how failure facilitates success

•••

Mark Wernsdorfer

August 23, 2018

proceeding

1. greater picture
2. problem setting
3. mental modelling
4. semiotic models

greater picture

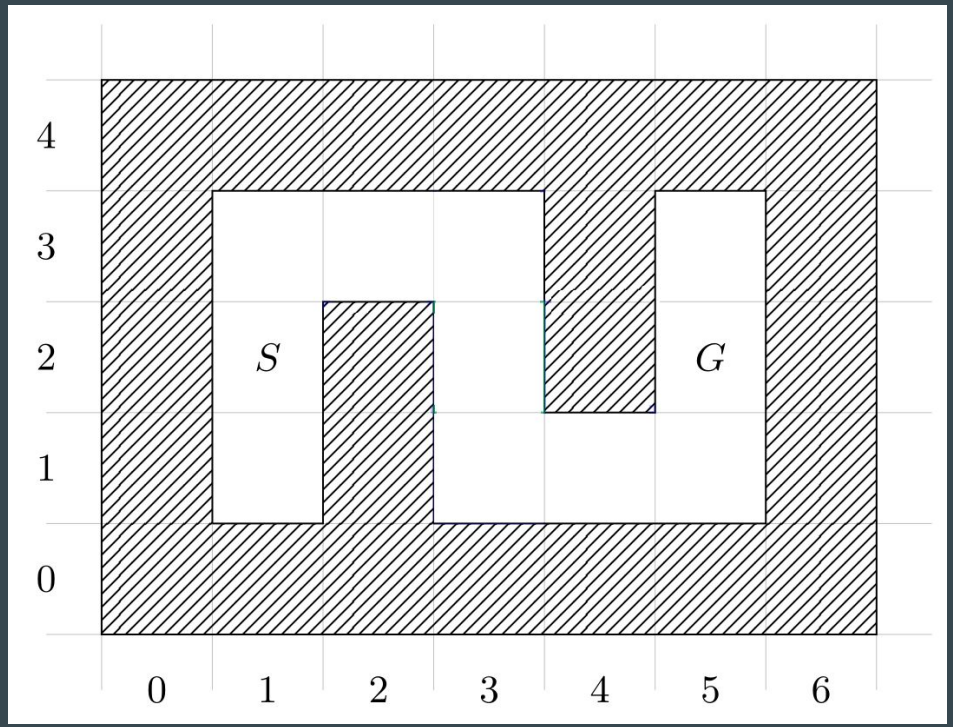
this paper is one of the results of a bigger project to simulate the generation of a mental model from the cognitive system's own perspective

consider *the process of mental modelling* under phenomenological aspects: how is the conversion of immediate feelings to abstract knowledge *experienced* by the system itself?

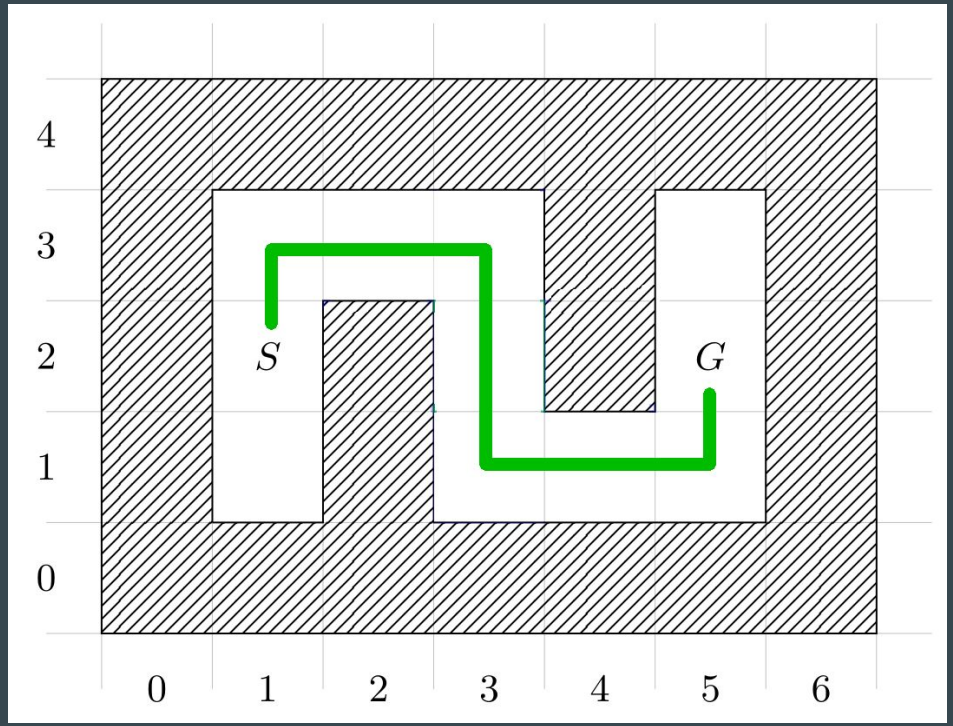
one insight was that a cognitive system's failures to predict *concrete* aspects of its environment (“breakdown”) enable it to describe *more abstract* aspects

the value of this hypothesis can be shown in a grid world environment

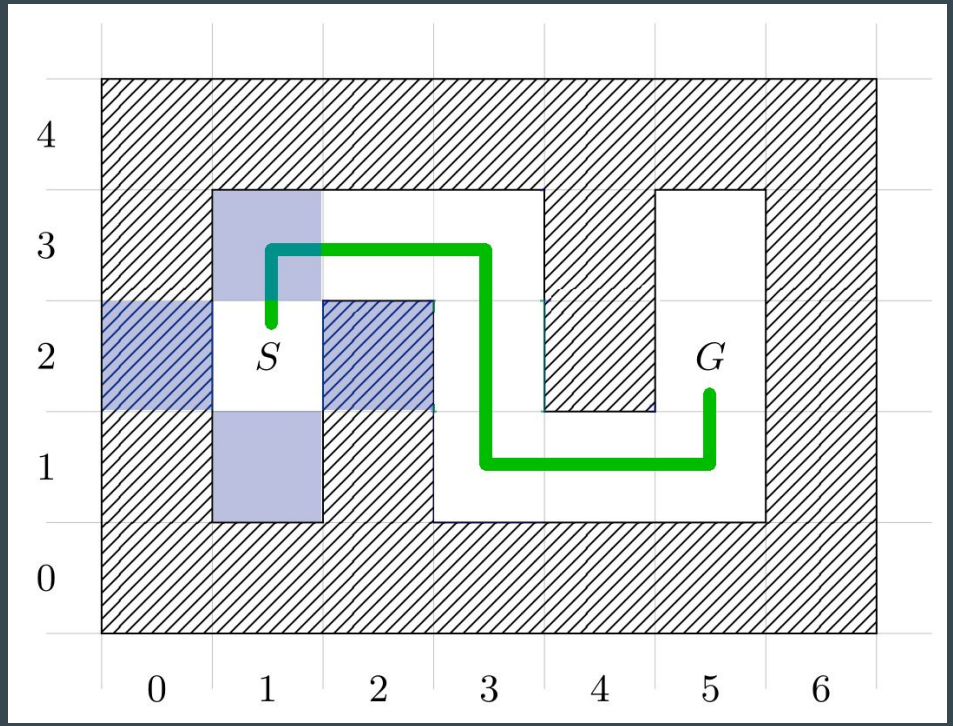
problem setting



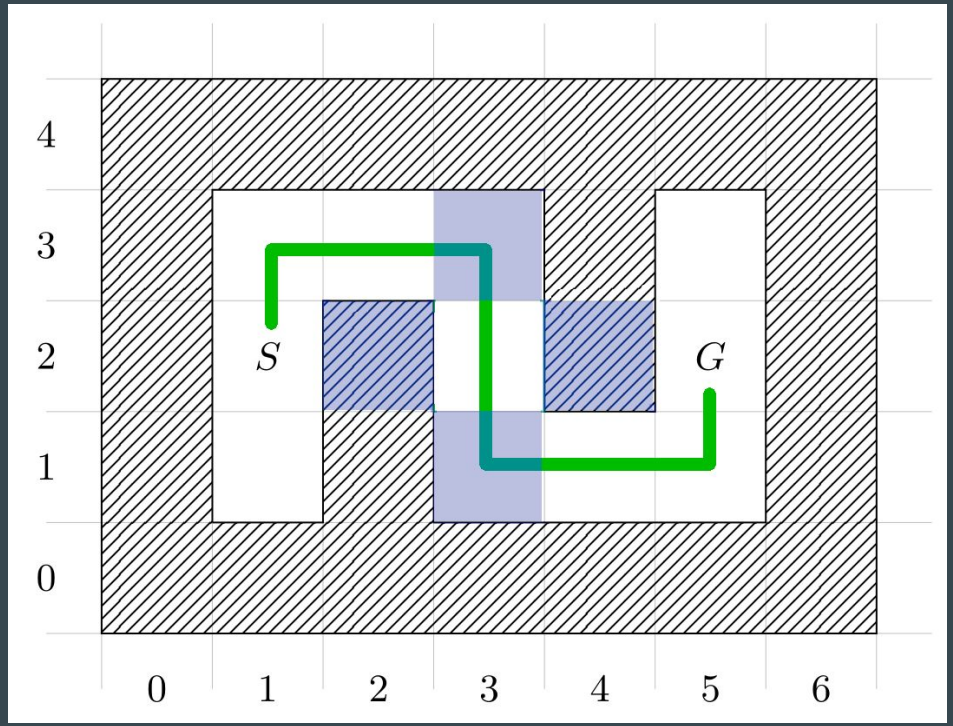
problem setting



problem setting (same perception, different action required)



problem setting (same perception, different action required)



problem setting

the agent can generate a model that allows the disambiguation of identical perceptions

the descriptive quality of the generated model can be quantified by its success in predicting the next perception given the current perception and action

the practical value of the model beyond a one-step-prediction can be quantified as *averaged reward accumulated over time*

the cognitive plausibility of the generated model can be compared against the default model in cases of partial observability: POMDPs

mental modeling

POMDPs are not ideal to simulate a subjective mental model because...

(1) they are too general to be generated on-the-fly

leads to: alternating train and test periods that *disconnect* learning from the present environment (e.g. experience replay)

practical problem: offline training on *past* data might always cause an adaptation to *past* circumstances

alternative: acquire and apply knowledge *in parallel* (i.e. during the same time step)

mental modeling

POMDPs are not ideal to simulate a subjective mental model because...

(1) they are too general to be generated on-the-fly

(2) they describe "naturalist" environments

> **leads to:** the procedure needs to find states and rules that *determine* transitions and observations
>
> **practical problem:** to determine the complete state and behavior of a system is almost impossible and probably unnecessary for any single task
>
> **alternative:** describe the environment as *indeterminate* or *erratic*

mental modeling

POMDPs are not ideal to simulate a subjective mental model because...

(1) they are too general to be generated on-the-fly

(2) they describe "naturalist" environments

(3) incomplete knowledge is usually baked into Markov models as probability

leads to: only the most probable option provides a useful prediction

practical problem: systematic prediction errors are not exploited

alternative: use the specific error to predict a more appropriate situation

semiotic model

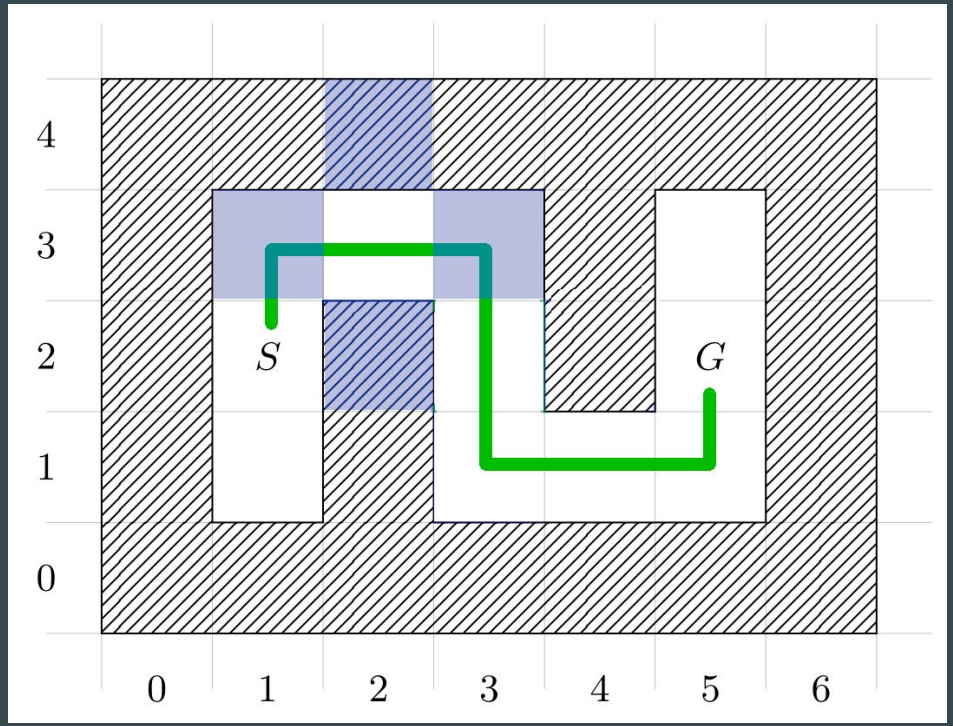
situations are encoded as Markov predictors that keep count of transition frequencies from the current perception and action to the immediately following perception

a prediction error indicates that, *from now on*, the situation has to change because now the same actions obviously lead to different perceptions

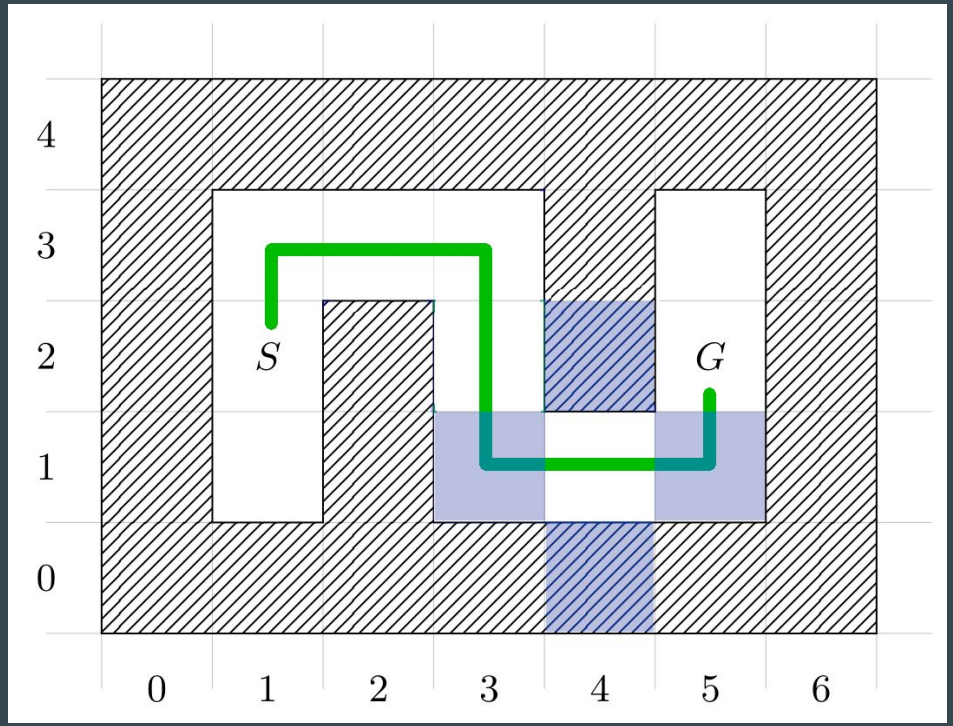
each unexpected transition enables the selection of a more appropriate situation

a *singular identifier* for the current situation provides the agent with perceptual information on the hidden state of the environment

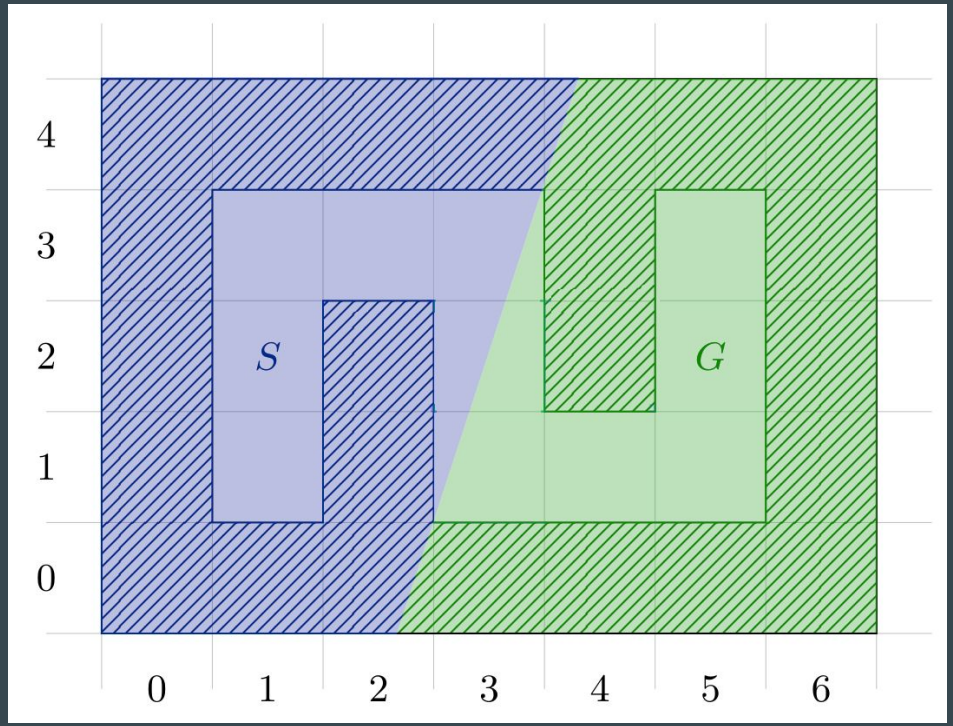
problem setting (same perception and action, different following perception)



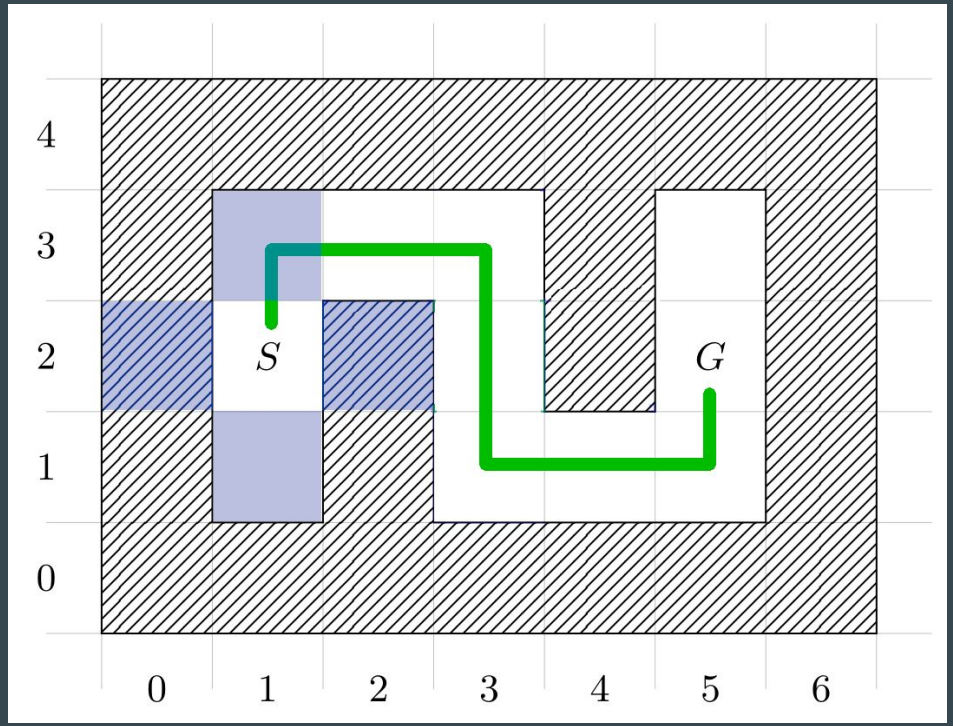
problem setting (same perception and action, different following perception)



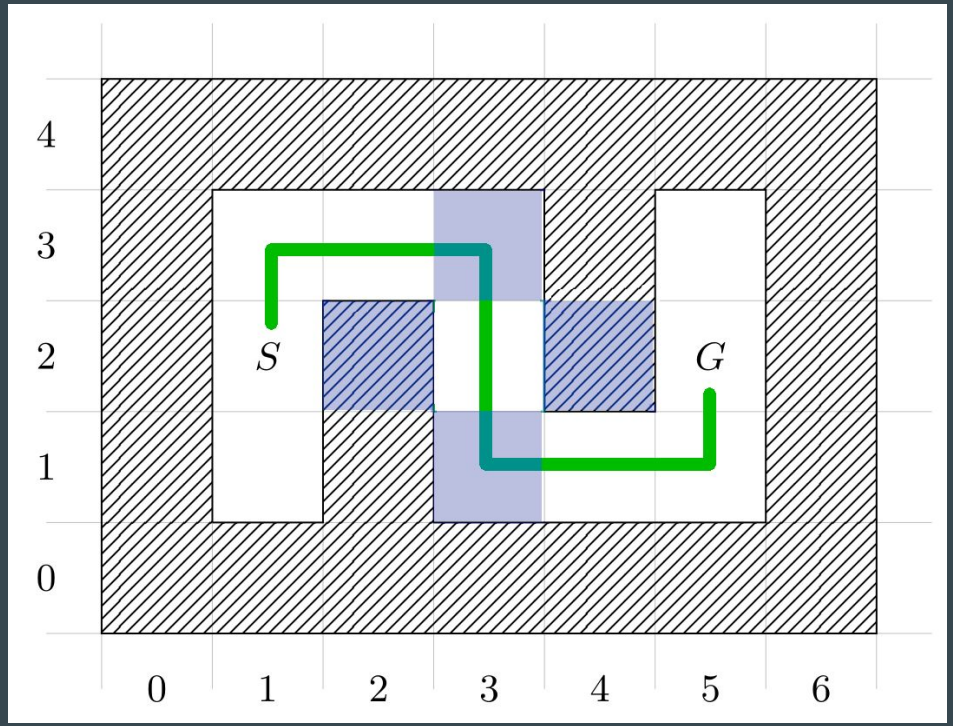
partitioning of environment



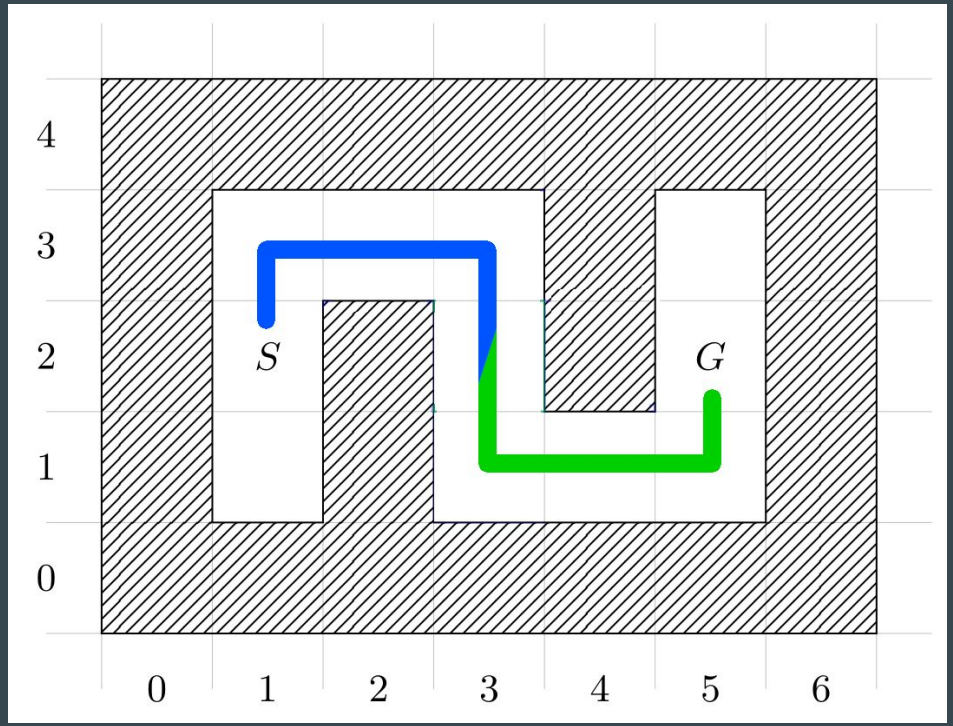
problem setting (same perception, different action required)



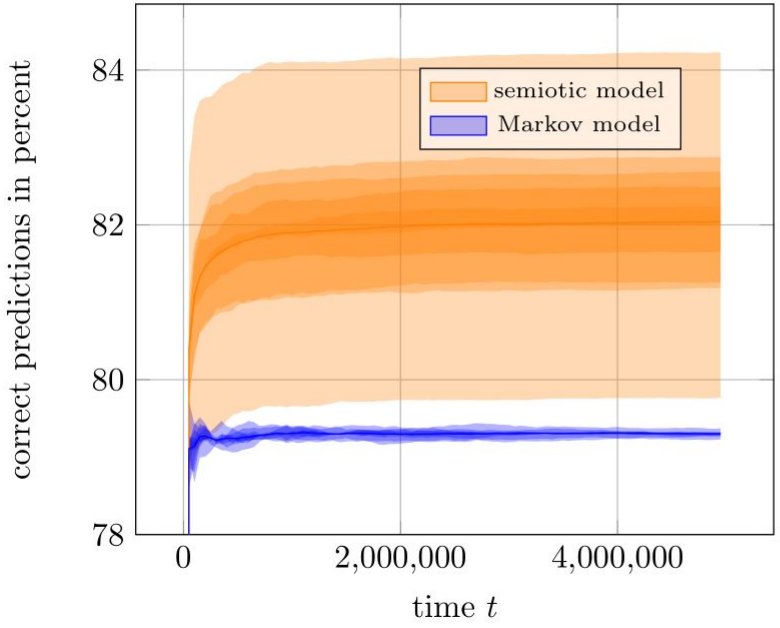
problem setting (same perception, different action required)

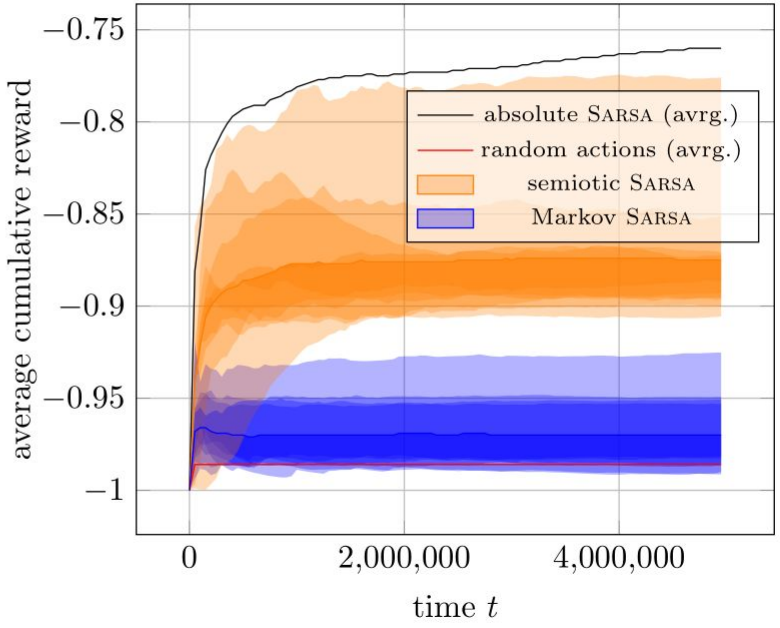


partitioning of path



evaluation results





related articles (containing additional information and complete bibliography)

Wernsdorfer, M. (2018). How failure facilitates success. In *International Conference on Artificial General Intelligence* (pp. 292-302). Springer, Cham.

Wernsdorfer, M. (2018). A phenomenologically justifiable simulation of mental modeling. In *International Conference on Artificial General Intelligence* (pp. 270-280). Springer, Cham.

Wernsdorfer, M. (2018). A time-critical simulation of language comprehension. In *International Conference on Artificial General Intelligence* (pp. 281-291). Springer, Cham.