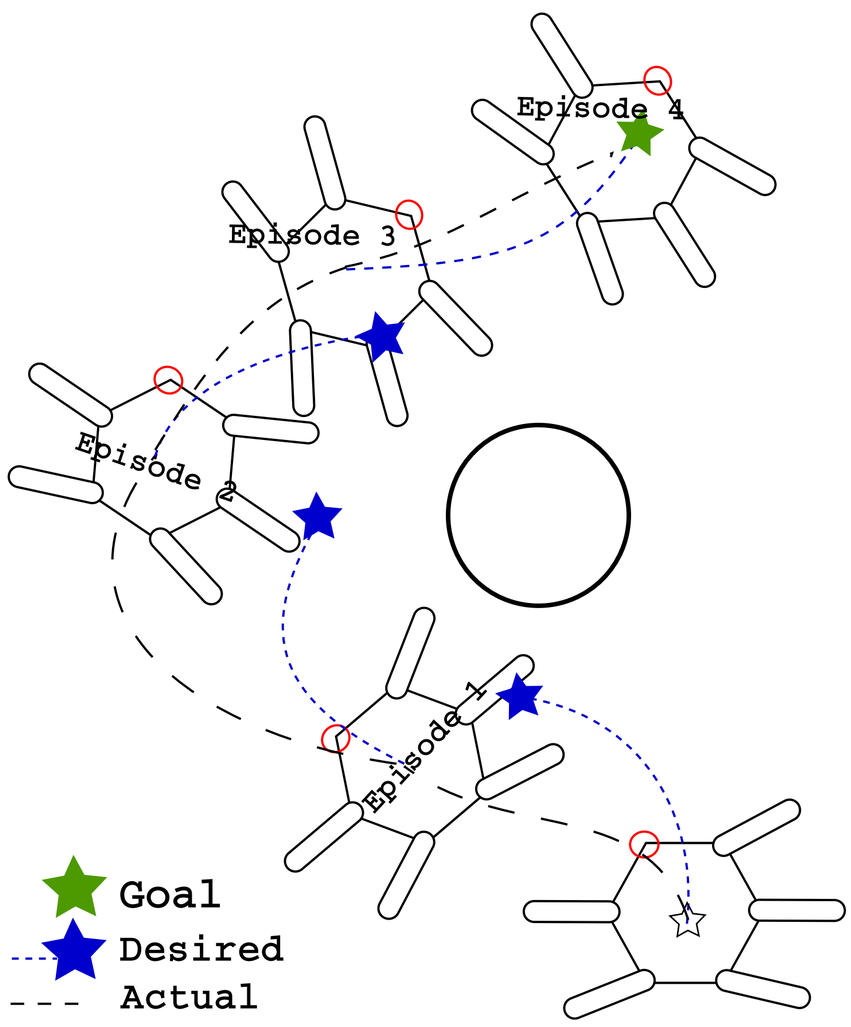
# Semi-episodic learning for robot damage recovery

■ "Towards semi-episodic learning for robot damage recovery." **AI for Long-Term Autonomy Workshop** - ICRA'16, 16 May 2016, Stockholm - Sweden

Konstantinos Chatzilygeroudis[1,2,3] • Antoine Cully[4] • Jean-Baptiste Mouret[1,2,3]
konstantinos.chatzilygeroudis@inria.fr / a.cully@imperial.ac.uk / jean-baptiste.mouret@inria.fr

- A major obstacle to the widespread adoption of robots is their fragility.
- A promising approach involves having robots learn appropriate behaviors in response to damage, but the learning process has to be fast and creative enough to be used in practice.
- The recently introduced **Intelligent Trial and Error algorithm (IT&E)**, by *Cully et al.* [1], allows autonomous robots to adapt to damage in a matter of minutes.
- **We extend the IT&E algorithm to allow for robots to learn to compensate for damages while executing their task(s).**
- Two steps: (1) Map creation (via MAP-Elites); (2) adaptation (via a variant of Map-based Bayesian Optimization)



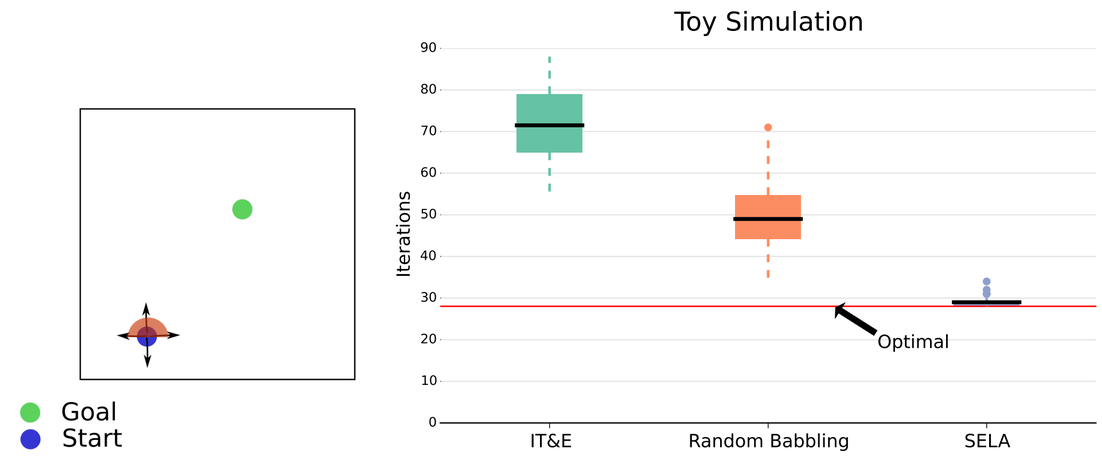

### Creating the behavior-performance map with MAP-Elites

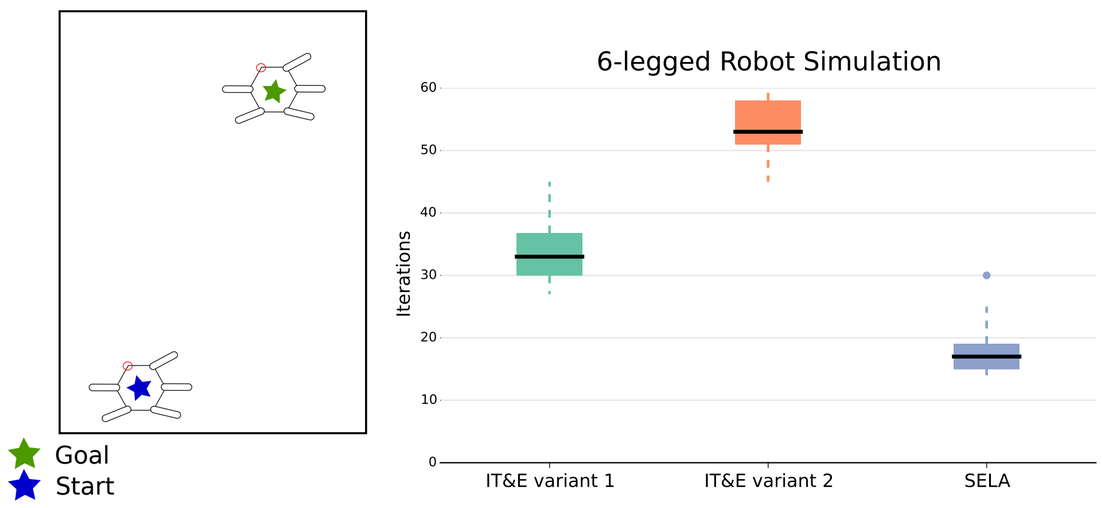
- In simulation, a "behavior-performance" map of the potential performance of each location in a user-defined (behavior) space is generated.
- In our hexapod experiments, the behavior space is 8D:
  - Six dimensions for walking diversity: duty factor.
  - Two dimensions for space diversity: (x,y) coordinates of end point.

### Adaptation Step with Bayesian optimization

- We use Gaussian Processes to learn a mapping from the atomic behaviors to the resulting relative outcomes (Generic Reward).
- After damage, the robot selects a promising behavior (according to a specific "Reward" function), executes it and updates the GPs.
- A "Specialized Reward Selection Layer" is responsible for selecting or updating the "Reward" function at each episode.
- This select/execute/update loop is repeated until the tasks of the robot are completed.

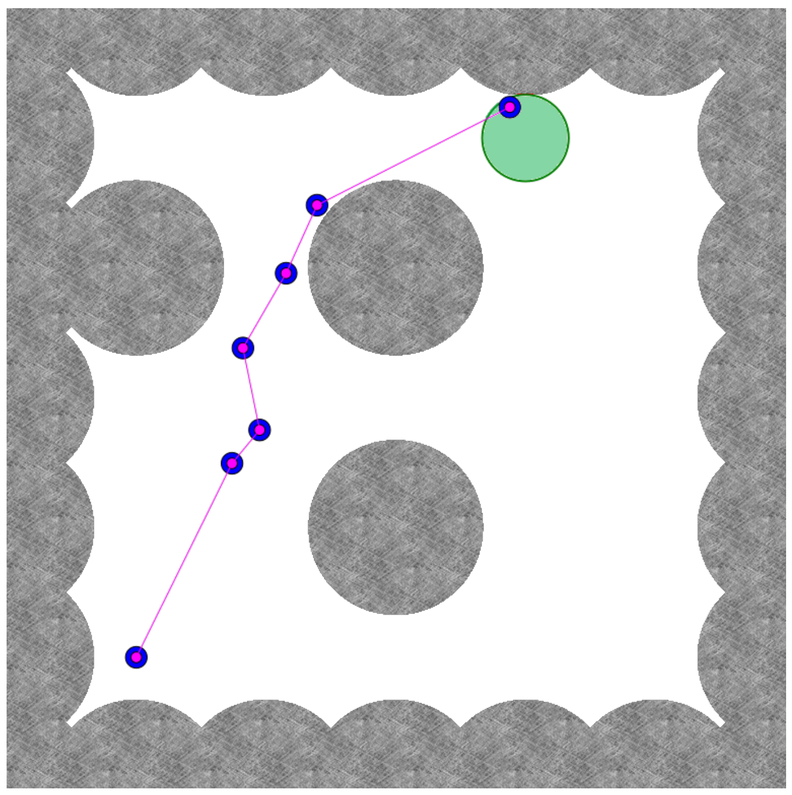
[1] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," Nature, vol. 521, no. 7553, pp. 503–507, 2015.




### Preliminary Results

- One toy locomotion task, in simulation, without MAP-Elites priors and one hexapod locomotion task, in simulation and real-robot, with 8D MAP-Elites archives (1 million behaviors).
- Successfully compensated for damages while completing the tasks.
- Outperforms "out of the box" IT&E and "random babbling".



### Work in Progress: Probabilistic Planning with Monte Carlo Tree Search

- "Specialized Reward Selection Layer" can be a planner.
- Take into account the uncertainty of each atomic behavior's outcome - "Monte Carlo Tree Search" (MCTS).
- First experiments yield promising results.

[1] Inria, Villers-lès-Nancy, F-54600, France
[2] CNRS, Loria, UMR 7503, Vandœuvre-lès-Nancy, F-54500, France
[3] Université de Lorraine, Loria, UMR 7503, Vandœuvre-lès-Nancy, F-54500, France
[4] Imperial College, London, UK

**Funding:** This work was supported by the ERC project "ResiBots" (grant agreement No 637972), funded by the European Research Council.