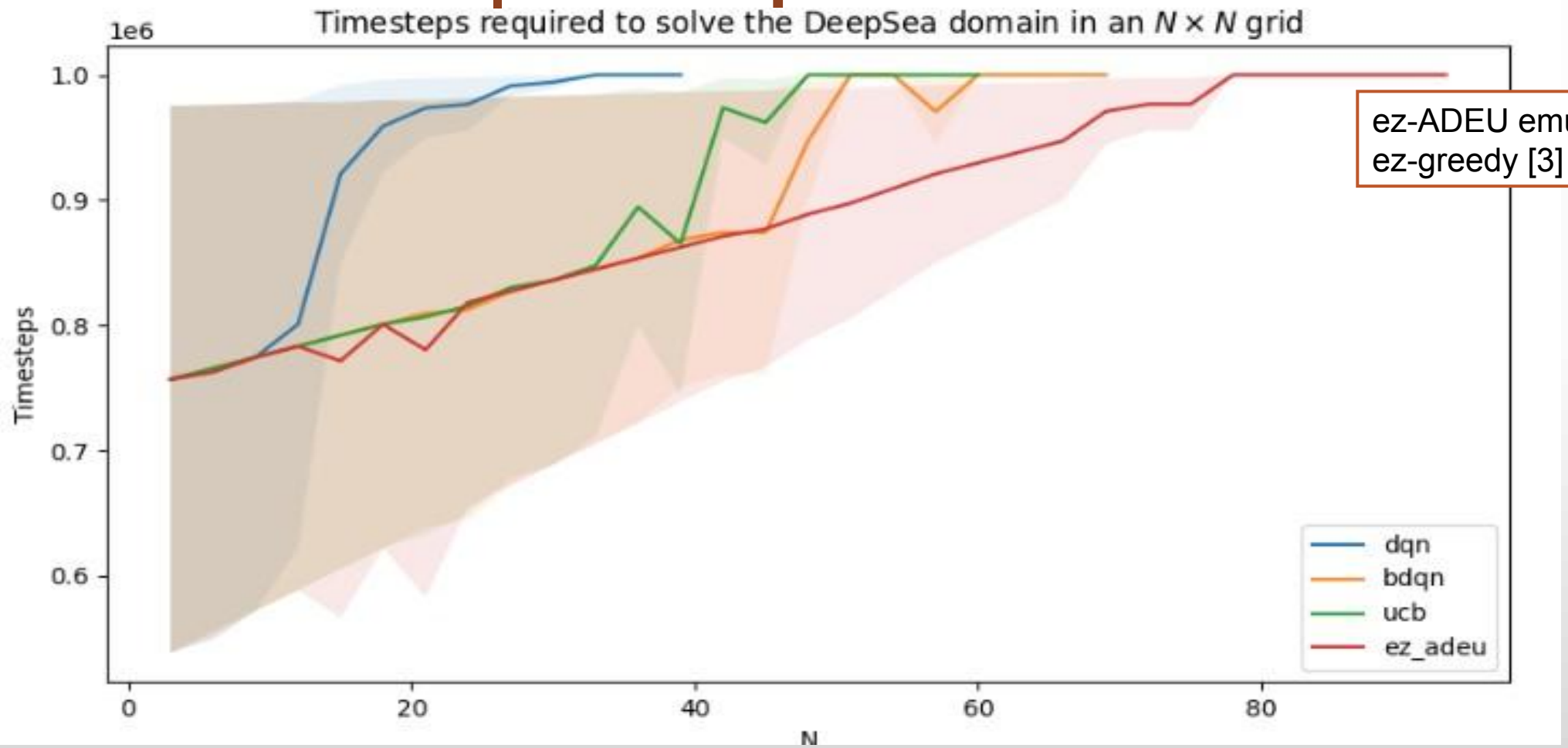
- The Deep Sea domain
  - Domain that requires Deep Exploration
    - Test one ADEU instance that emulates a previous adaptive exploration work (*ez-greedy*)



Results | MuJoCo

Domain /Method	ADEU + IM	ADEU + UCB	Naïve ADEU	UCB	IM	Noisy Nets
Walker2d	**745;947**	413;500		355;397	1;23	
Hopper	**733;1167**	293;514		245;422	10;35	
Swimmer	**52;62**	36;39		32;36	-3;-2	
Ant	**1072;1341**	**1134;1500**		585;790	-2658;980	
Humanoid	313;345	**512;658**		89;149	85;132	
Hum.Stand	**73895;79978**	**72978;80139**		54928;56527	55211;56057	

Results | Deep Sea

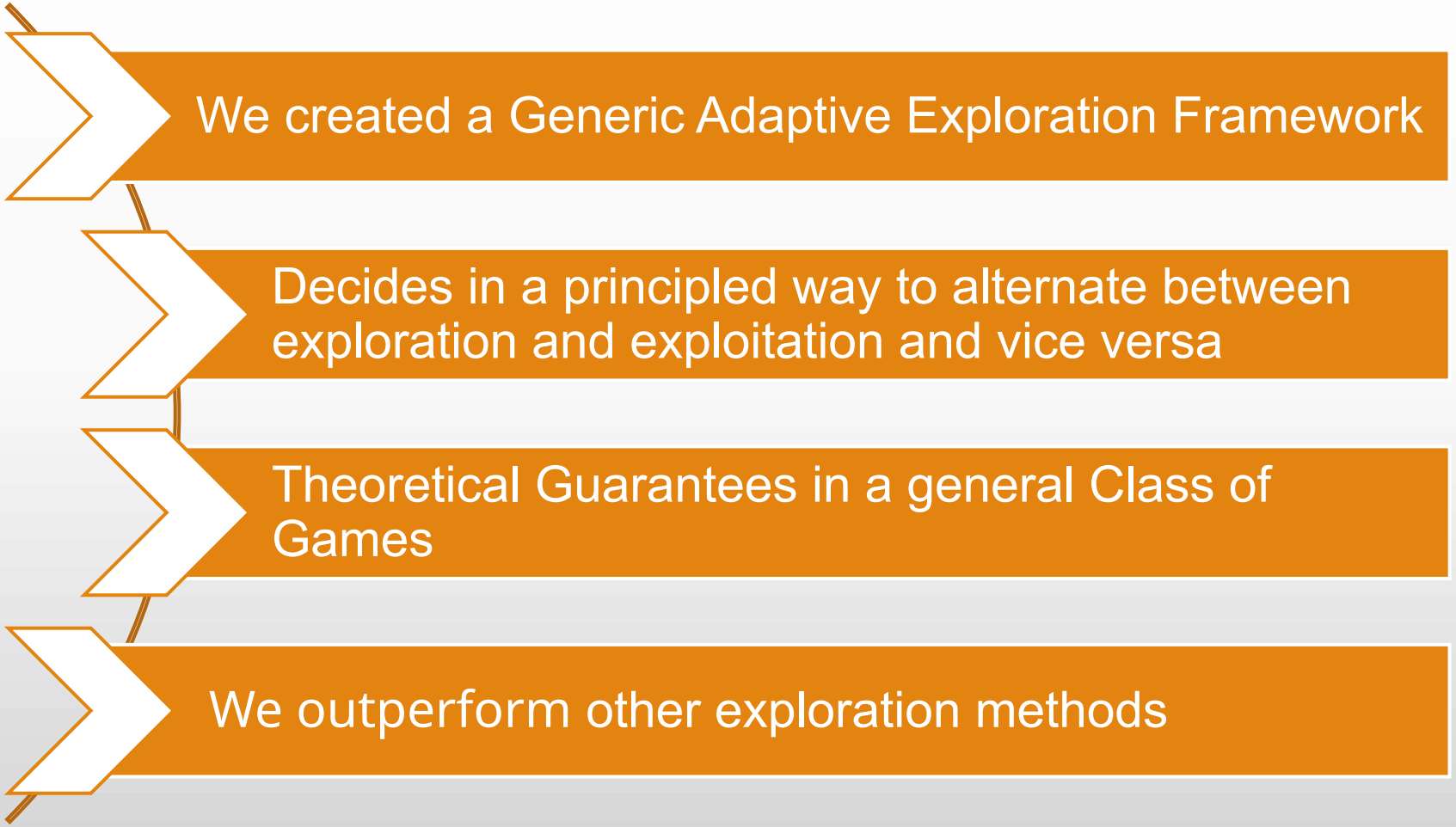


ez-ADEU emulates ez-greedy [3]

[3] Dabney et al. "Temporally-Extended ε-Greedy Exploration. " *International Conference on Learning Representations*.

Conclusions
and
Ongoing / Future Work

FSummary


We created a Generic Adaptive Exploration Framework
Decides in a principled way to alternate between exploration and exploitation and vice versa
Theoretical Guarantees in a general Class of Games
We outperform other exploration methods

Agent Self-Evolution…
ADEU is extending its own knowledge

Exploration is decided per state, not globally

Policies it is *confident* on are replicated; novel states are explored

Knowledge is extended only when required

…and now…Safety

ADEU for Safe Exploration

From uncertainty to safety

Building on $a \sim \mathcal{D}\left(\pi(s), g(f(s))\right)$

we can create safety aware instances, *by **pretraining*** π(s) on a **safe mentor**, and using the following *f(s)*:

1. **Standard Uncertainty?** a pretrained, uncertainty-driven *f(s);*
2. **Known cost function?** explore inversely to it; i.e., use its inverse as *f(s)*
3. **Unknown cost function?** learn it online; explore inversely to it

Safety-aware action selection

Unknown cost: we create two new "safety-aware" ADEU instances:

$$a \sim \mathcal{D}\left(\pi(s), g(1/f_{\text{safety}}(s))\right)$$

High expected cost → low variance → exploit π(s)

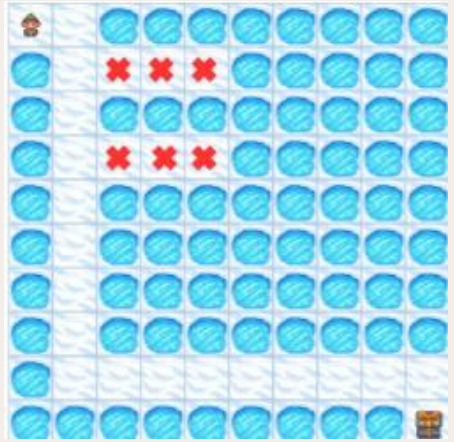
$$a \sim \mathcal{D}\left(\pi(s), g(f_{\text{uncertainty}}(s)/\lambda(f_{\text{safety}}(s))\right)$$ (Eq. 2)

Uncertainty-driven exploration when the expected cost is low; otherwise, safe (Eq. 3)

Eq 3: Best of both worlds: safe where it matters, uncertainty-driven exploration everywhere else.

Safe Exploration: Results

Safety Frozen Lake



ADEU instances, **pretrained** with a **safe policy** near unsafe states:

Eq. 1: Standard ADEU using f(s) = visitation frequency

Eq. 2, where $f_{safety}(s)$ is a learned cost

Eq. 3, where $f_{uncertainty}$ = visitation freq. and f_{safety} = learned cost

Method	Mean ; Max reward	Violations
ε-greedy	210 ; 739	0.34
IM	267 ; 780	0.55
UCB	12,605 ; 65,687	0.10
ADEU (Eq. 1, with pretraining)	59,865 ; 104,487	0.01
ADEU (Eq. 2)	7 ; 7	0.00
ADEU (Eq. 3)	**19,255 ; 85,087**	**0.00**

Eq. 1: ***highest reward (fastest to the goal)*** ~10× fewer violations than baselines

Eq. 2: ***perfectly safe*** - but too conservative.

Eq. 3: **Best of both worlds**: **perfectly safe** & **quite efficient**

Safe Exploration **is** Adaptive (ADEU) Exploration

- Safe exploration can be cast as ADEU exploration
- Bounds, guarantees, good results

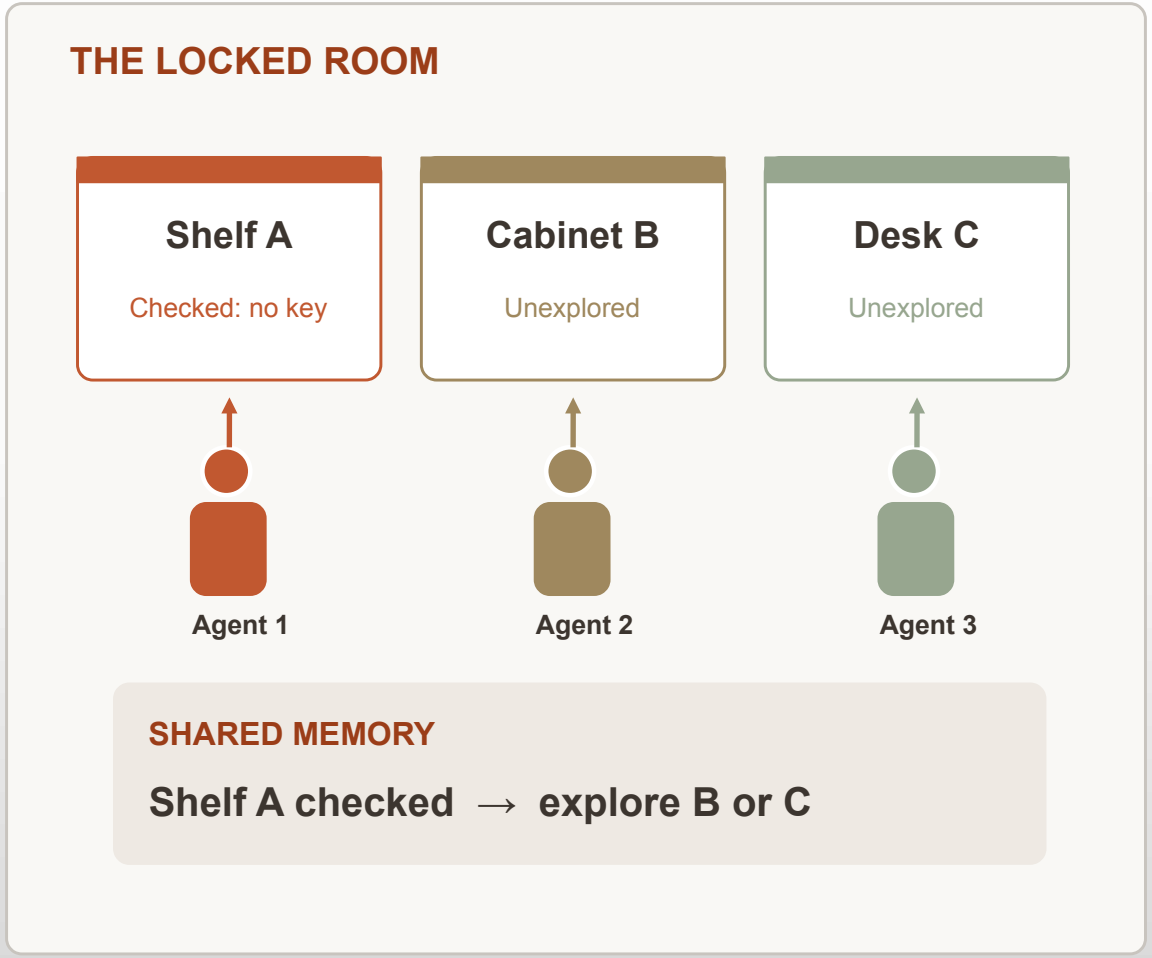
…moving to multiple agents…

Extension: Multiagent exploration

Example

- All agents are trapped in a locked room.
- Goal: at least one agent finds the key to unlock the door.
- The agents have a shared goal and shared memory.

Shared experience matters:
an unsuccessful search can **prevent** another agent from repeating the same action



ADEU in a Multiagent Setting

Vanilla multiagent exploration

Independent exploration can produce duplicated/ redundant experience.

Example: two agents search the same location for the key.

Proposed Multiagent ADEU

Agents share a **policy** and
a mechanism for estimating (**collective**) **uncertainty**.

Exploration reflects what the whole team knows.

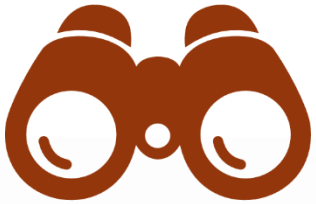
Individual uncertainty → Collective uncertainty → No redundant exploration

From Individual Experience to Organisational Knowledge

1	**Generate experience**	Agents act in the environment and record observations, including failed searches.
2	**Build shared memory**	Consolidate experiences into a common record of evidence (Replay Buffer).
3	**Estimate collective uncertainty**	Use shared knowledge to identify unresolved questions and gaps.
4	**Coordinate exploration**	Prioritise useful experiences rather than duplicating already reliable evidence.

Organisational knowledge: reusable evidence that improves the team's next decisions.

Future Directions



- Test background knowledge
- Escape suboptimal knowledge
- Create novel uncertainty measuring mechanisms
- Guide exploration via modifying higher moments of the distribution
- Interpret safe exploration as an adaptive one

Uncertainty-Driven Adaptive Exploration

Thank you!
Any Questions?

Download published proceedings paper version at:
https://www.ifaamas.org/Proceedings/aamas2026/pdfs/HQII4250.pdf

{lbakopoulos, gchalkiadakis}@tuc.gr

ADEU website

TECHNICAL UNIVERSITY OF CRETE

MAs-ML

2026 AAMAS PAPHOS • CYPRUS 25 YEARS