FraMoTEC:
Modular Task-Environment Construction Framework for Evaluating Adaptive Control Systems

Thröstur Thorarensen[1], Kristinn R. Thórisson[1,2], Jordi Bieger[1] and Jóna Sigurðardóttir[2]
[1] School of Computer Science | Center for Analysis and Design of Intelligent Agents
[2] Icelandic Institute for Intelligent Machines

Overview

- Motivation
- Task Theory
- FraMoTEC
- Examples
- Conclusion

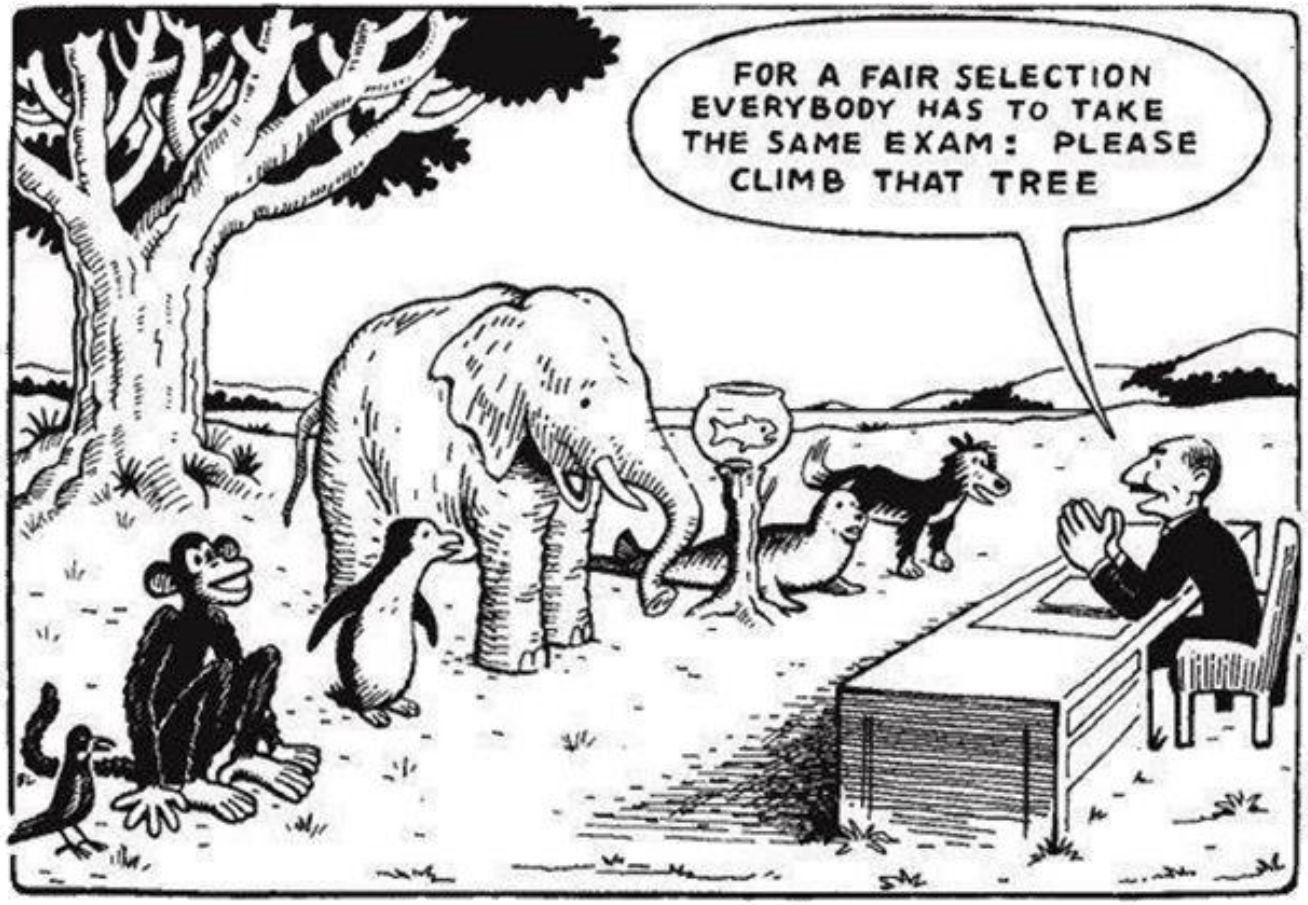
Long Term Goal

- Evaluation of general-purpose AI:
  - ***Predicting general cognitive abilities of adaptive control systems now and in the future.***

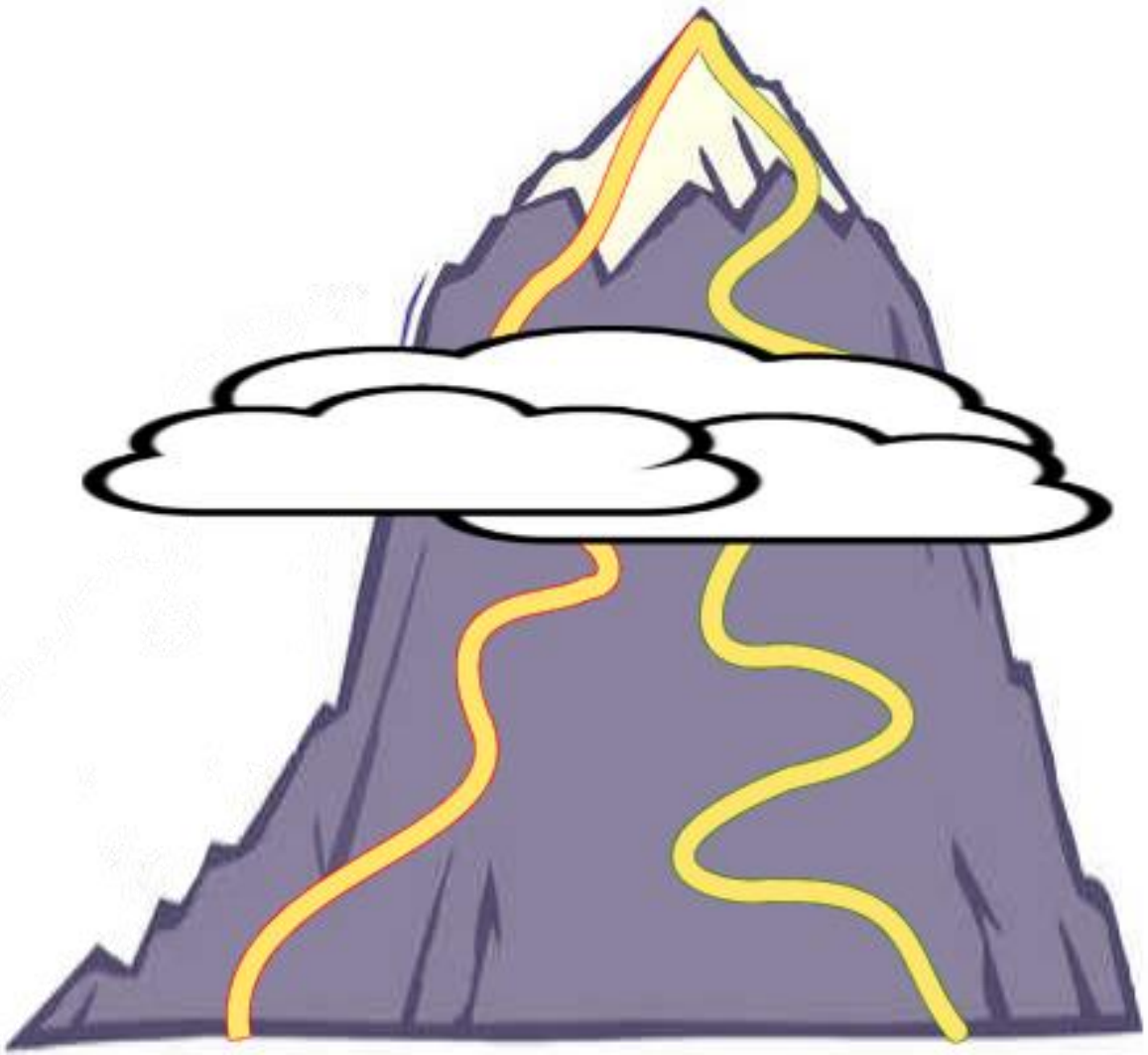
Motivation

- Behavioral testing is done through observing a system's performance in **task-environments**.
  - Solution: get interesting task-environment
- **Generality** cannot be evaluated in a single task-environment (or a small set).
  - Solution: get range of task-environments that require desired ability
- We want to know more than single-point-of-time performance: we want to make predictions about **future behavior**.
  - Solution: perturb and scale task-environments
- Ability to deal with **novel situations** cannot be evaluated using a fixed, known test battery.
  - Solution: continually build new task-environments
- No AGI exists yet and **different approaches** have different strengths and weaknesses and attain different kinds of cognitive abilities first.

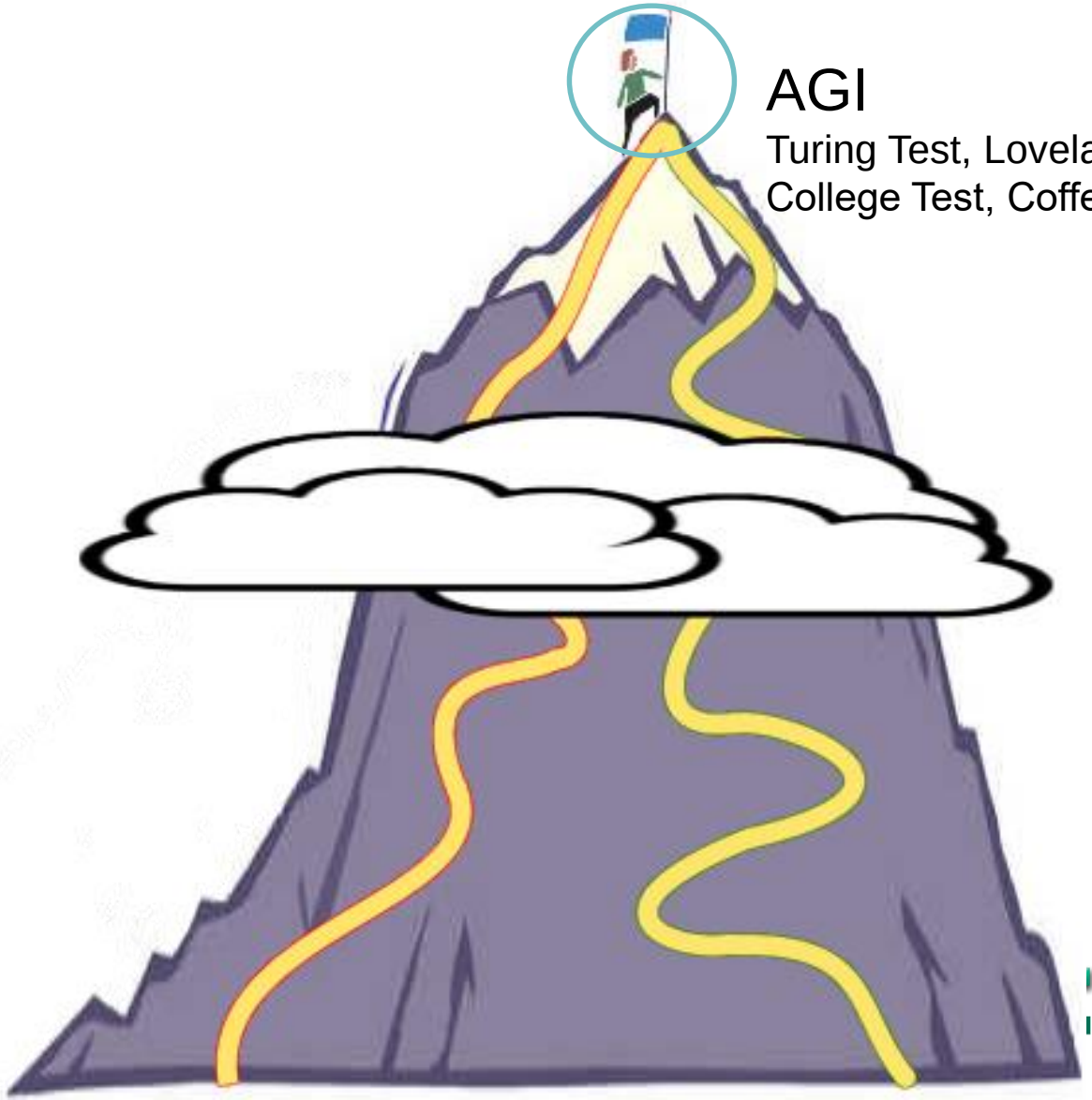
Motivation – No AGI Yet



Motivation – Multiple Paths to AGI



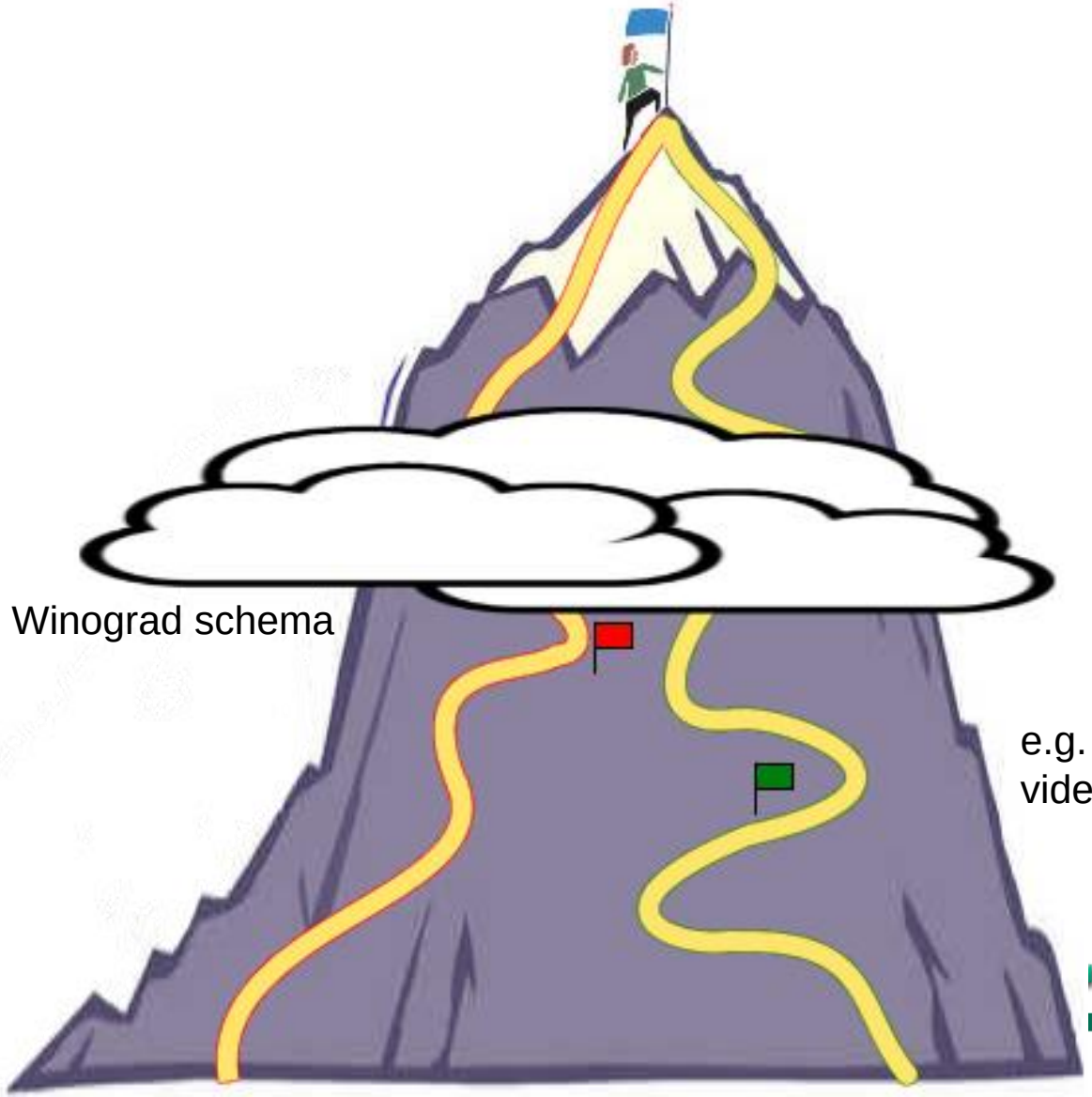
Motivation – Multiple Paths to AGI



AGI

Turing Test, Lovelace Test, College Test, Coffee Test…

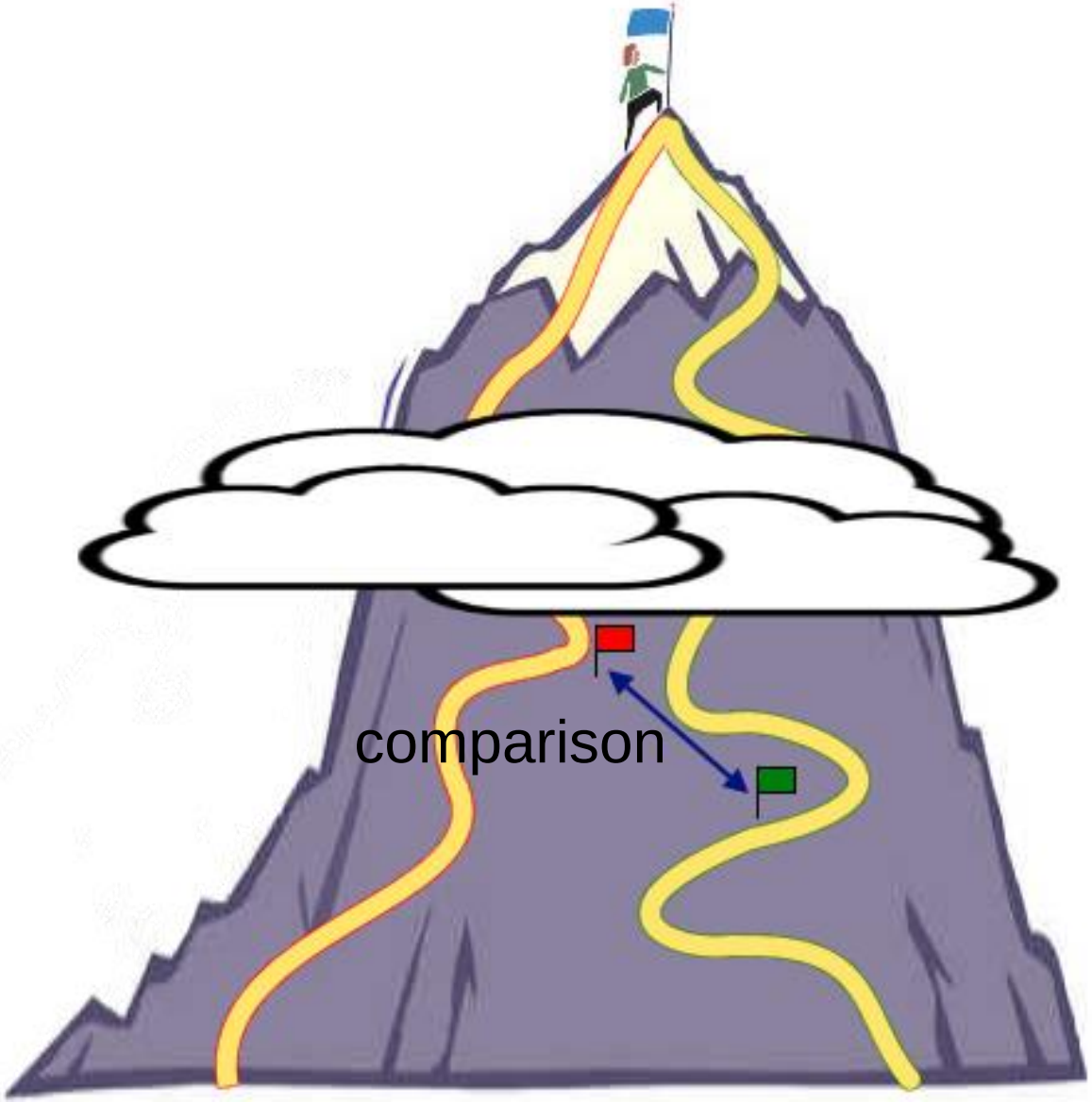
Motivation – Multiple Paths to AGI



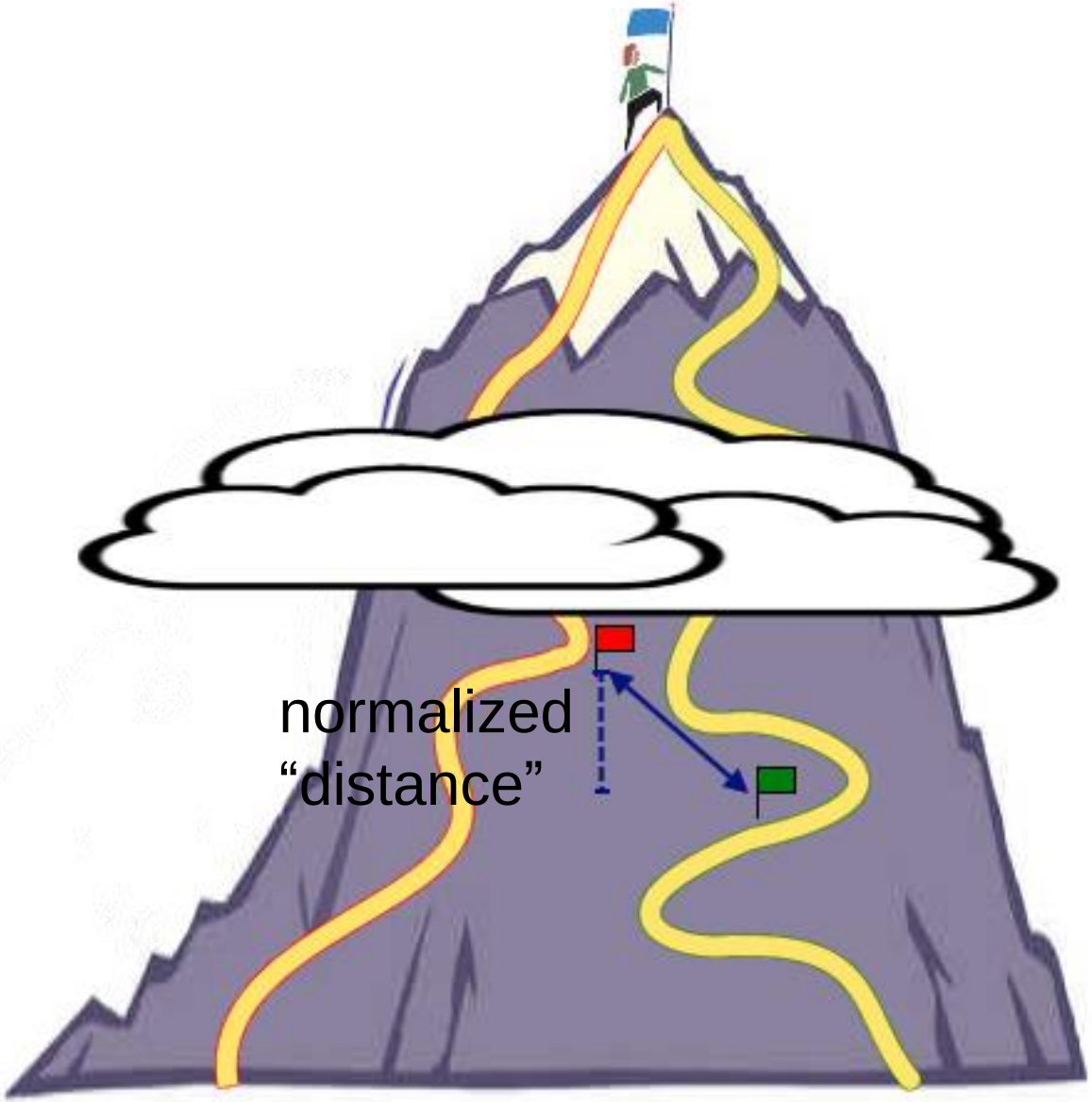
e.g. Winograd schema

e.g. simple video games

Motivation – Multiple Paths to AGI



Motivation – Multiple Paths to AGI



Motivation

- Behavioral testing is done through observing a system's performance in **task-environments**.
- **Generality** cannot be evaluated in a single task-environment (or a small set).
- Ability to deal with **novel situations** cannot be evaluated using a fixed, known test battery.
- We want to know more than single-point-of-time performance: we want to make predictions about **future behavior**.
- No AGI exists yet and **different approaches** have different strengths and weaknesses and attain different kinds of cognitive abilities first.

- Solution:
  - Build comparable task-environments with characteristics that can be varied depending on the system and test goal.

Requirements for AGI Evaluation Framework

- Easy construction of task-environments and variants:
  - (de)composition
  - scaling
  - tuning
- Automated generation of task-environments with specific features, constraints, etc.
- Analysis of task-environments in terms of parameters of interest

Thórisson, K.R., Bieger, J., Schiffel, S. and Garrett, D., 2015. Towards Flexible Task Environments for Comprehensive Evaluation of Artificial Intelligent Systems and Automatic Learners.

Task Theory

- Comparison of similar and dissimilar tasks
- Abstraction and concretization of (composite) tasks and task elements.
- Estimation of time, energy, cost of errors, and other resource requirements (and yields) for task completion.
- Characterization of task complexity in terms of (emergent) quantitative measures like observability, feedback latency, form and nature of information/instruction provided to a performer, etc.
- Decomposition of tasks into subtasks and their atomic elements.
- Construction of new tasks based on combination, variation and specifications.

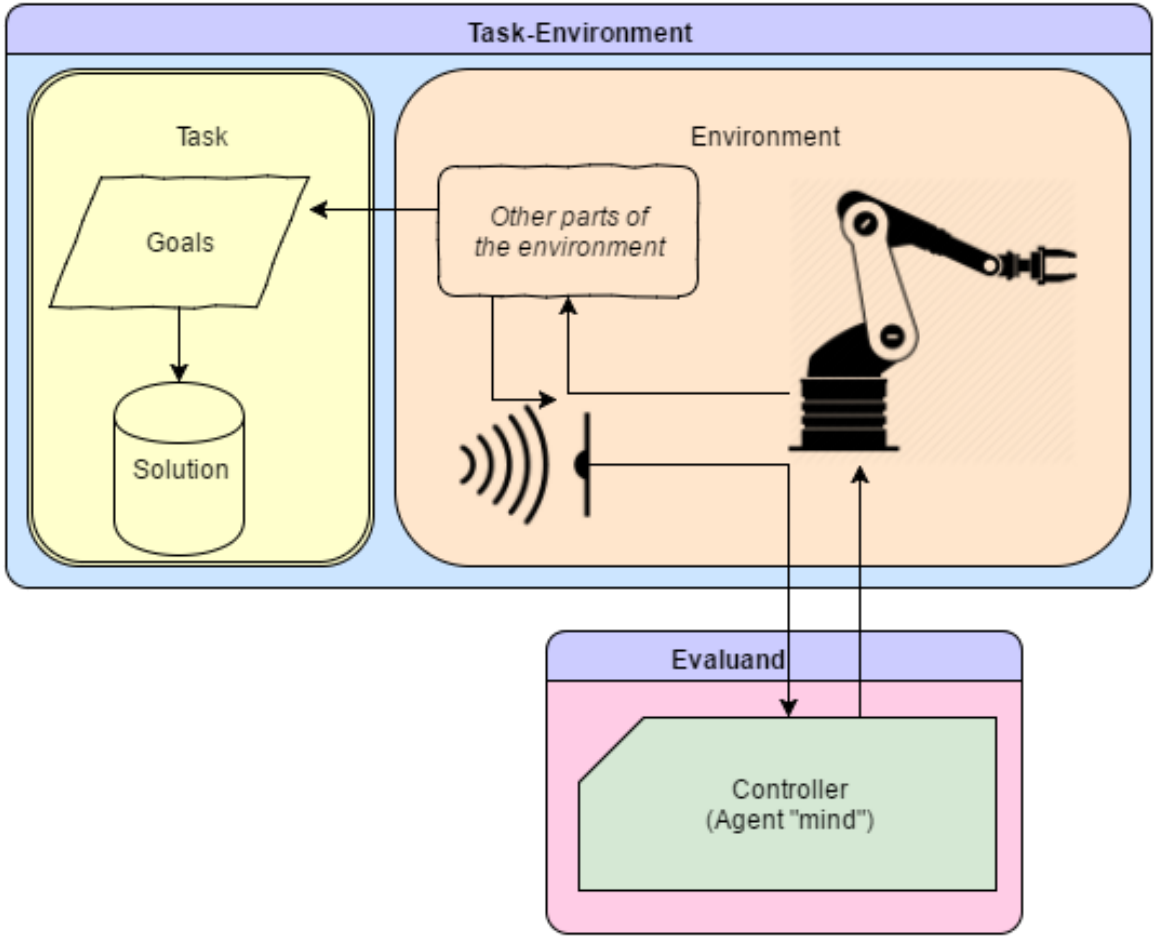
Thórisson, K.R., Bieger, J., Thorarensen, T., Sigurðardóttir, J.S. and Steunebrink, B.R., 2016. Why Artificial Intelligence Needs a Task Theory---And What It Might Look Like.

FraMoTEC's Task Theory

- Grounded in physics
- Closed problems
- Estimation of difficulty based on required time and energy
- Estimation of learning complexity

FraMoTEC

Framework
for
Modular
Task-
Environment
Construction.



https://github.com/ThrosturX/task-env-model

Building blocks

- Objects
- Transitions
- Sensors
- Actuators
- Systems (abstractions)
- Goals

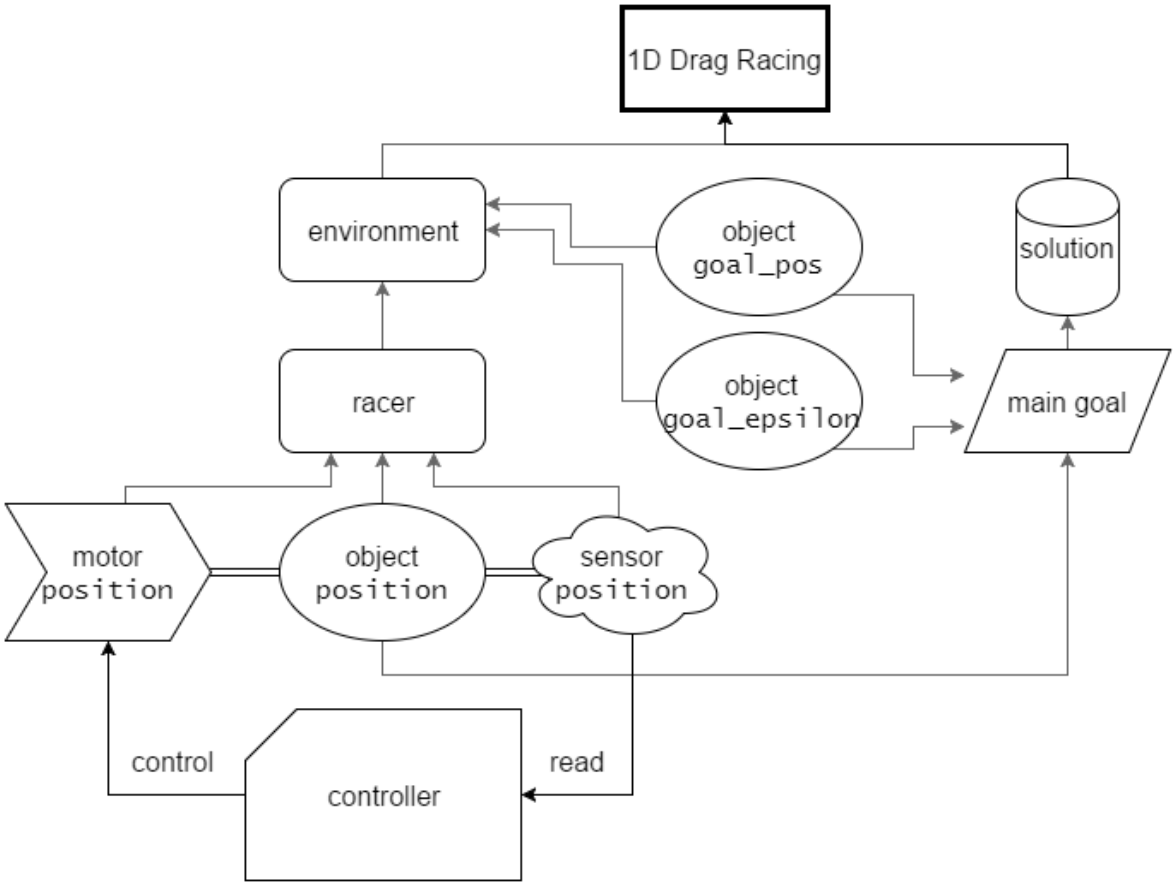
Environment

- Objects
  - Affected by transitions
  - Sensors enable inspection
- Transitions
  - Built-in
    - Activated motors
    - Gravity, friction
  - Custom

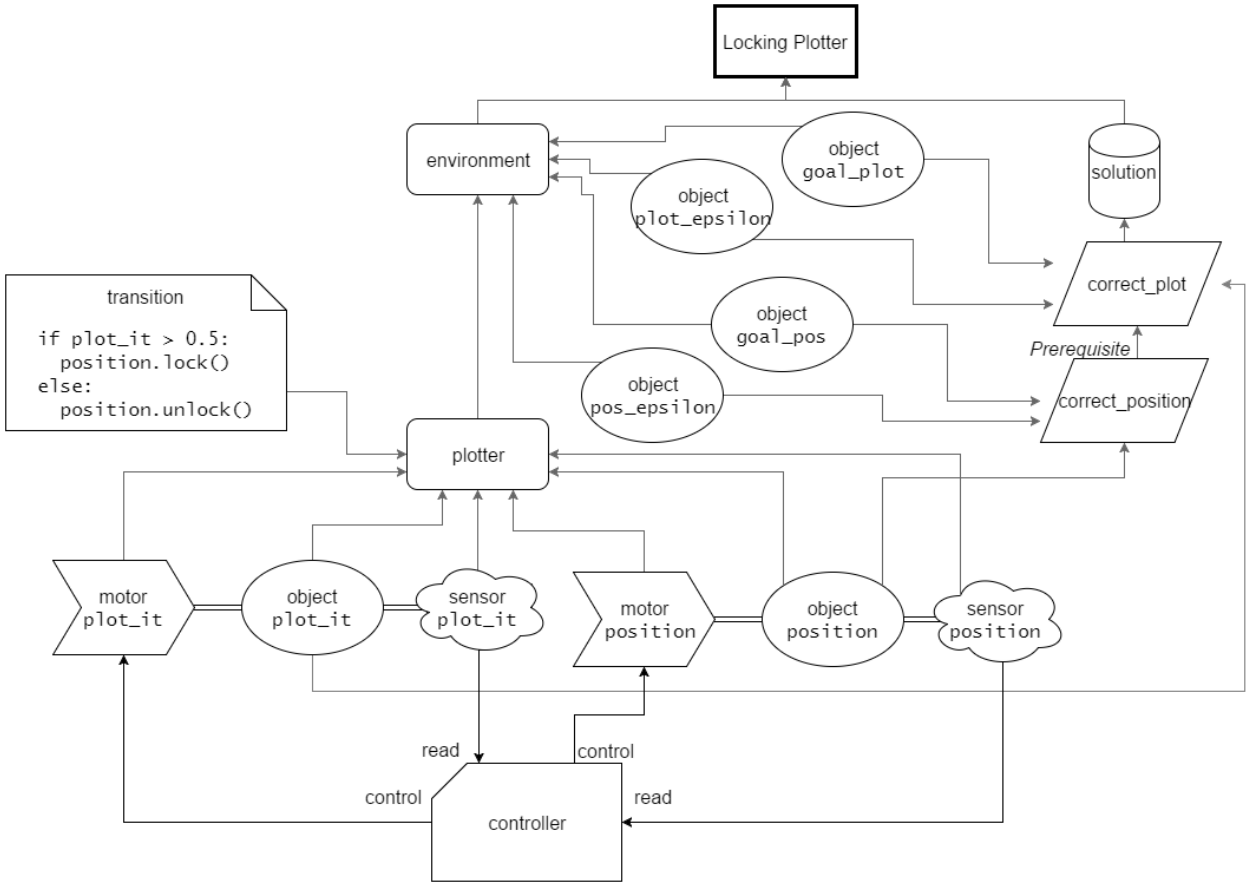
Task

- Problems
- Goal states
- Goals can depend on other goals
- Time and energy limitations

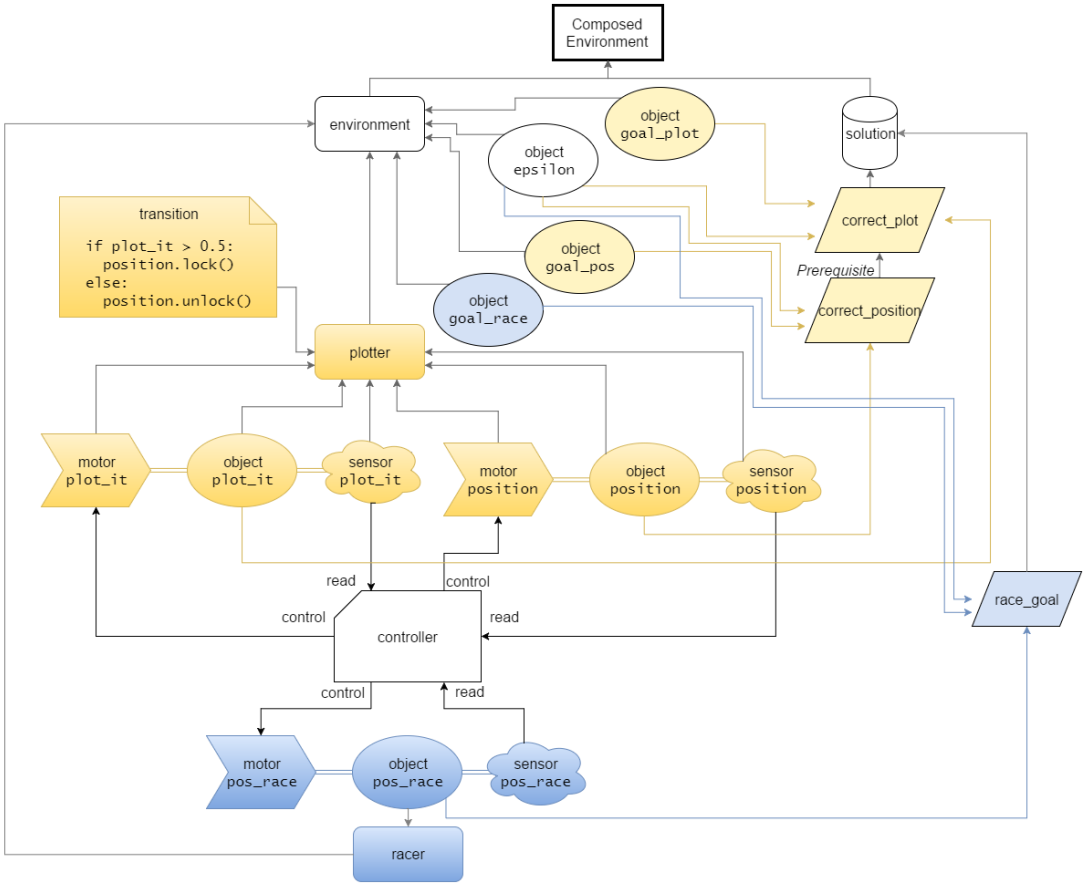
Example 1: Race



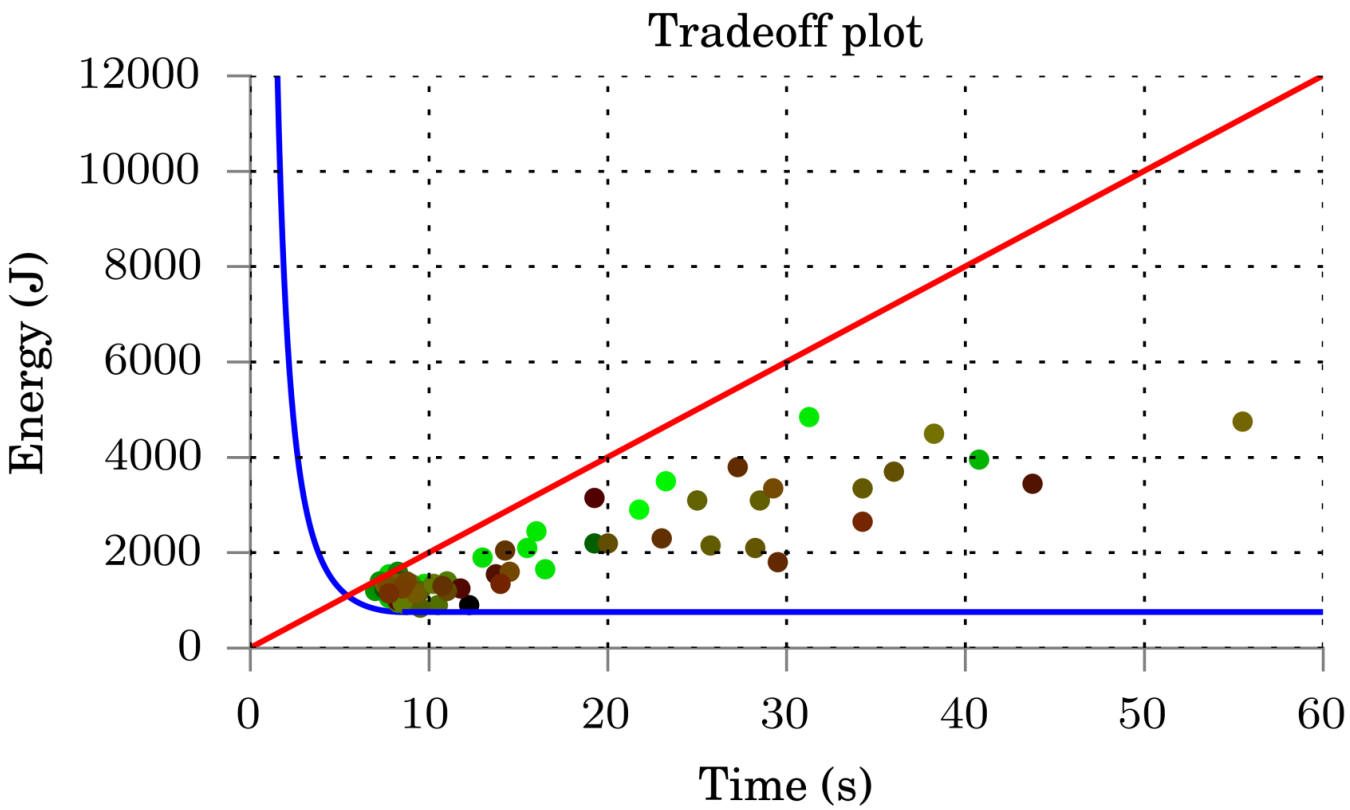
Example: plotter



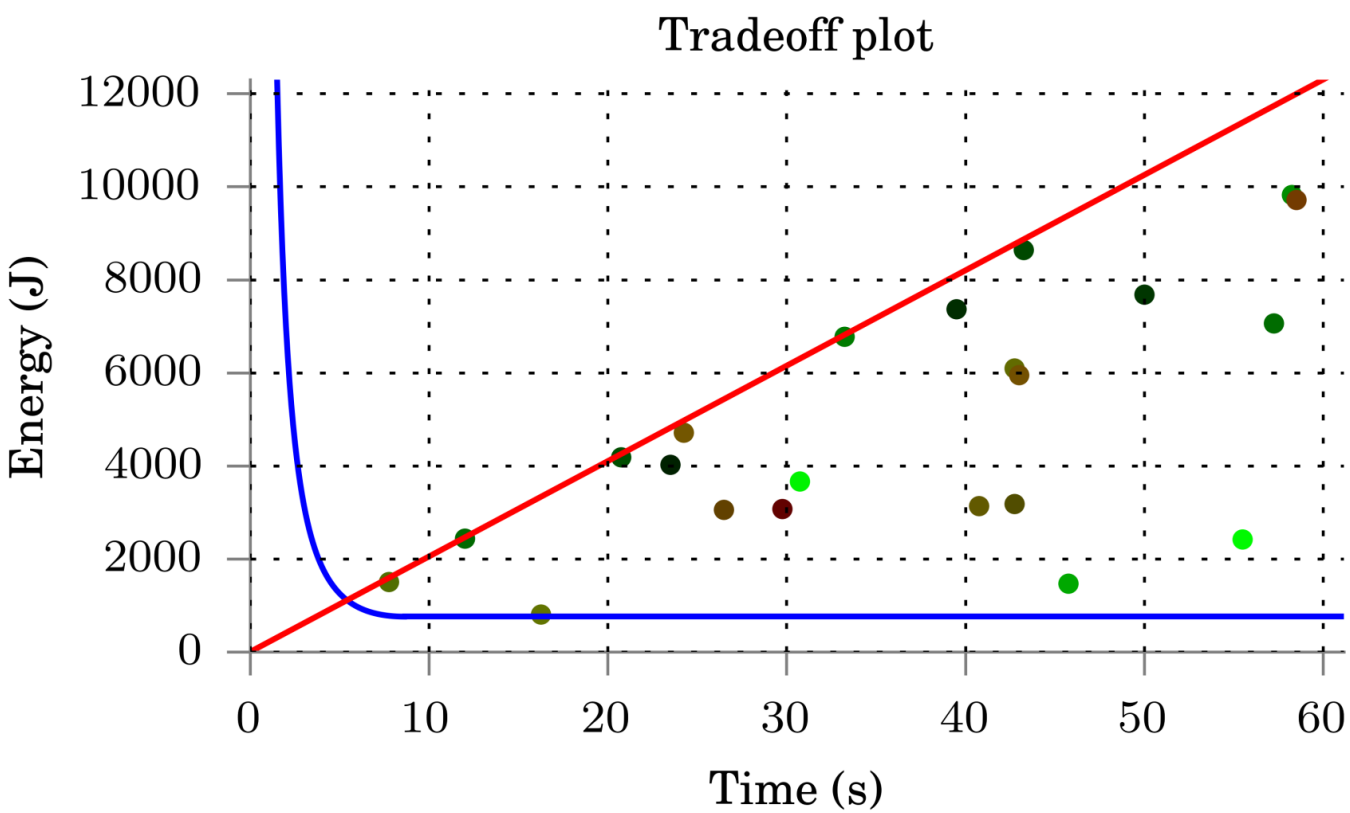
Example: combined



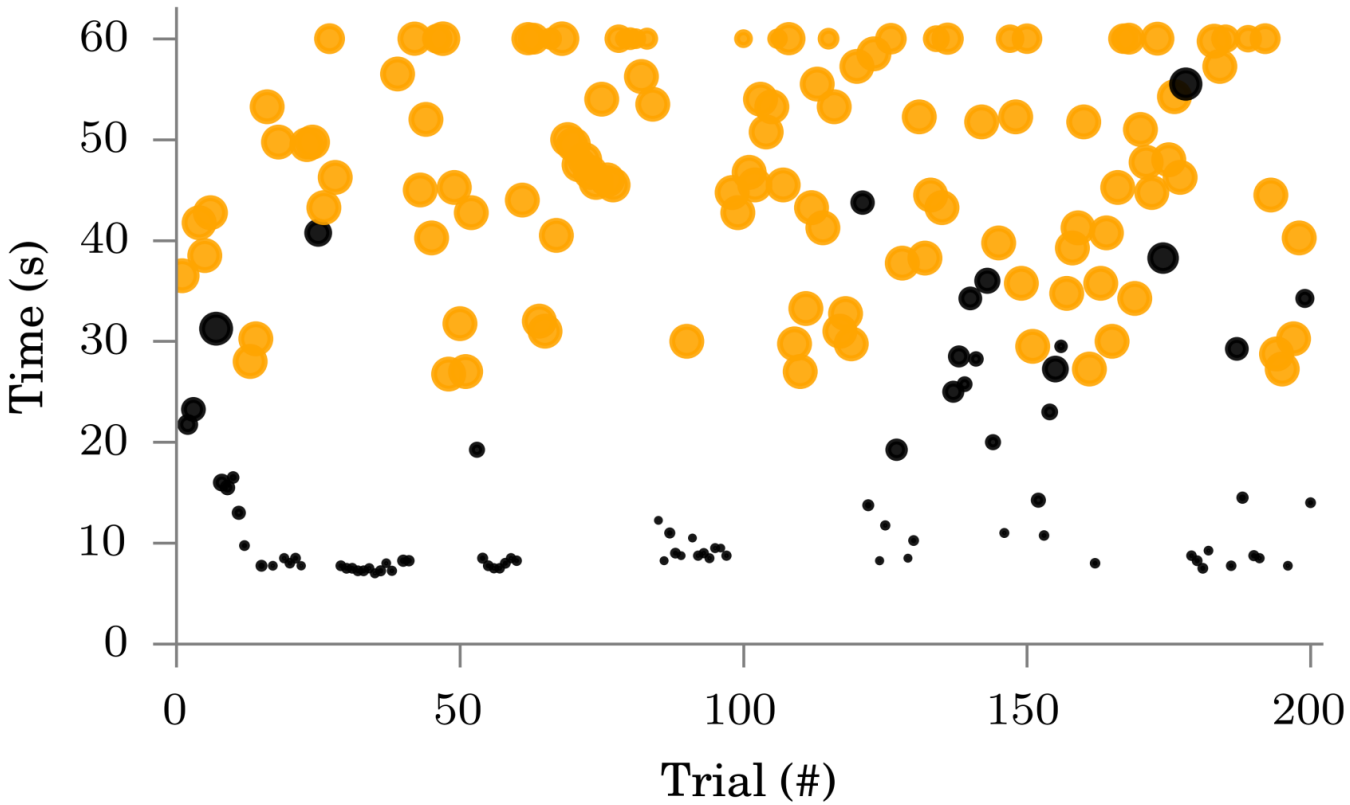
Sarsa Race Results



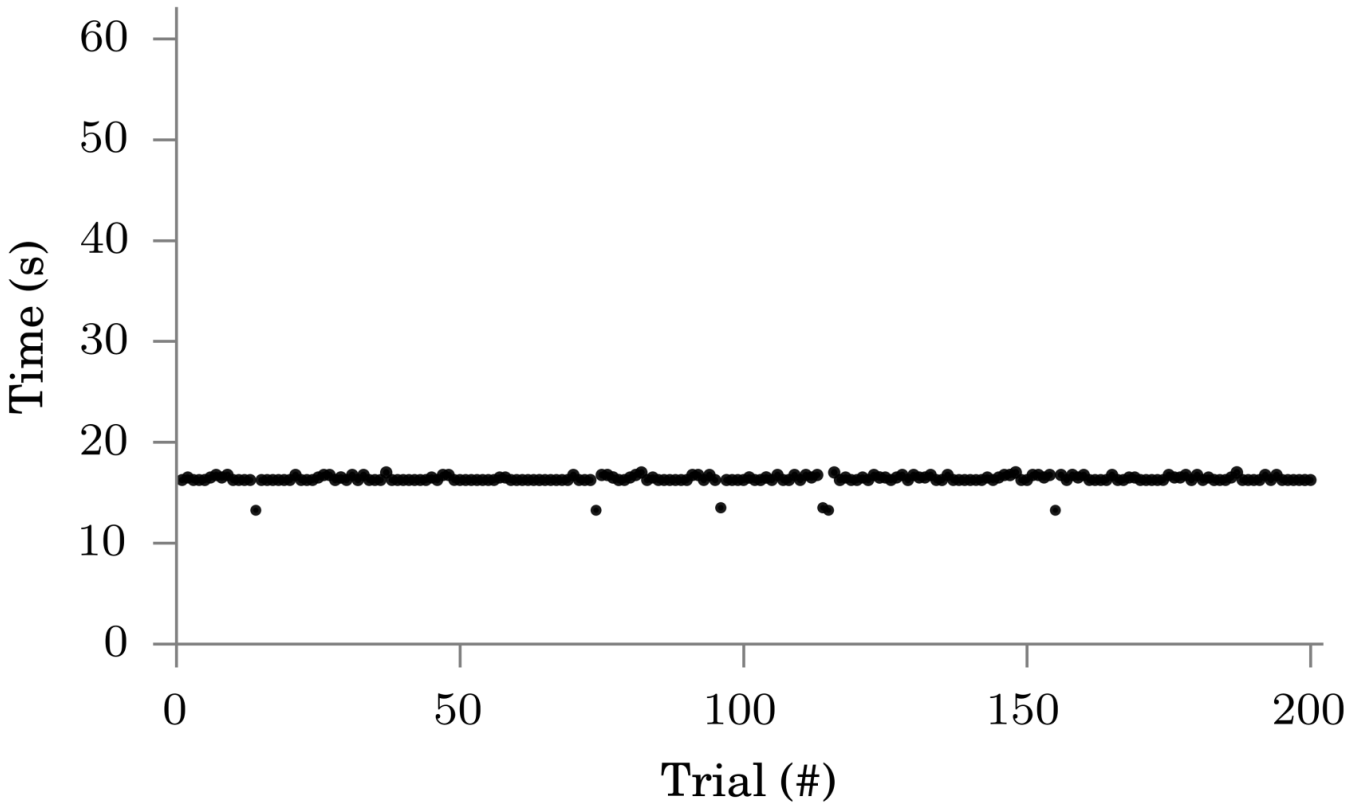
Sarsa Plotting Results



Sarsa Race Results



Specialized Plotting Results



Conclusion

- Evaluation of general-purpose AI requires building large numbers of appropriate task-environments.
- For this we need a task theory.
- FraMoTEC is our first implementation of some of these ideas.
- Much future work is needed on task theory and the construction of tasks according to specification.

https://github.com/ThrosturX/task-env-model

Questions?

- Thröstur Thorarensen <throstur11@ru.is>
- **Kristinn R. Thórisson <thorisson@ru.is>**
- **Jordi Bieger <jordi13@ru.is>**
- Jóna Sigurðardóttir <jona@iiim.is>

https://github.com/ThrosturX/task-env-model