On the interplay between goals and action-orientedness

Manuel Baltieri - 18th October 2024

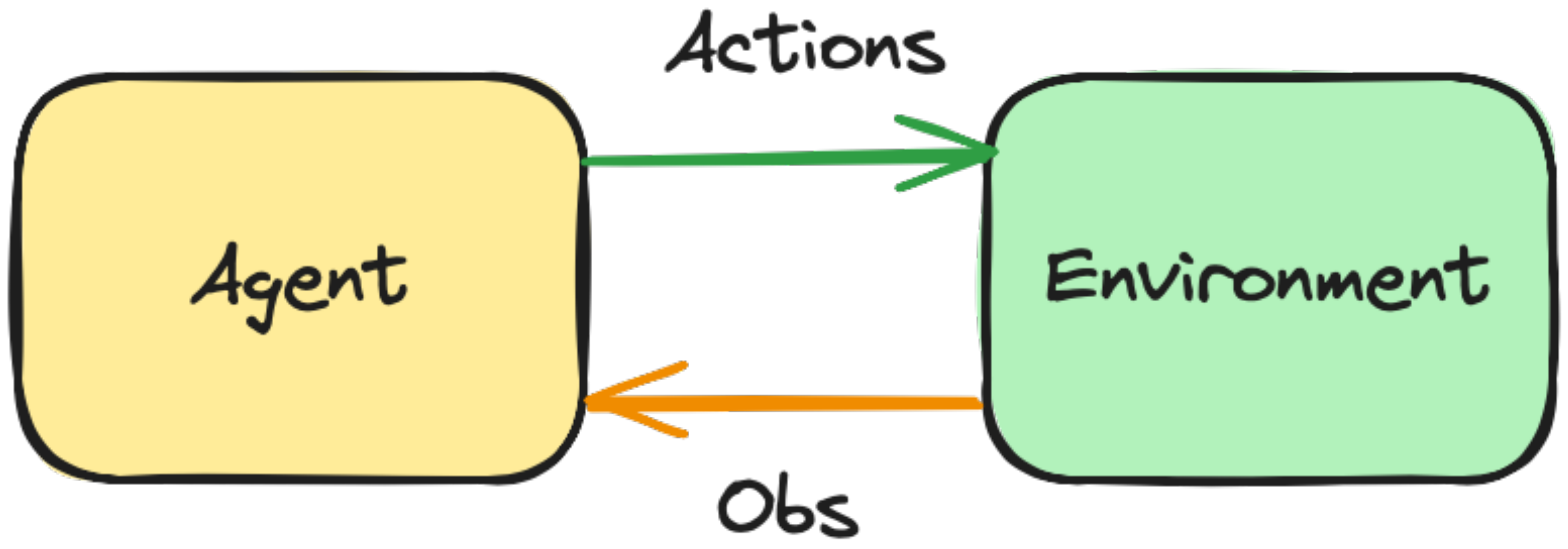
ARAYA

Contents

- Agents and (simple) beliefs
- A detour to Fristonland
- Simple beliefs via compression of MDPs (?)
- (Tentative) Compressed MDPs for minimal cognition

What I am interested in

Background



«[...] the rule "collect truth for truth's sake" may be justified when the truth is unchanging; but when the system is not completely isolated from its surroundings, and is undergoing secular changes, the collection of truth is futile, for it will not keep.»

(Ashby, 1958)

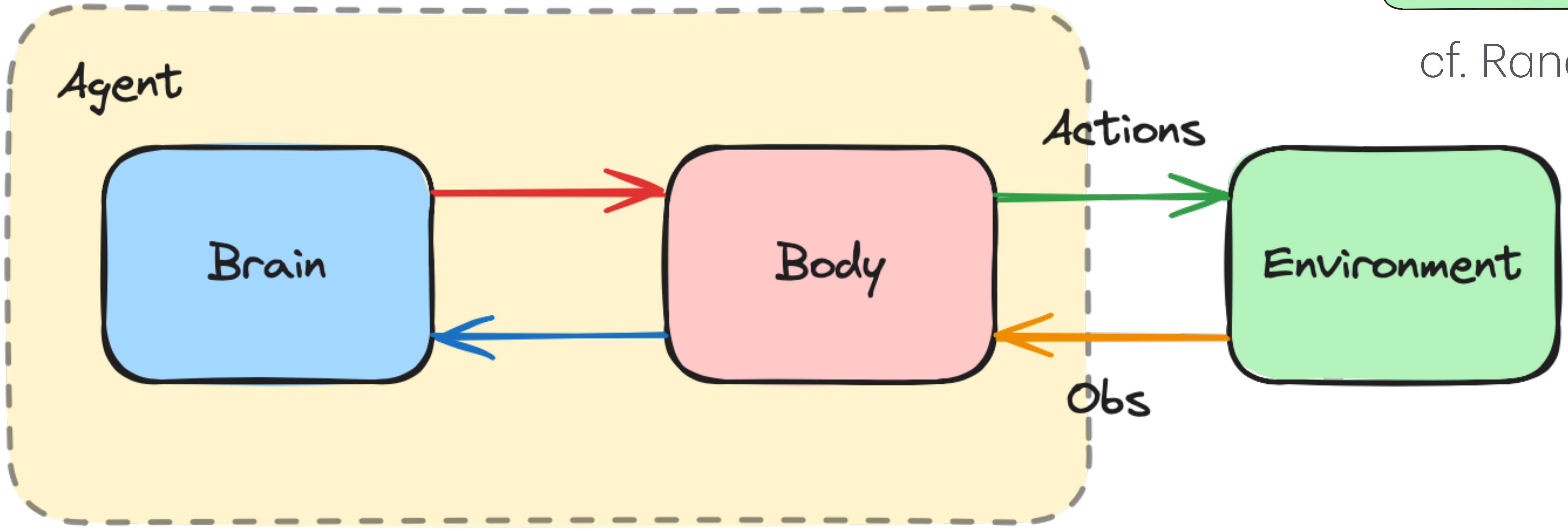
What agents “know” about their environment

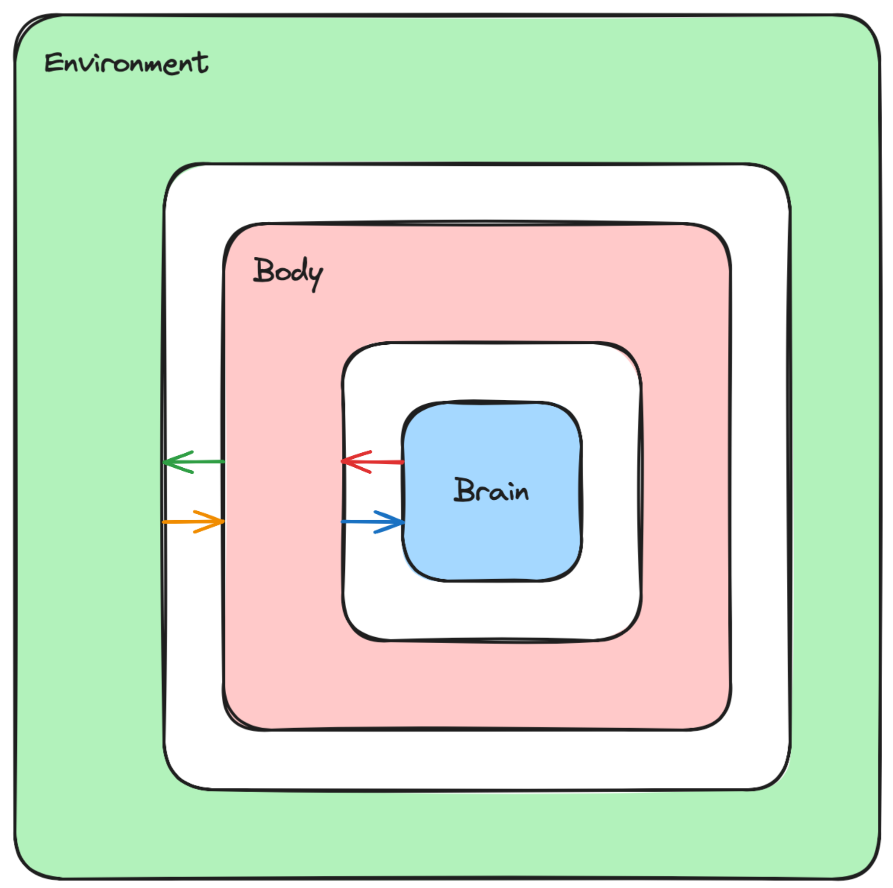
Or rather, what we should believe agents “know”

- What beliefs can we attribute to an agent solving a task?
- What are some interesting (minimal?) classes of such beliefs?
- What goals can we attribute to an agent?
- What is the relation between goals and beliefs we attribute to a system?
- ...

Unpacking that a little

Factorising the agent

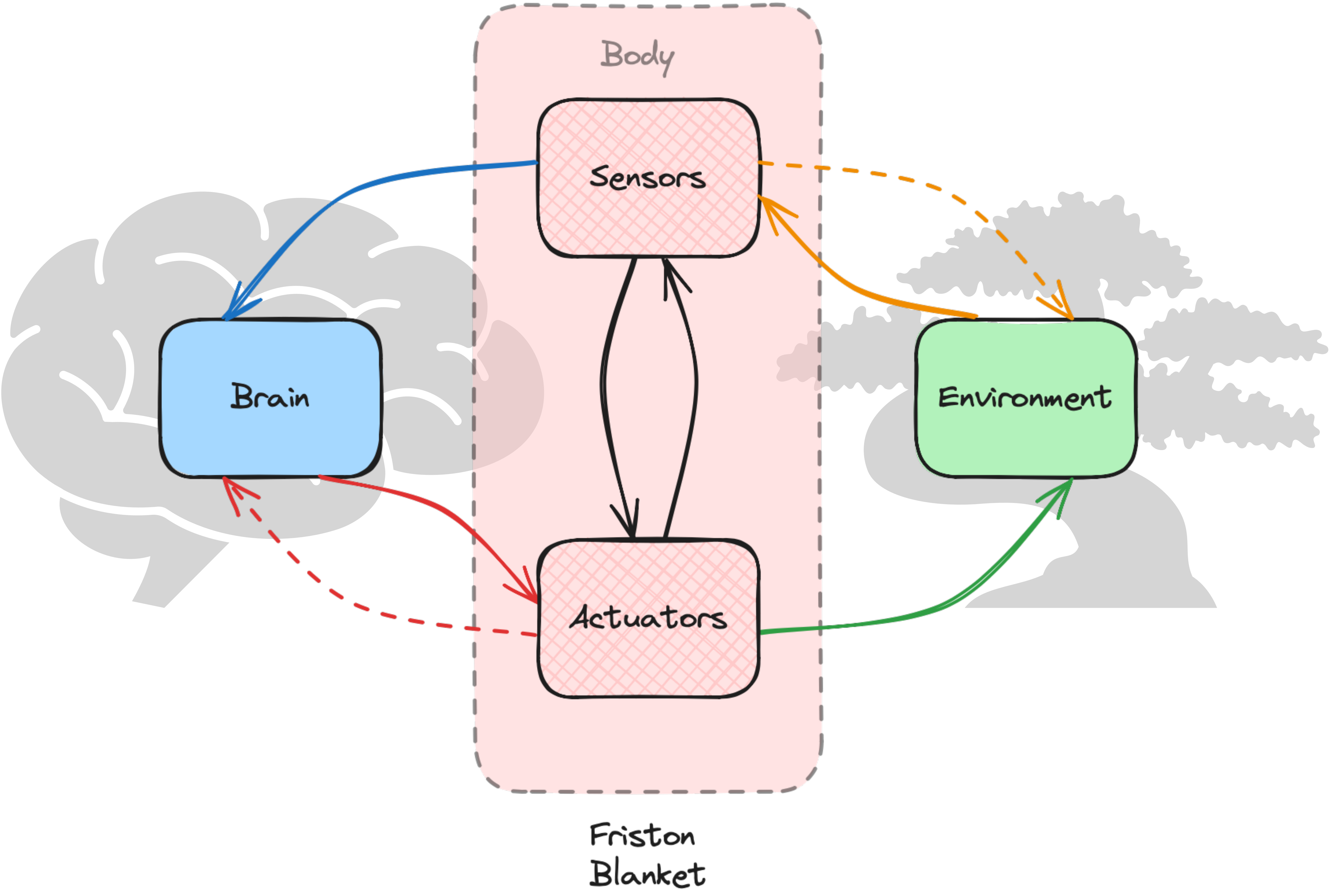


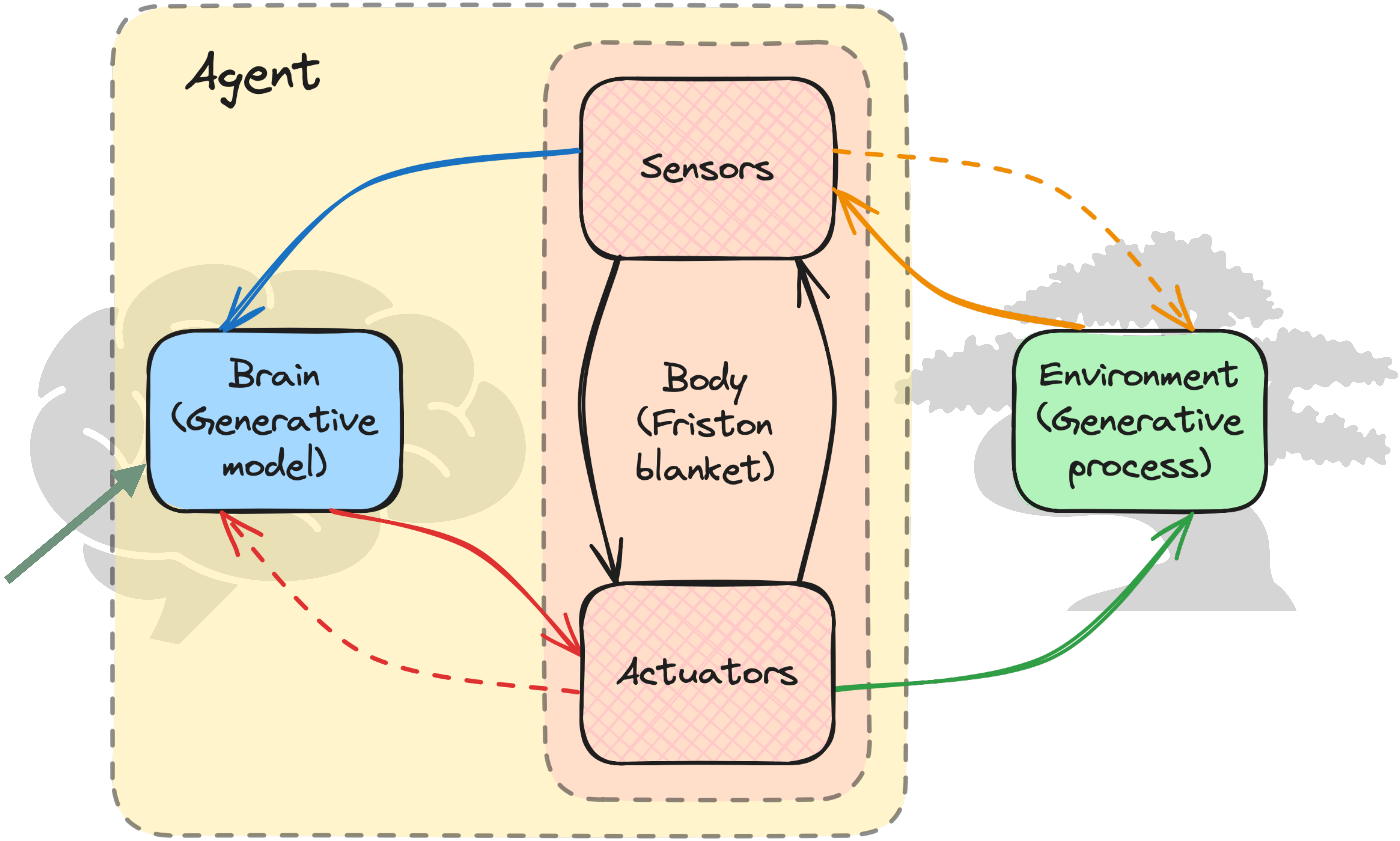


cf. Randy Beer

~~Alice~~ Manuel in Fristonland

Friston blankets, boundary factored into sensors and actuators



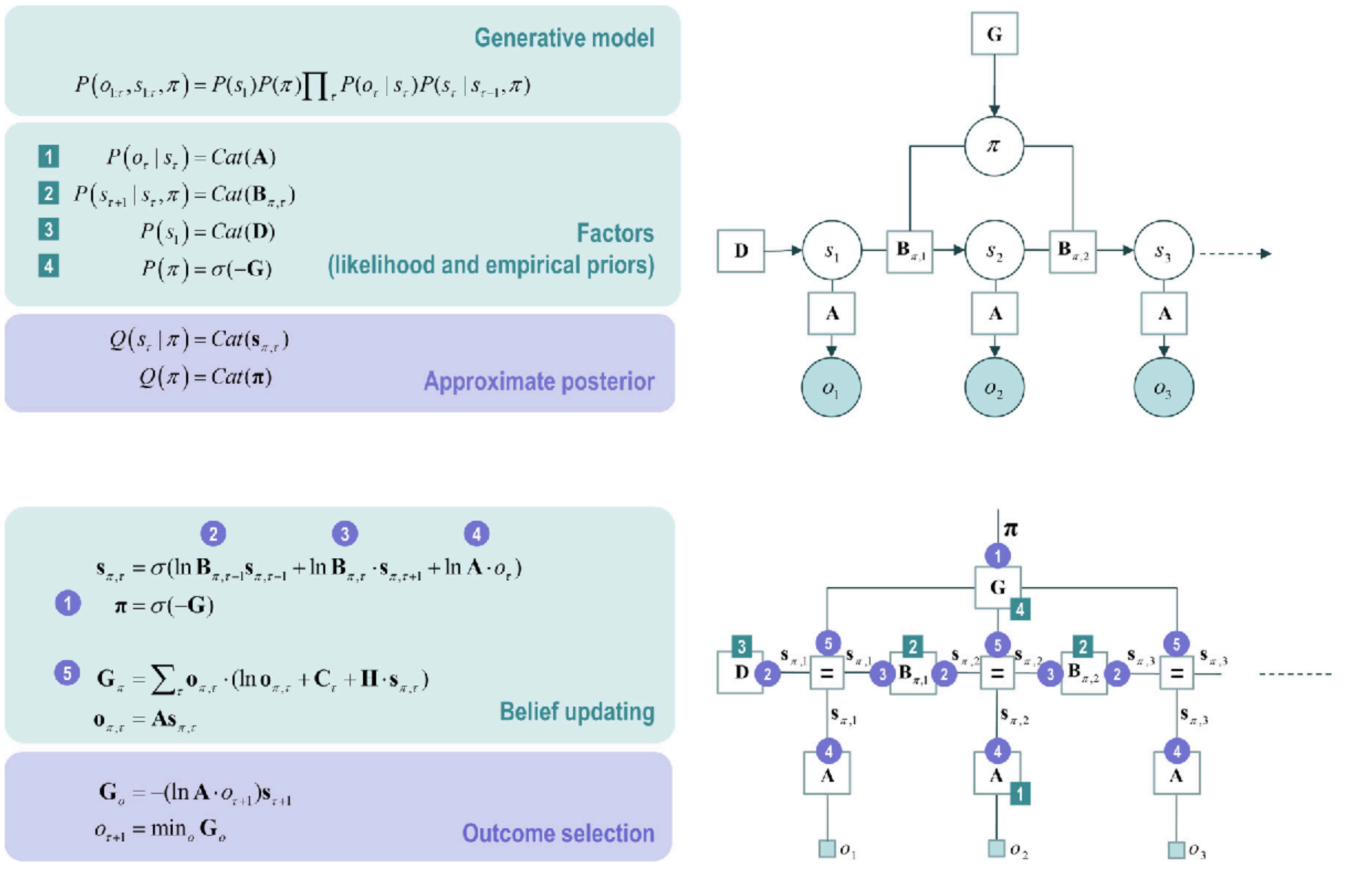


Brain states parametrise beliefs about “external” states, GM describes such beliefs

Active inference

What agents DO

- Perception, decision making, planning and learning based on approximate Bayes
- Assumes POMPDs/state-space models problem structure (~ RL setup)
- Provides an alternative cost function (expected free energy)
- ...ideally one that is derived from the FEP, but it can stand without it

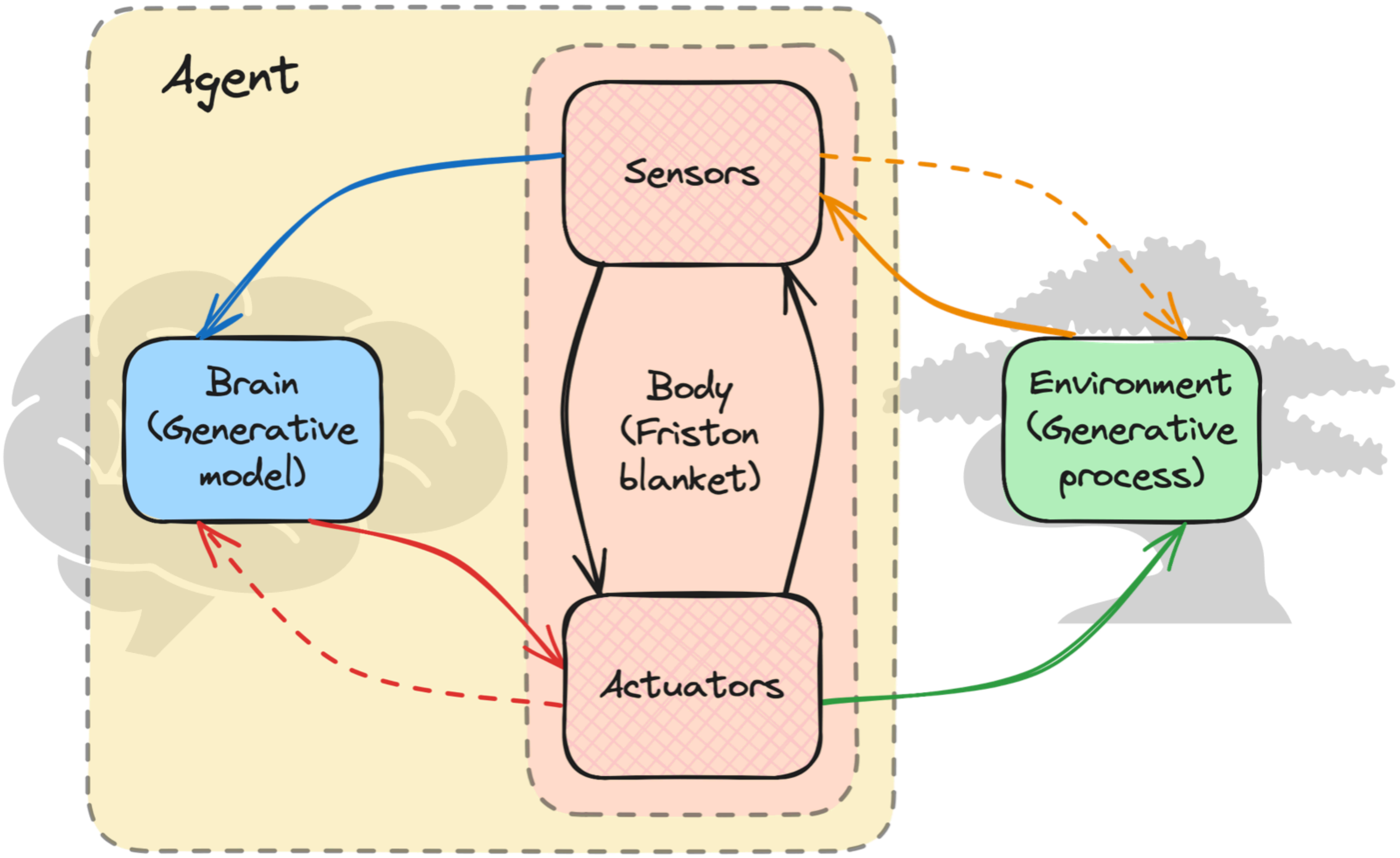


...nerative models for discrete states and outcomes. Upper left panel: These equations specify the generative model. A generative model i...

The free energy principle

What agents ARE

- A foundational theory of agents, (living) systems, “things”
- A thing is a “thing” if and only if it (appears to) minimise(s) free energy
- Friston blankets as a “veil” that separates internal from external states
- Internal states parametrise beliefs about external states in some contexts



«Neural representations, this work has suggested, are not action neutral mirrors of the world. Instead they are in some deep sense 'action-oriented' (Clark 1997, Engel et al. 2013). They are geared to promoting successful, fast, fluent actions and engagements for a creature with specific needs and bodily form. Such representations will be as minimal as possible, neither encoding nor processing information in costly ways when simpler routines, combined with world-exploiting actions, can do the job.»

(Clark, 2015)

Different types of generative models?

- Gathering knowledge vs. achieving a goal
- Simplified generative models, encoding sensorimotor information/Umwelt

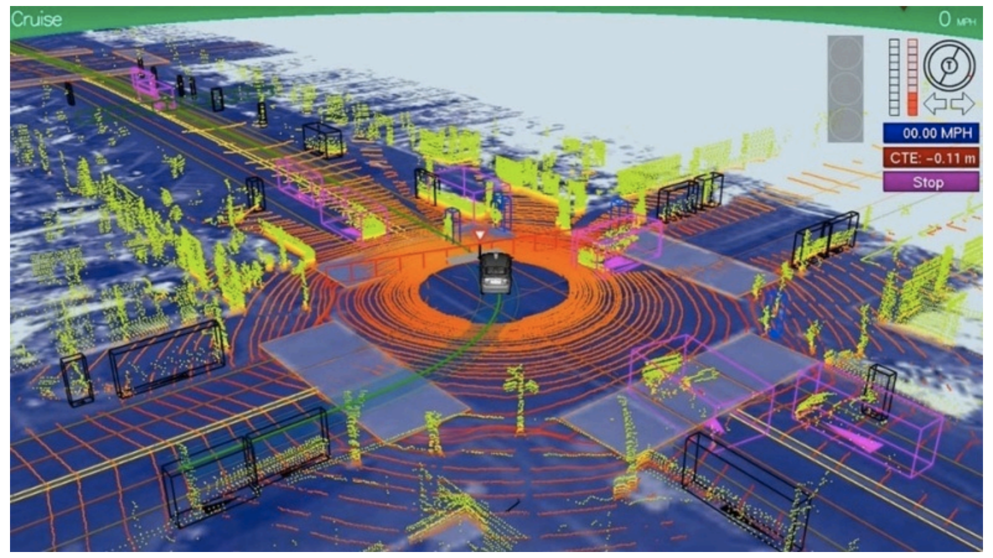
Example: Outfielder problem (Fink et al., 2009)

1) ~~Trajectory prediction (TP)~~
2) Optical Acceleration Cancellation (OAC)

Action-oriented generative models

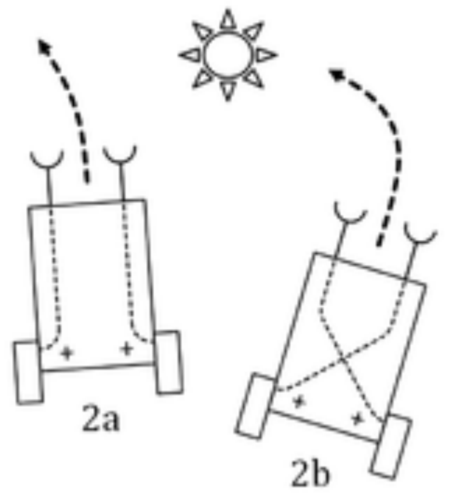
Example task: agent performing phototaxis

Perception-oriented



e.g. SLAM

Action-oriented



e.g. Braitenberg vehicles

The linebot

McGregor et al. (2015) look at FEP to understand what it can say about an agent's beliefs.

This agent is trying to reach a goal position when the only information available is high/low concentration of a certain chemical.

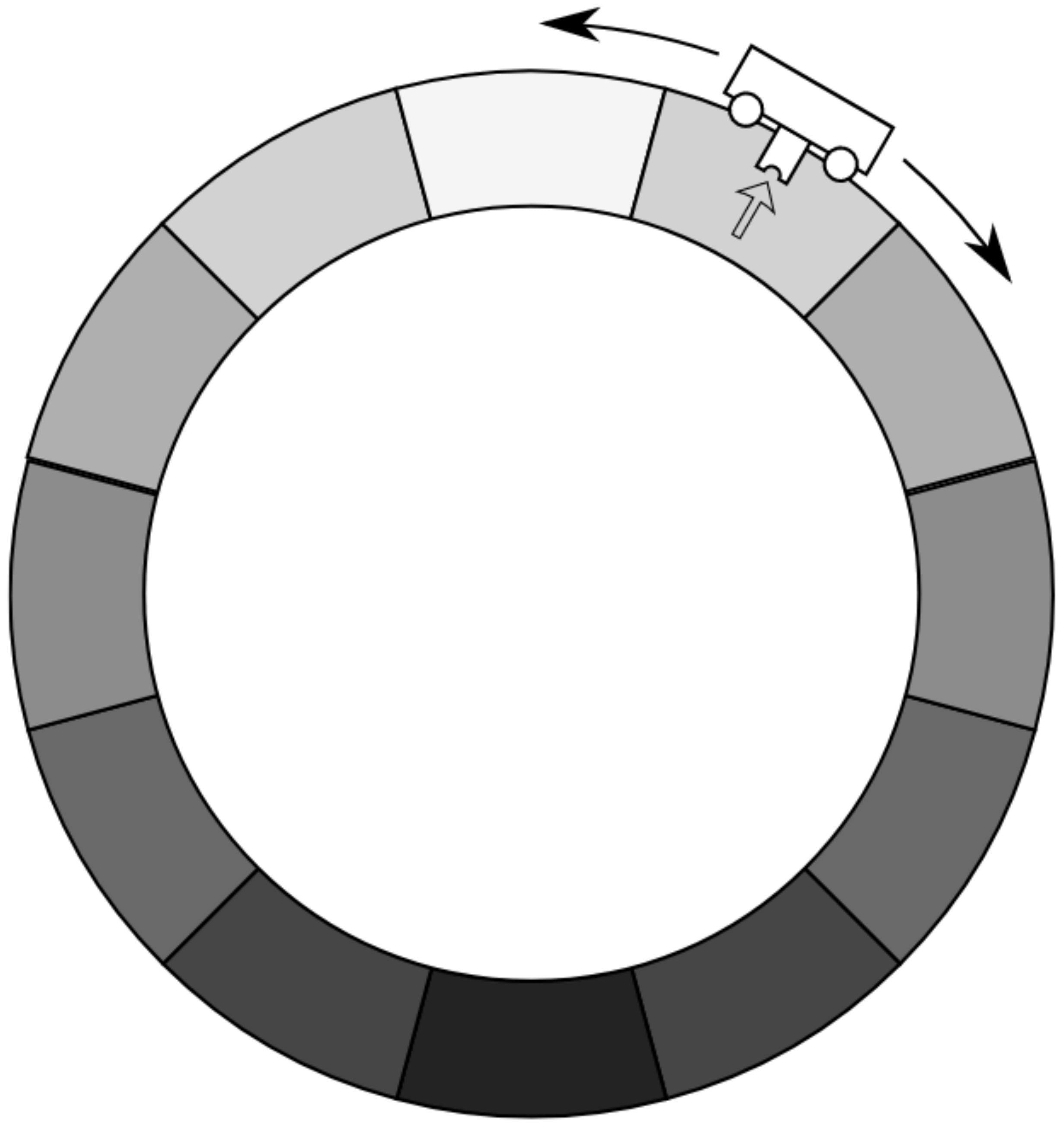


Figure 1: Illustration of agent-environment system. The agent has a sensor which reads $High$ or Low and is sensitive to chemical concentration. The agent's motor can attempt to move the agent clockwise or anticlockwise.

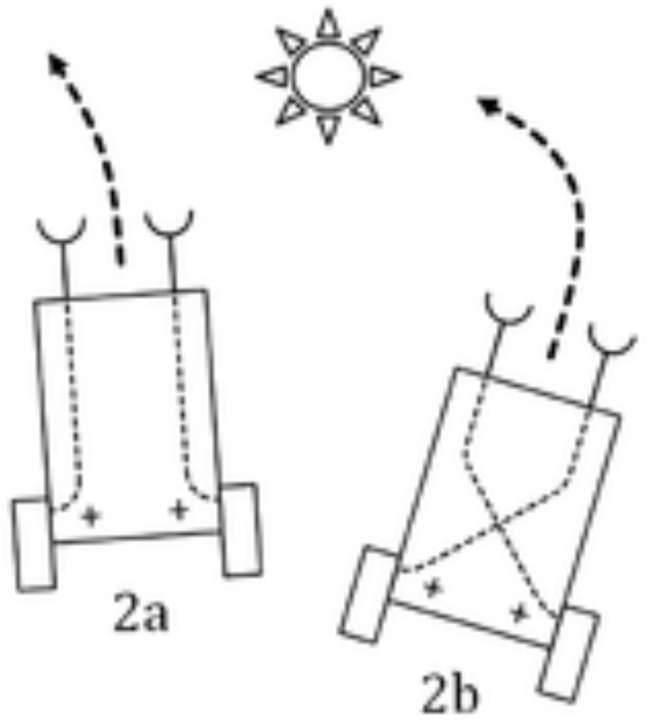
The linebot

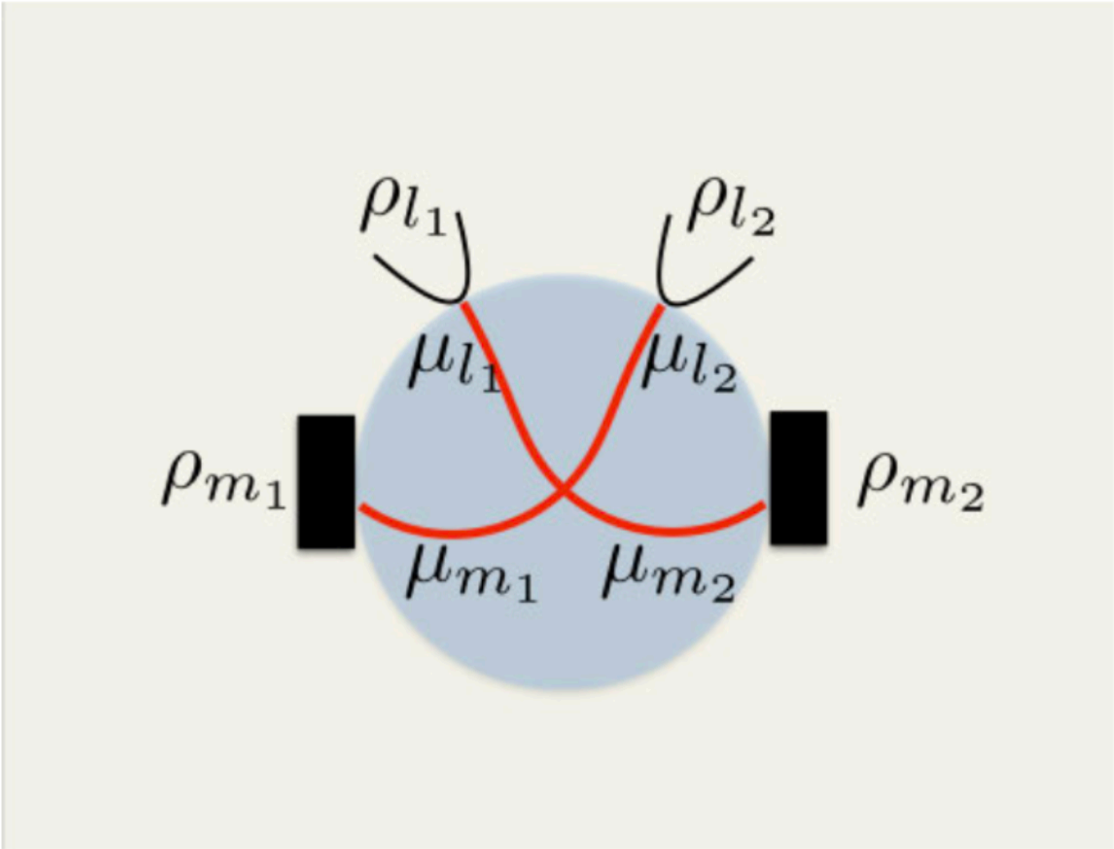
…with simplified beliefs

My master dissertation: what if the agent beliefs were “simplified” (hierarchical model with two levels: half circle + left/right)

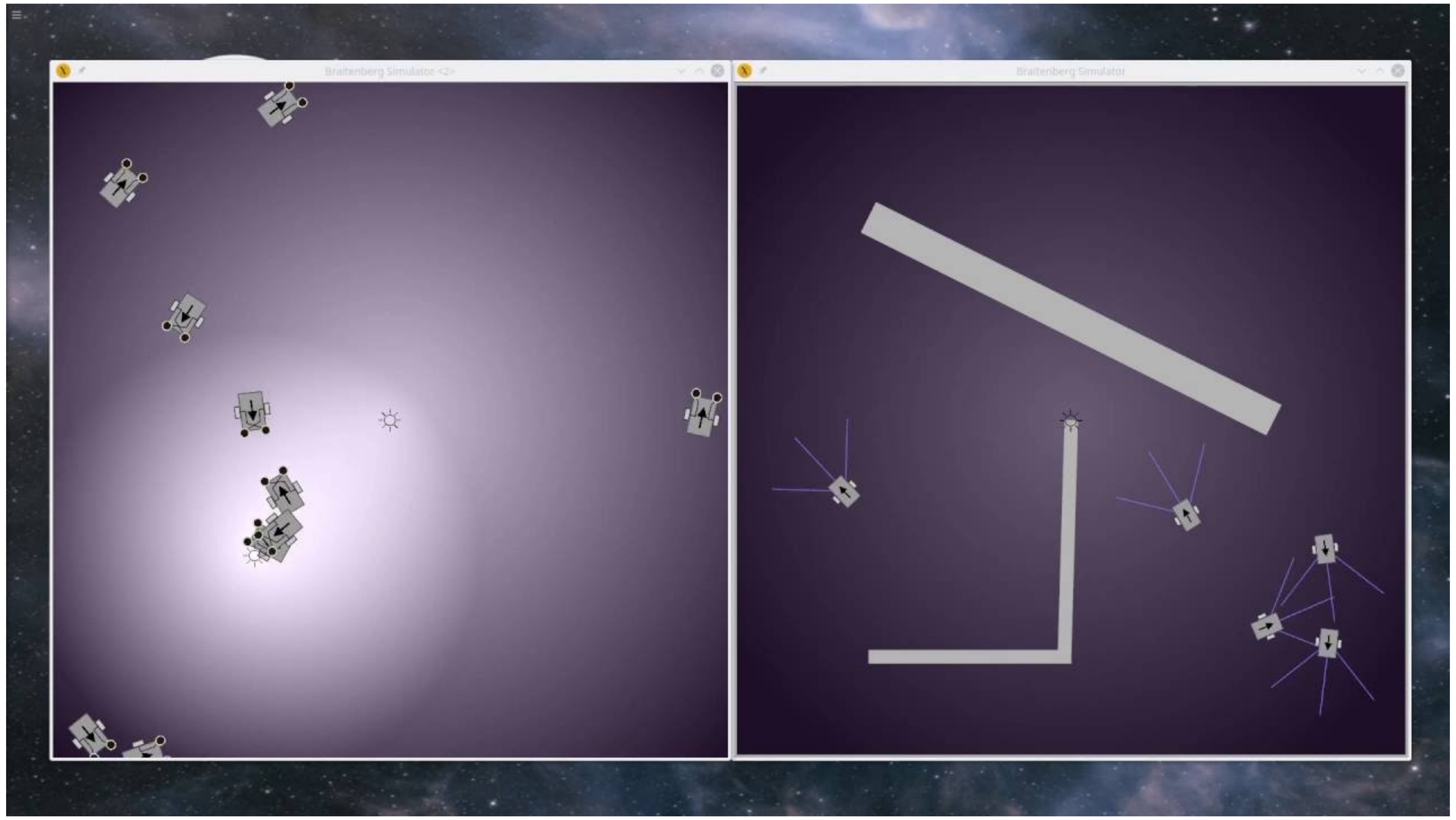
Braitenberg vehicles

Photo/chemo/rheo/tropo/… taxis





- Vehicles “2”
- Agent with two sensors and two wheels
- Sensors and wheels connected by wires
- Implementation: (Left/right) Wheel rotational velocity = constant * (right/left) sensory reading



5-10 years later...

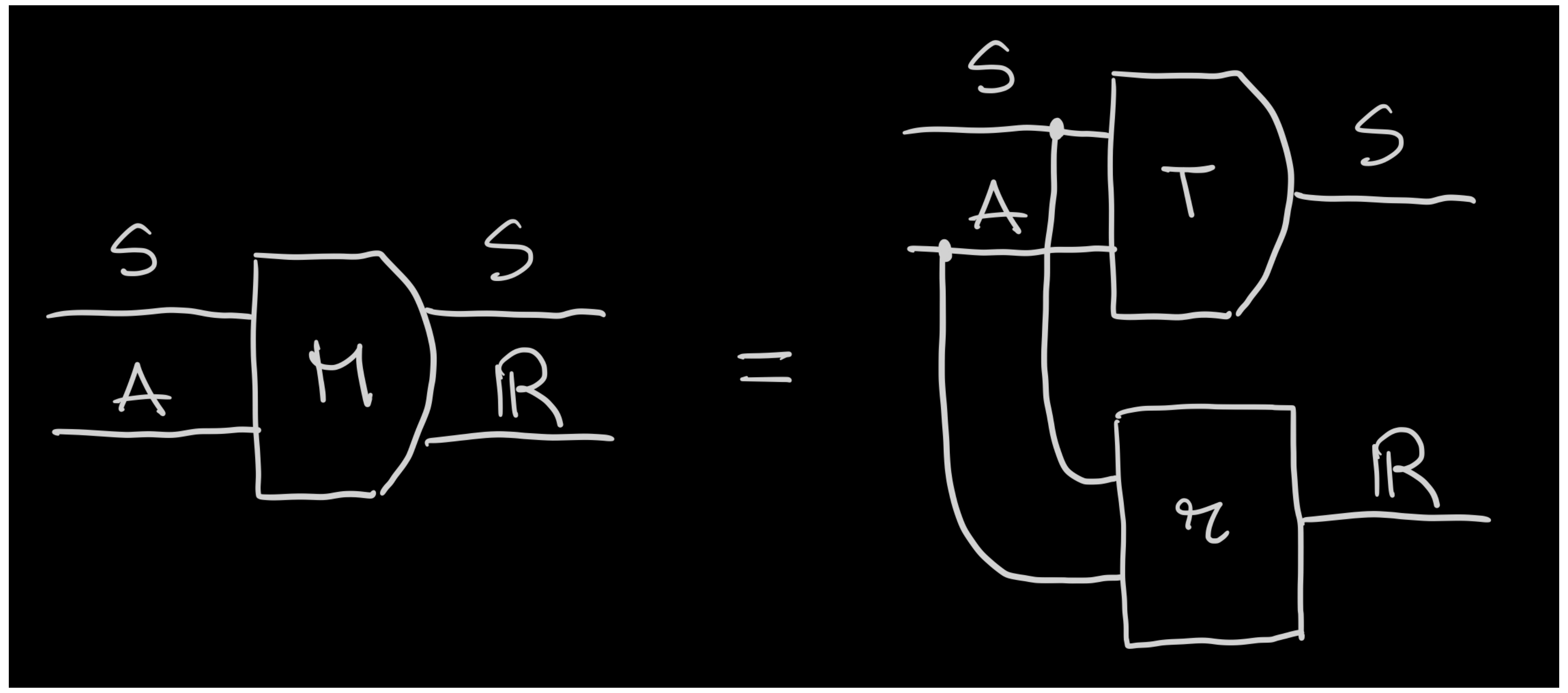
Markov decision processes

Cheap ways towards goals

A Markov decision process is a tuple (S, A, T, γ, r), where:

- S is the state space,
- A is the action space,
- $T : S \times A \to P(S)$ is the transitions dynamics, often written as $P(s_{t+1} \mid s_t, a_t)$,
- $\gamma \in [0,1)$ is called the discount factor,
- $r : S \times A \to \mathbb{R}$ is the reward function, giving a reward every time a transition is taken.

$$S \longrightarrow P(S)^A \times \mathbb{R}$$



Probabilistic bisimulation equivalence

Givan et al. (2003), but only one of the many definitions

Let (S, A, T, γ, R) be a Markov Decision Process, $s_i, s_j \in S$ states of the state space S and $a \in A$ an action of the action space A.

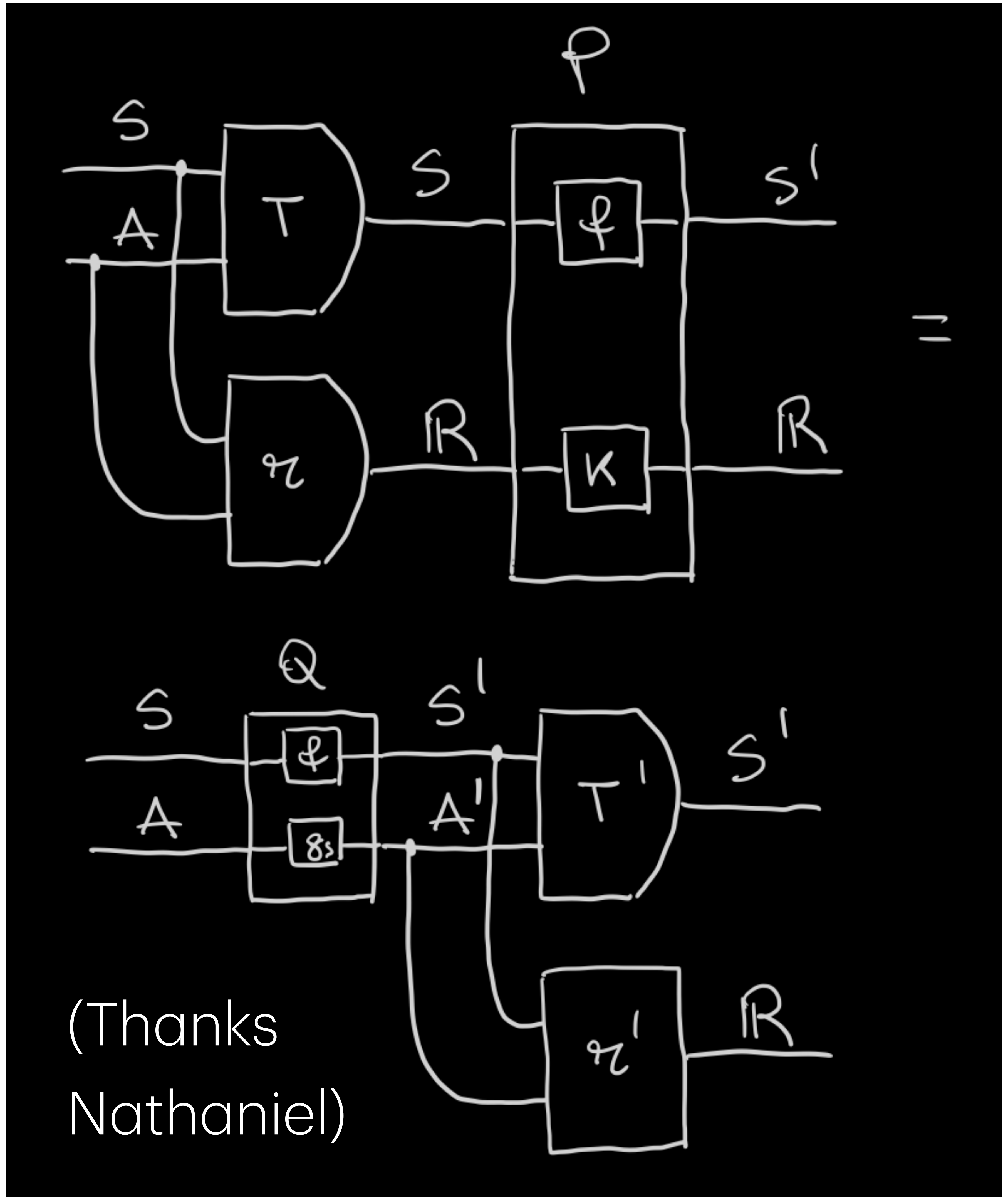
A probabilistic bisimulation equivalence is an equivalence relation $B \subseteq S \times S$ such that:

$$(s_i, s_j) \in B \ \ (\text{or} \ \ s_i \, B \, s_j) \implies P(E \,|\, s_i, a) = P(E \,|\, s_j, a) \quad \text{and}$$
$$R(s_i, a) = R(s_j, a)$$

for all equivalence classes $E \in S/B$, i.e. where $P(E \,|\, s, a) = \sum_{s' \in E} P(s' \,|\, s, a)$, and for all actions $a \in A$.

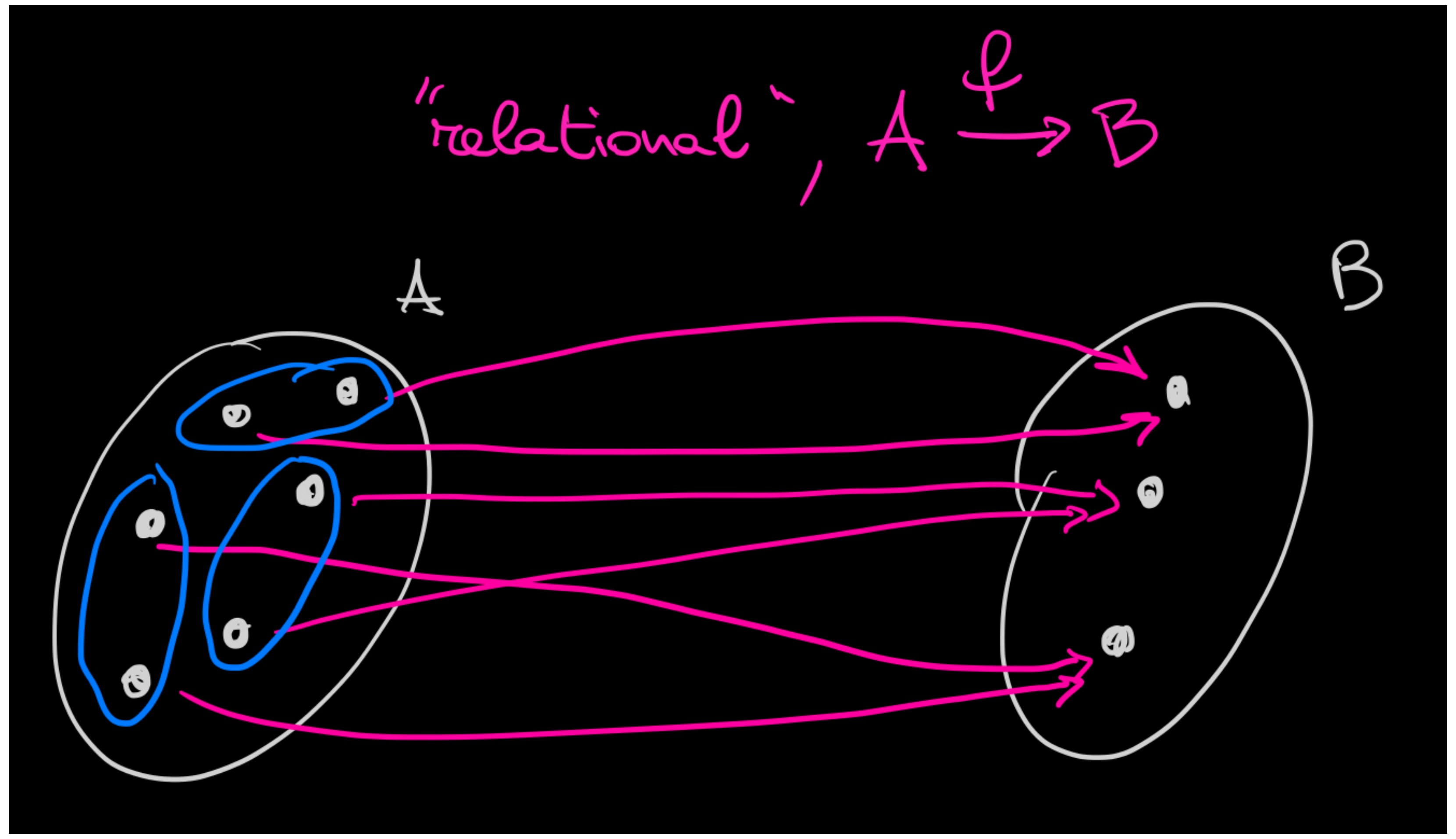
f, g: surjective

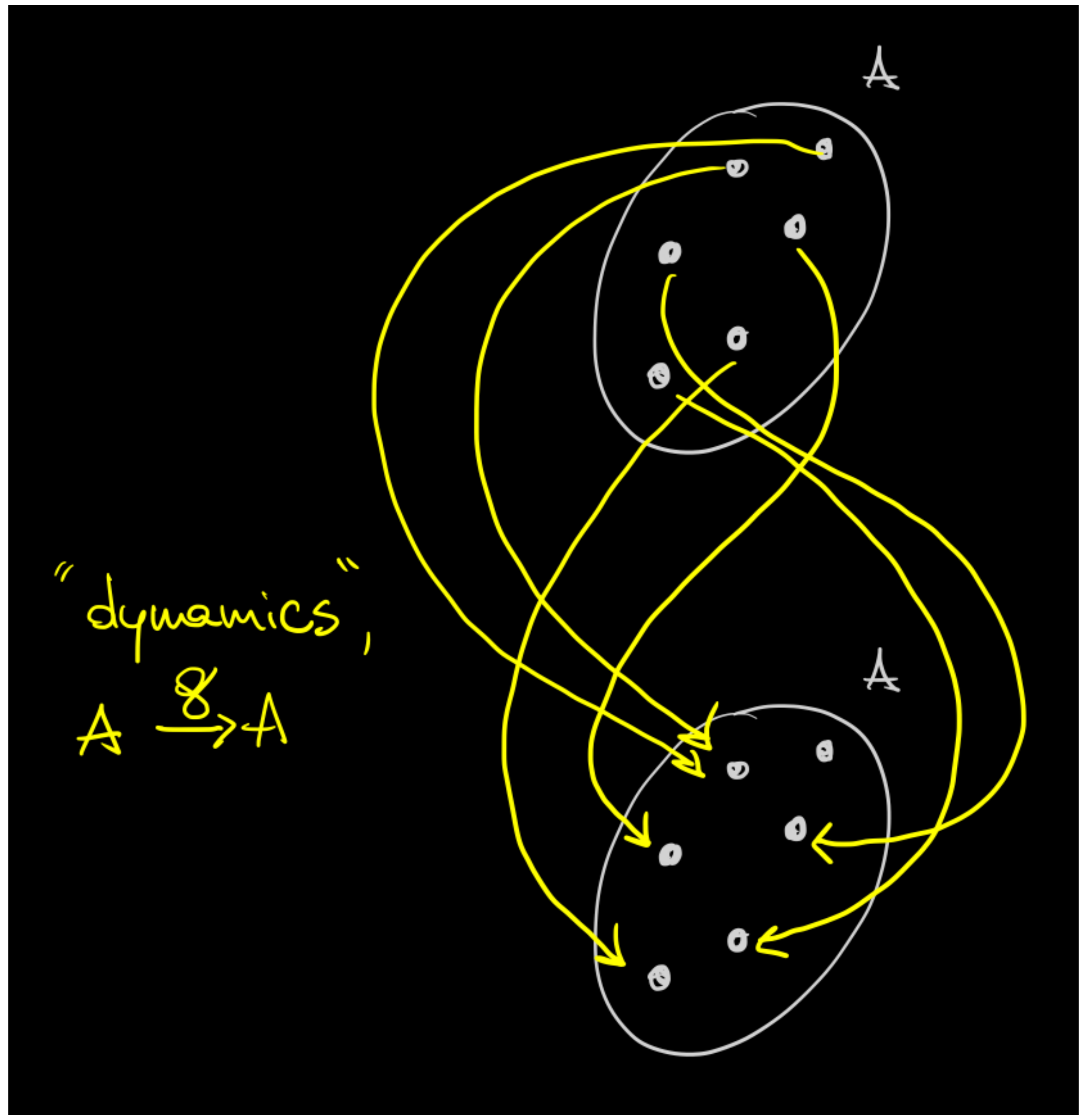
k: identity

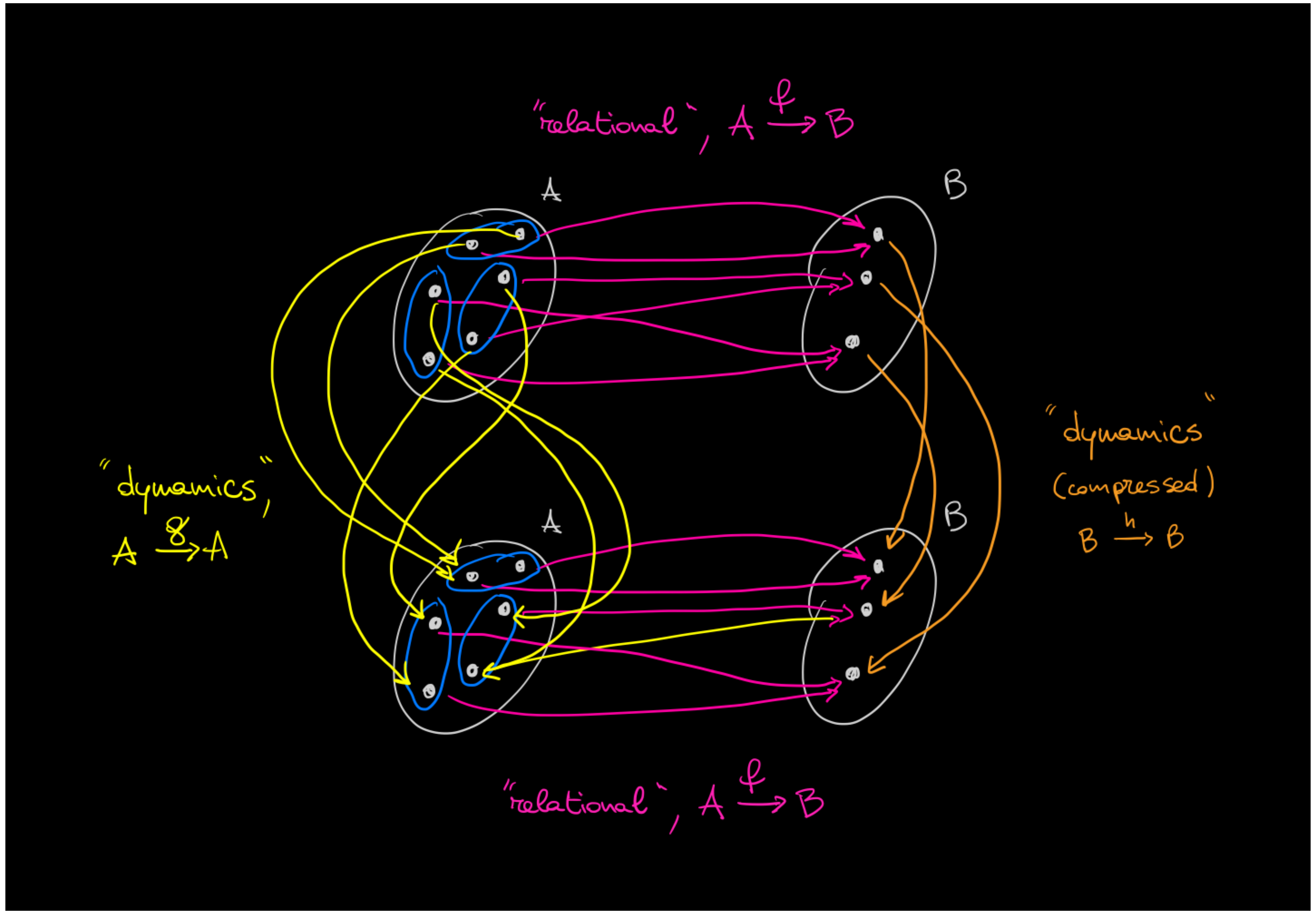


Same idea, simplified

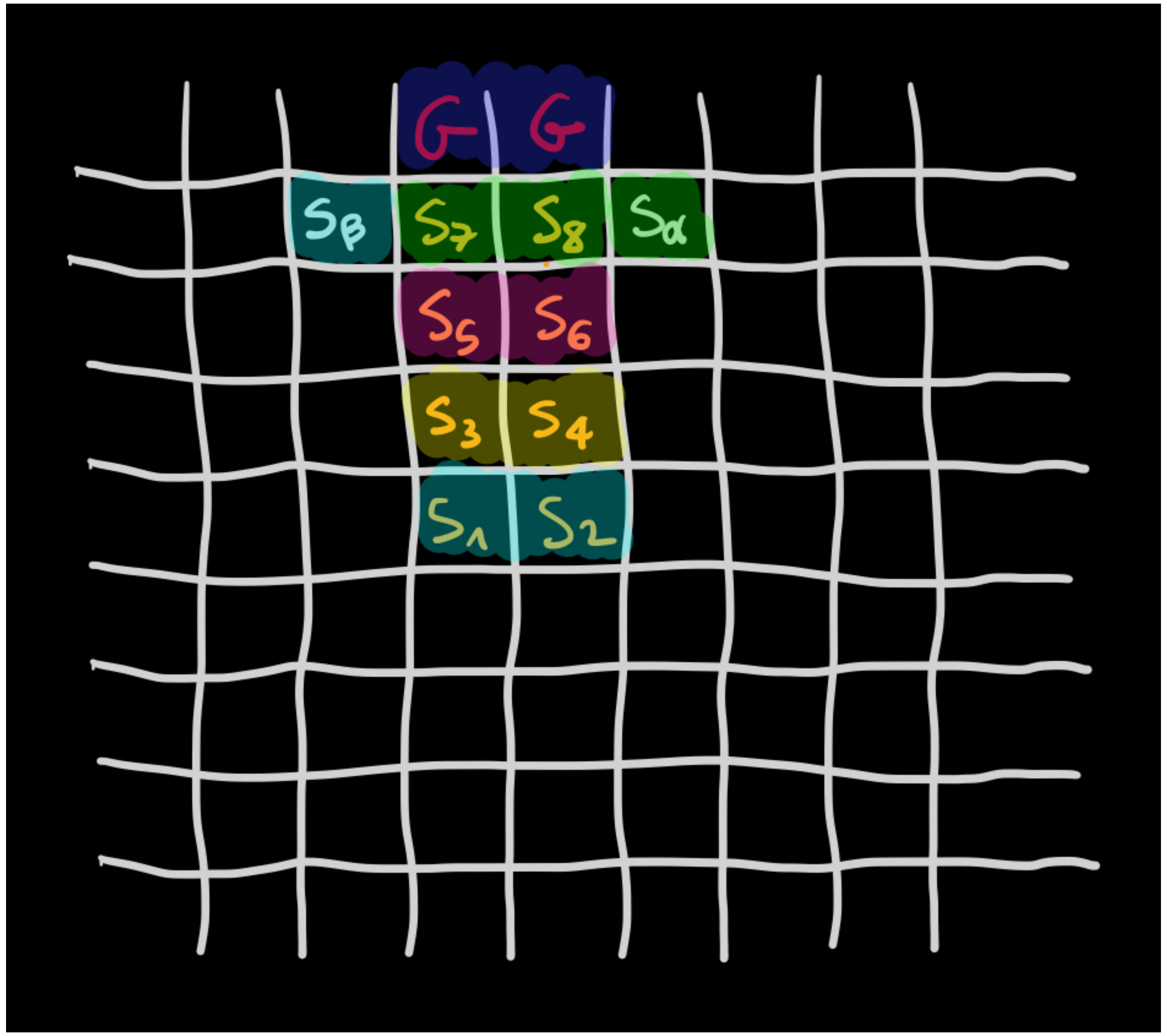
For closed dynamical systems, no MDPs

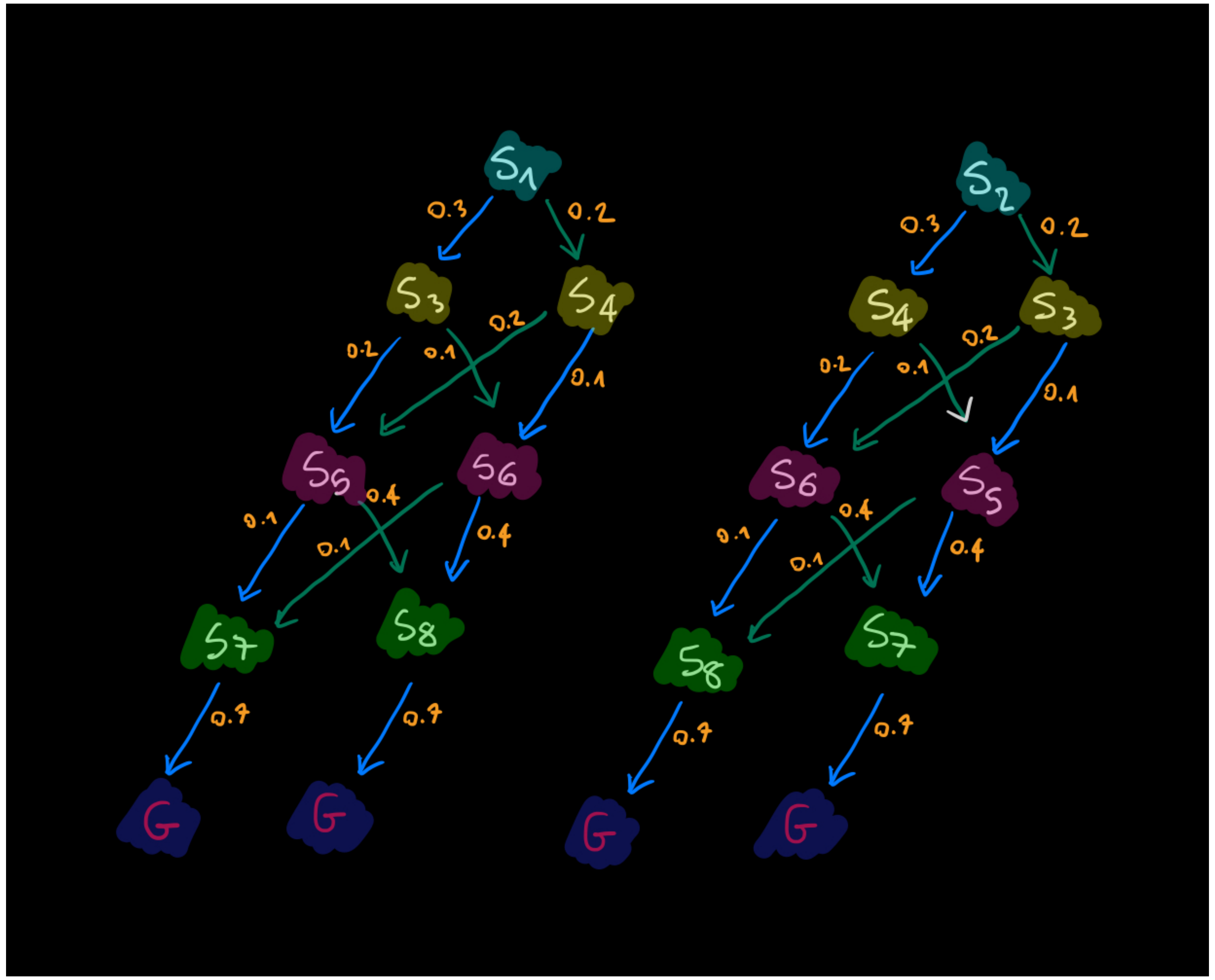


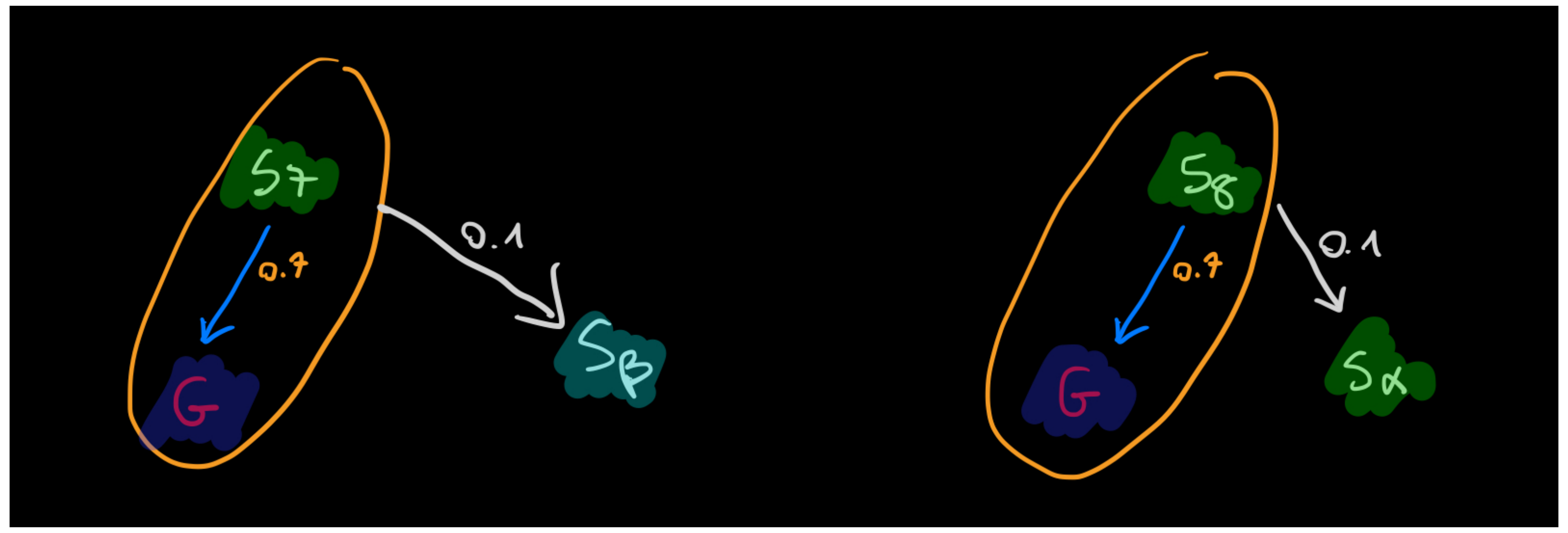


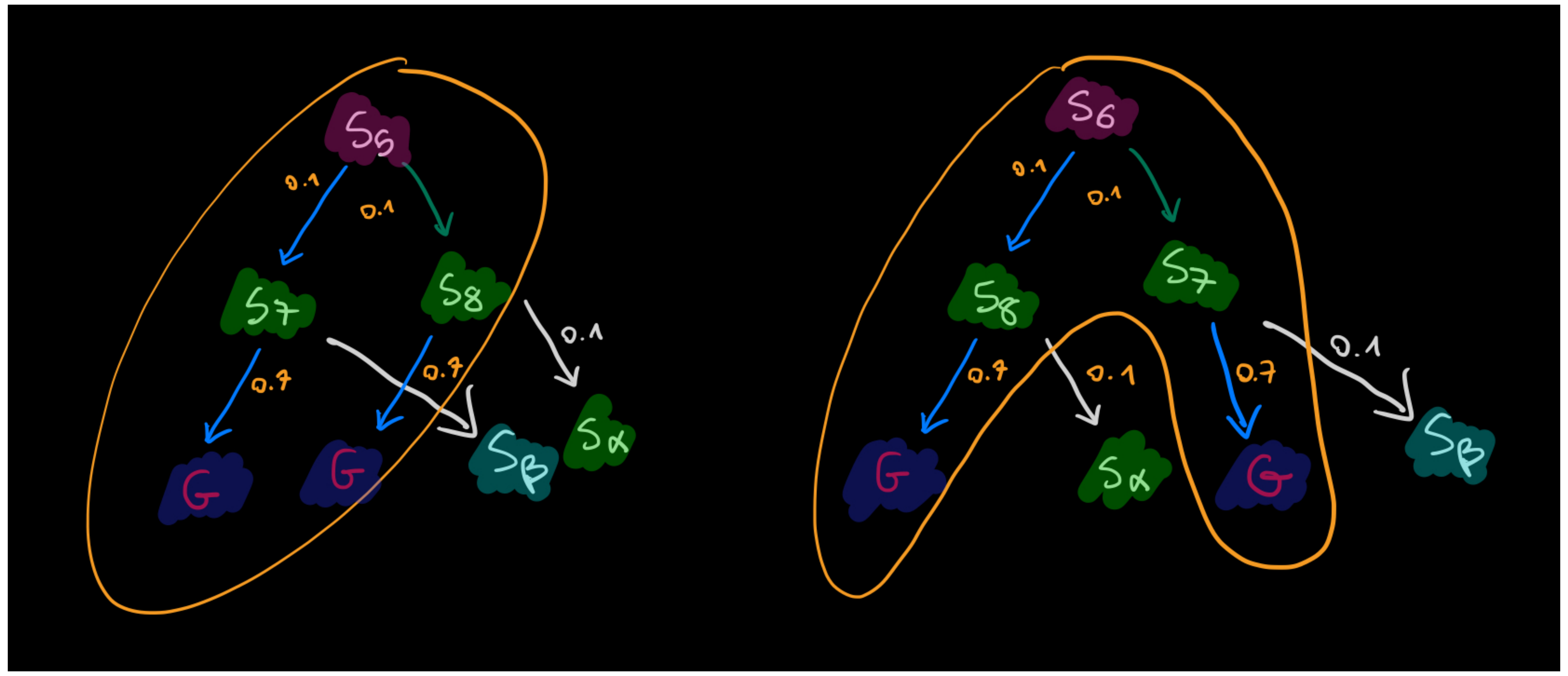


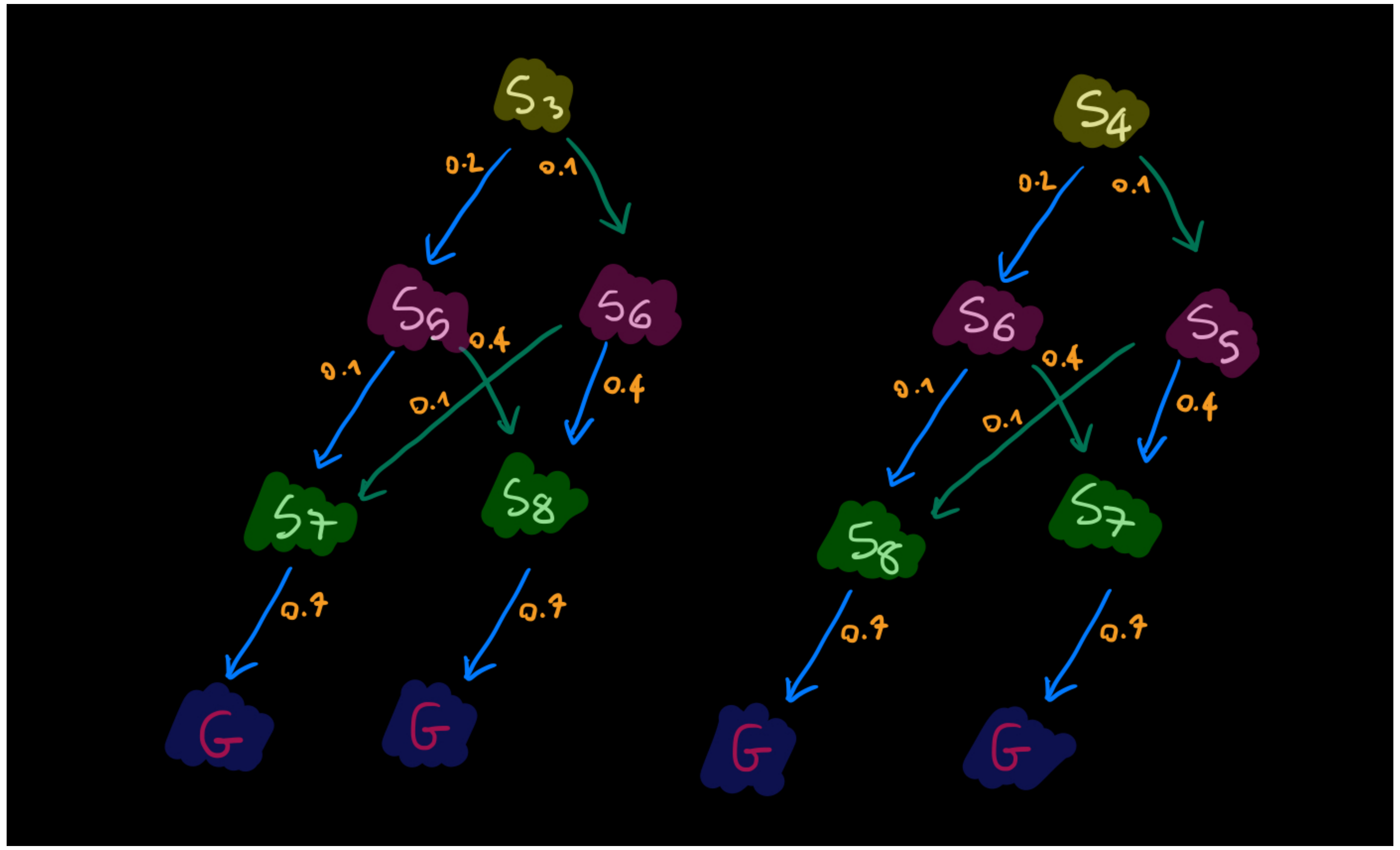
A worked out example

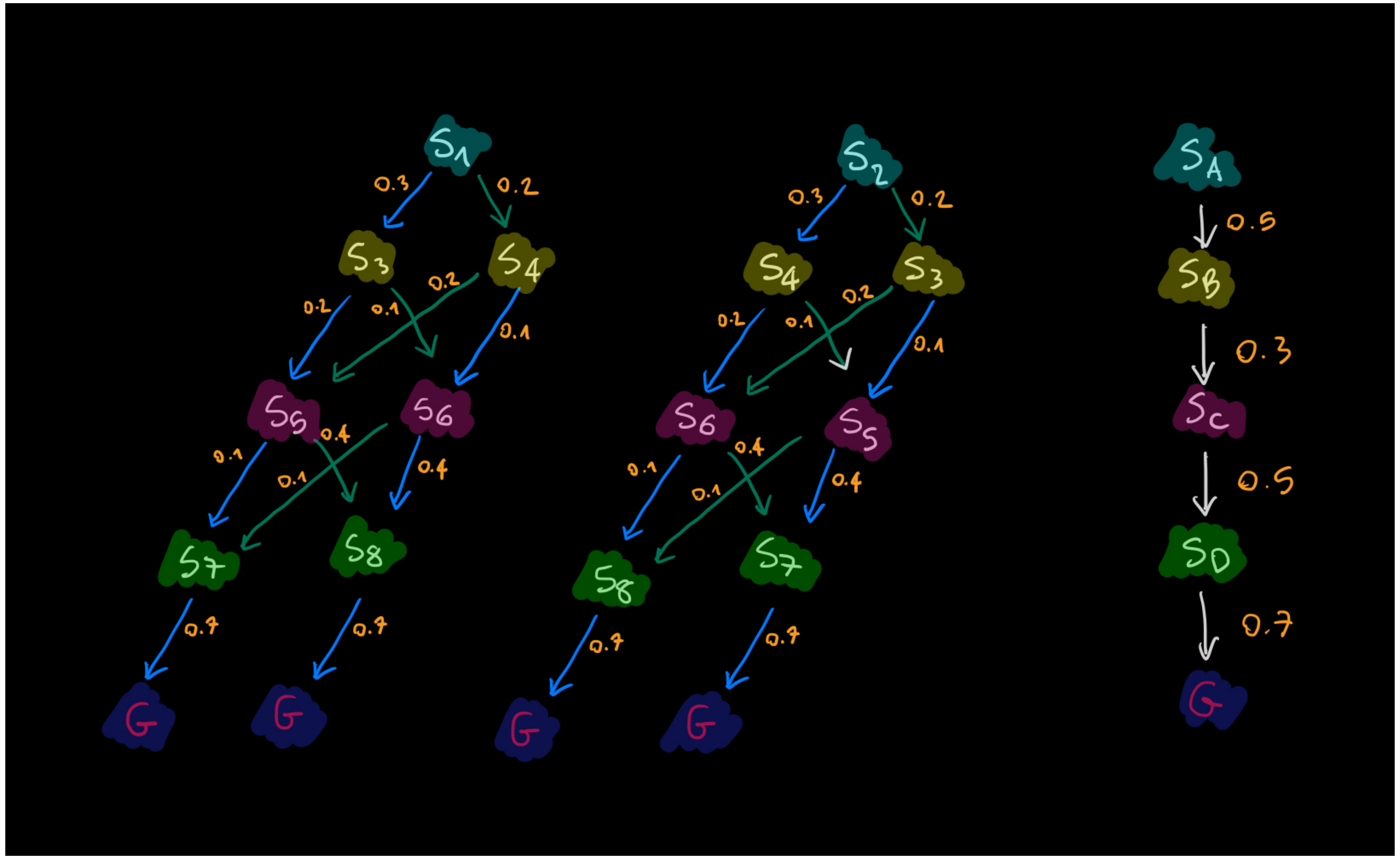


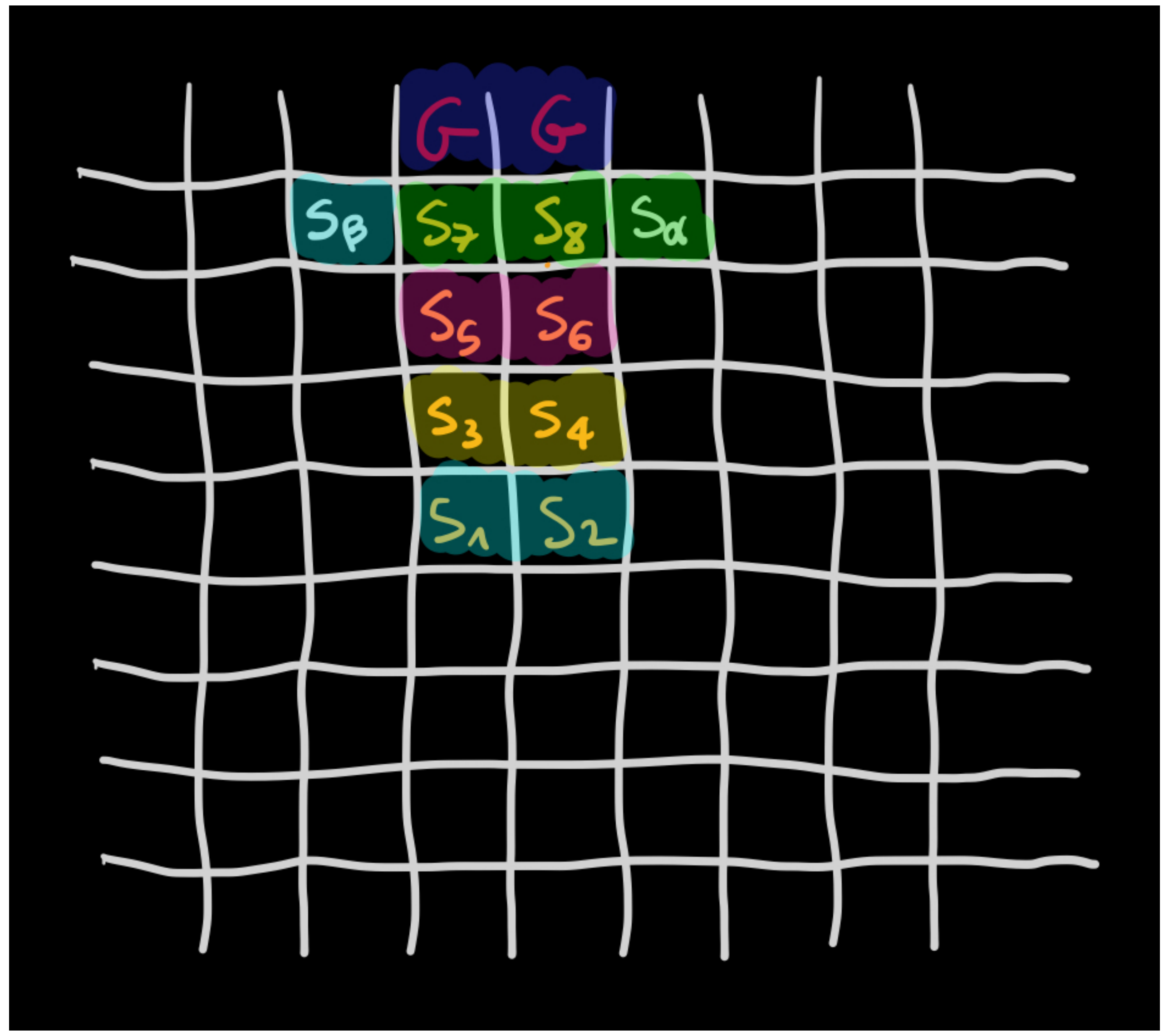



S7
0.7
G
0.1
Sβ
S8
0.7
G
0.1
Sα


S5
0.1
0.1
S7
S8
0.7
0.7
0.1
G
G
Sβ
Sα
S6
0.1
0.1
S8
S7
0.7
0.1
0.7
0.1
G
Sα
G
Sβ


S3
0.2
0.1
S5
S6
0.1
0.4
0.1
0.4
S7
S8
0.7
0.7
G
G
S4
0.2
0.1
S6
S5
0.1
0.4
0.1
0.4
S8
S7
0.7
0.7
G
G

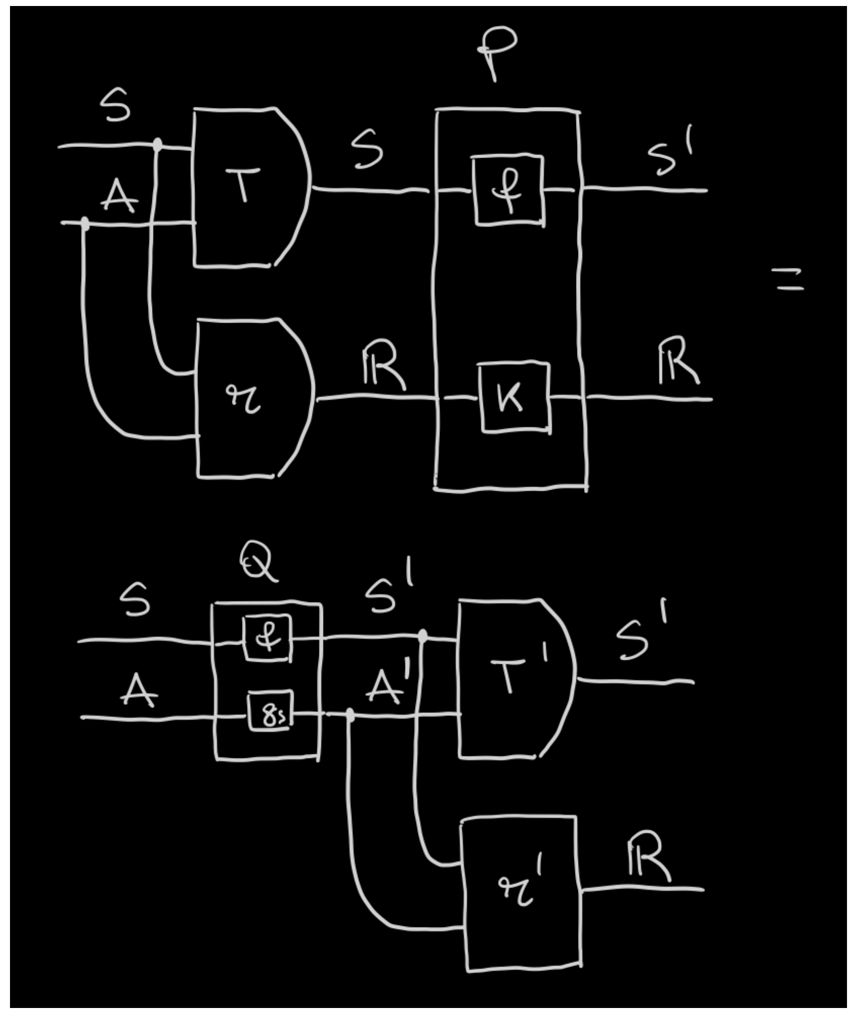

S1
0.3
0.2
S3
S4
0.2
0.1
0.2
0.1
S5
S6
0.1
0.4
0.1
0.4
S7
S8
0.7
0.7
G
G
S2
0.3
0.2
S4
S3
0.2
0.1
0.2
0.1
S6
S5
0.1
0.4
0.1
0.4
S8
S7
0.7
0.7
G
G
SA
0.5
SB
0.3
SC
0.5
SD
0.7
G

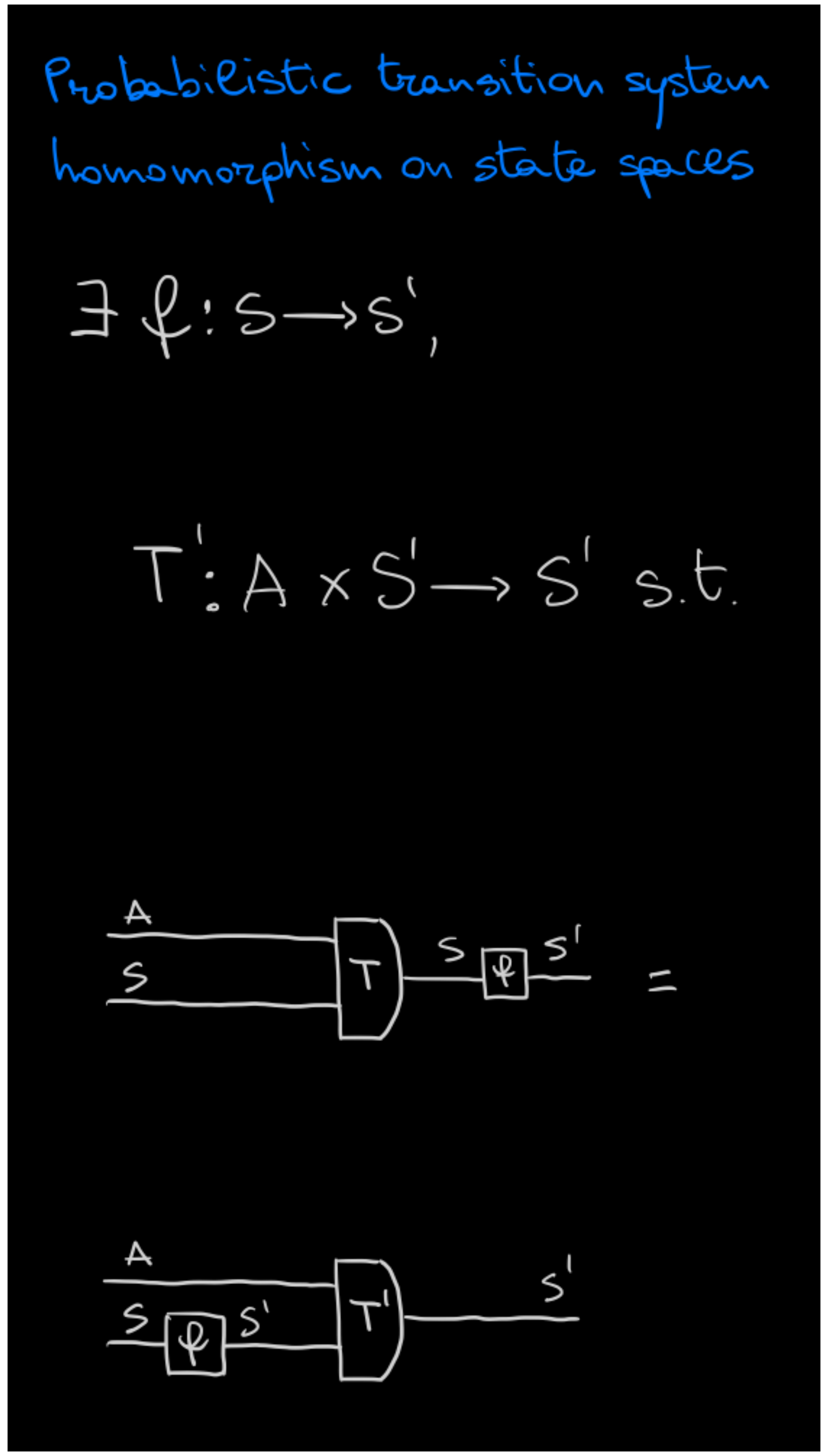

G
G
Sβ
S7
S8
Sα
S5
S6
S3
S4
S1
S2

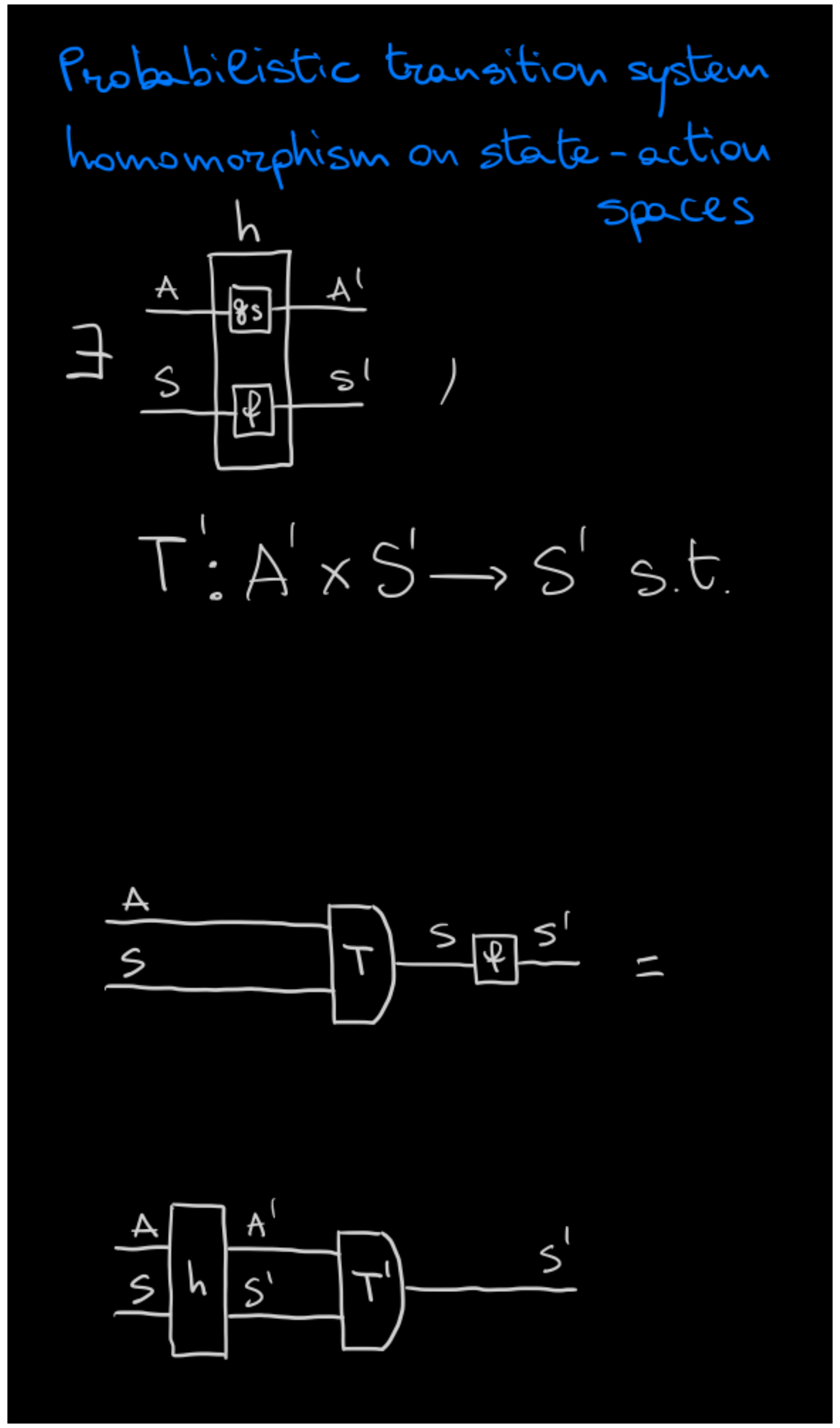
Bisimulations without rewards

Task-independent compression

- Coarse-graining state space
- Coarse-graining state-action space
- f, h: surjective
- cf.



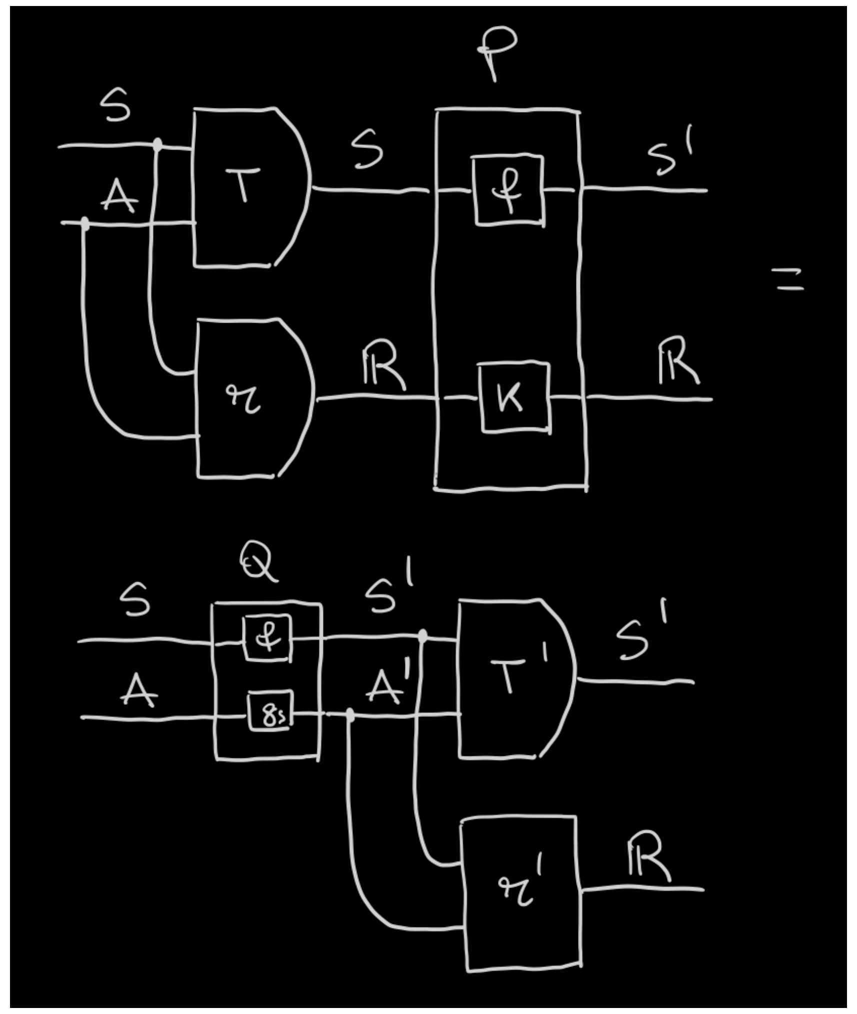


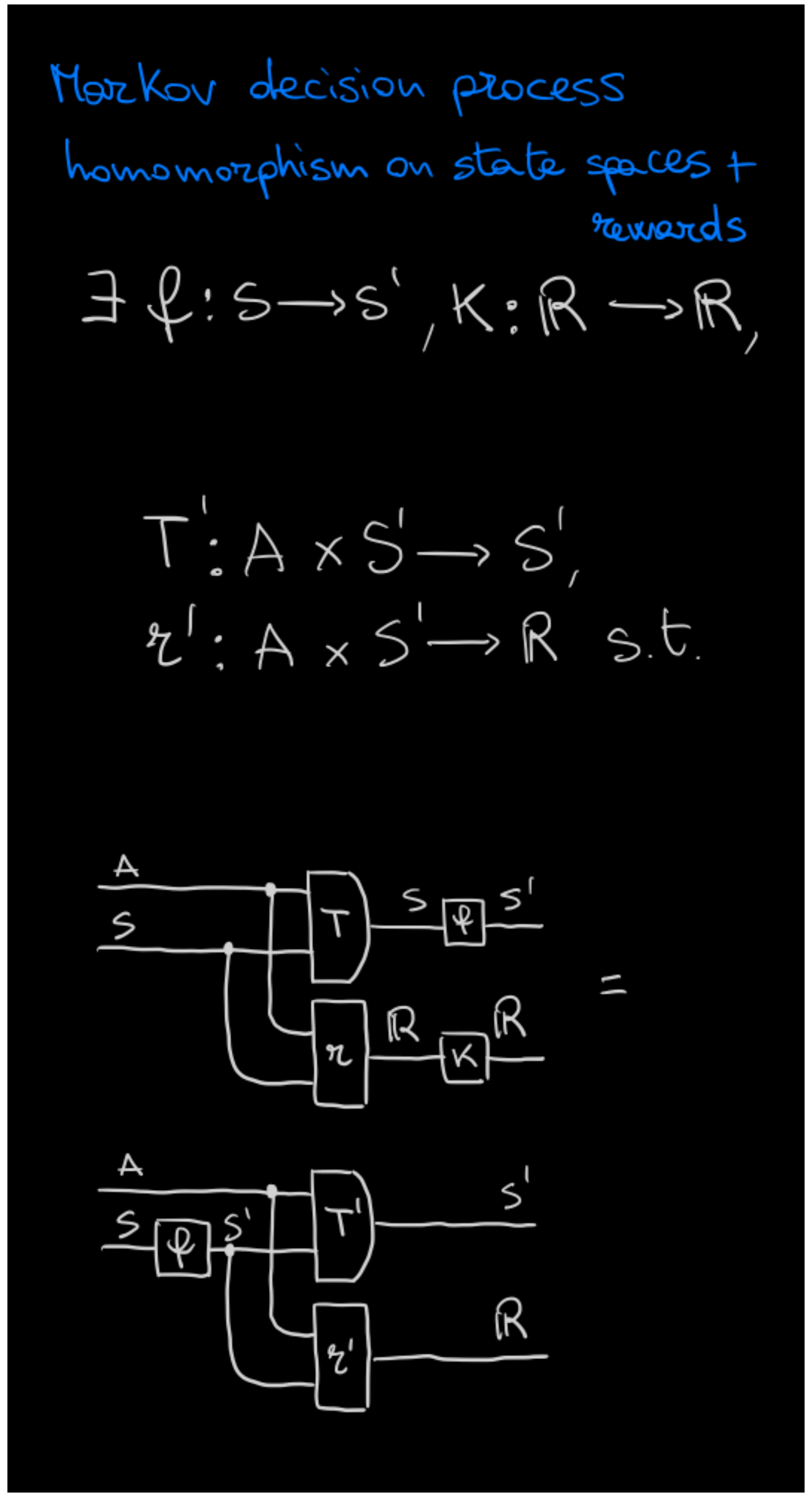


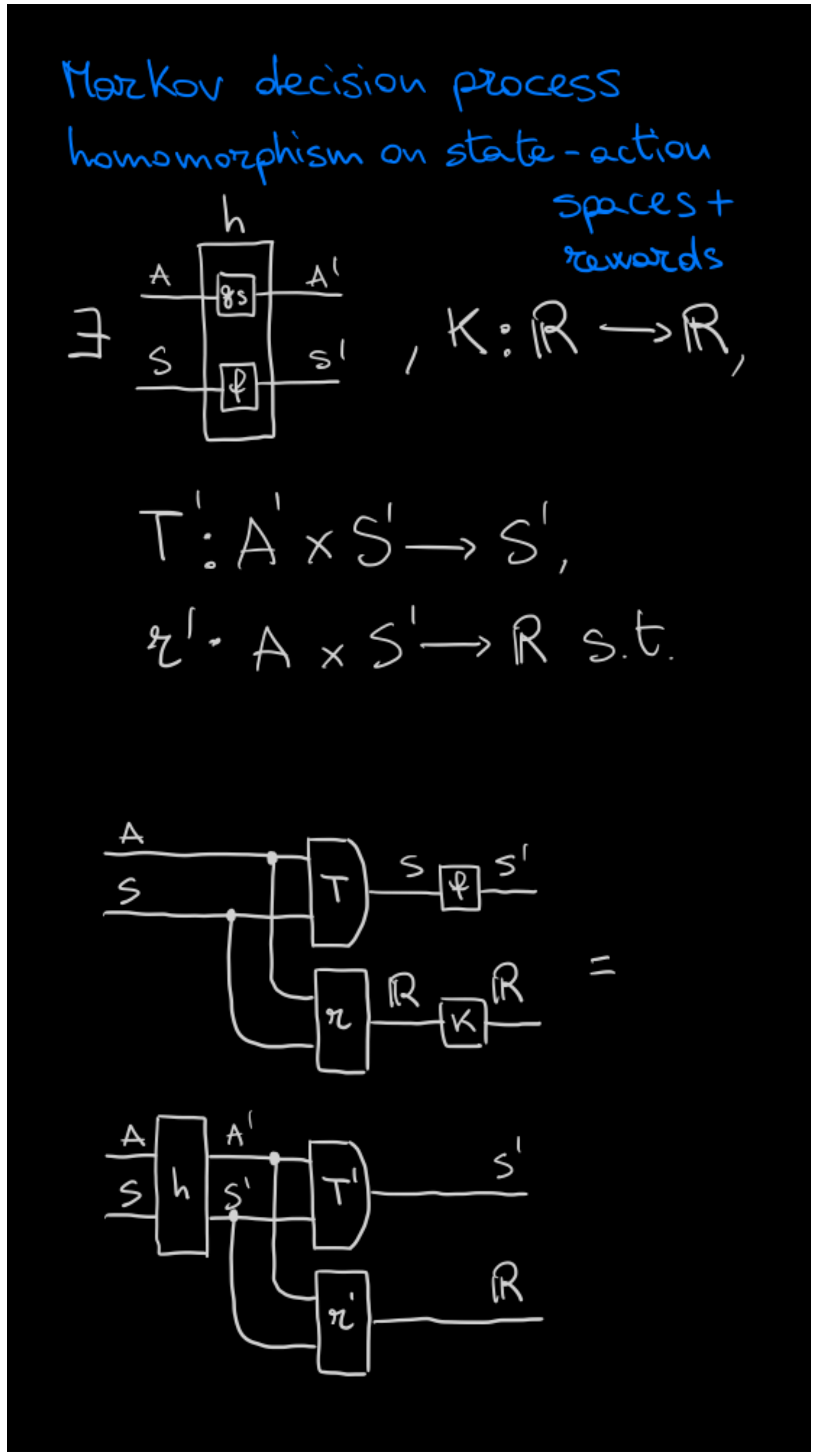
Bisimulations with rewards

Task-relevant vs task-irrelevant information?

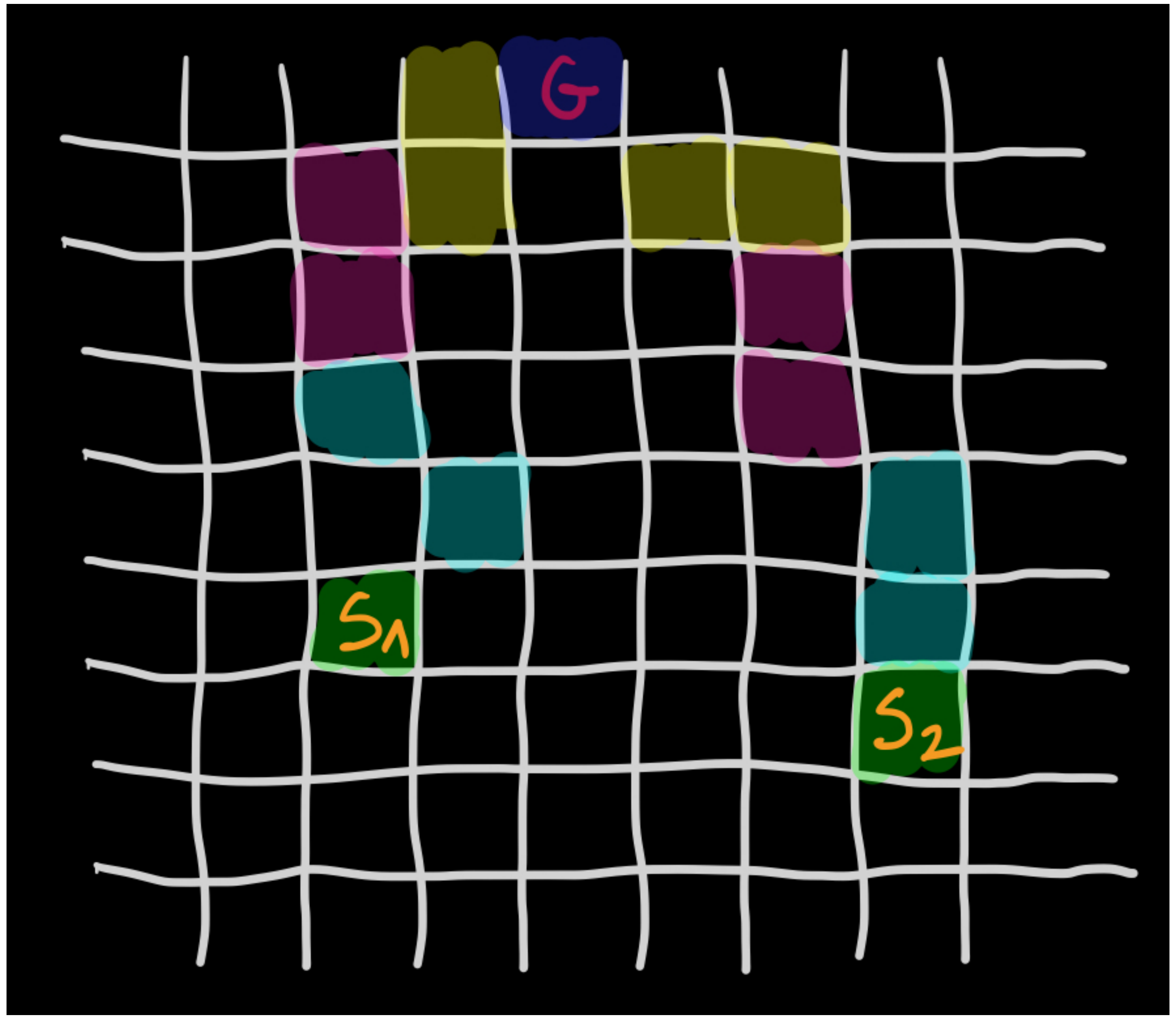
- Coarse-graining state space
- Coarse-graining state-action space
- f, h: surjective
- cf.







What about policies?



On-policy bisimulations

Policy-dependent compression

- Given a policy, π

the dynamics induced by each action. We first define:

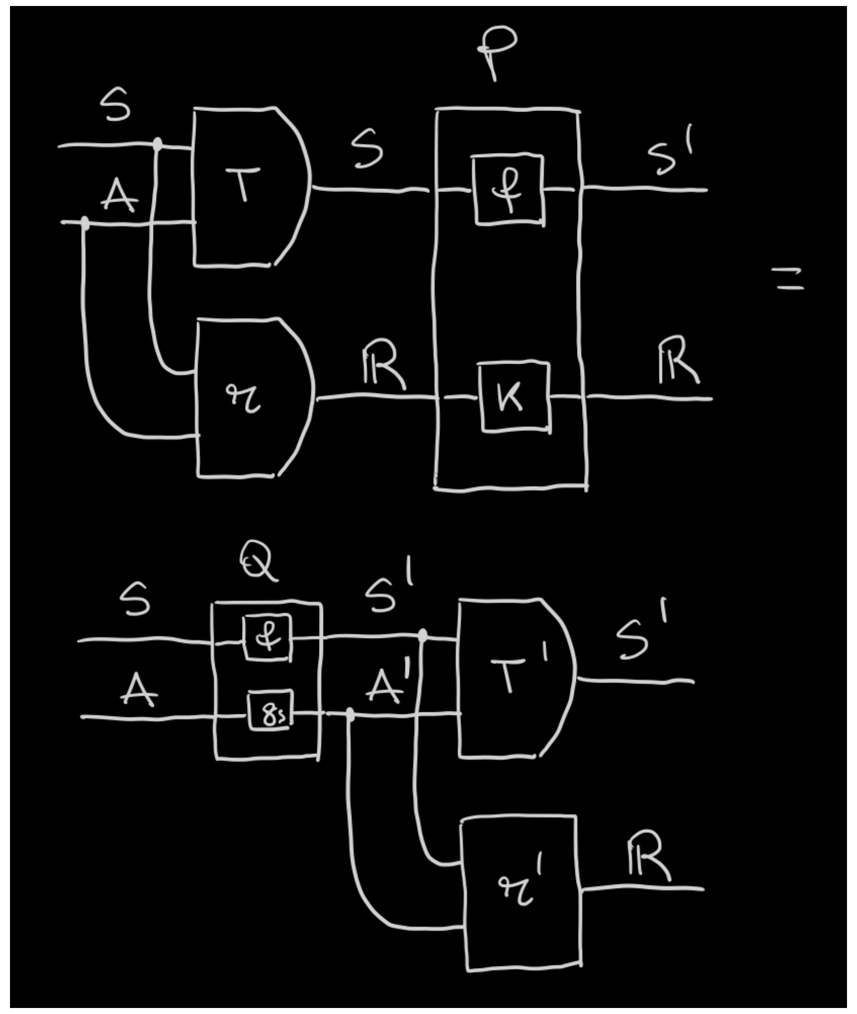
$$\mathcal{R}_s^\pi := \sum_a \pi(a|s)\mathcal{R}(s,a)$$

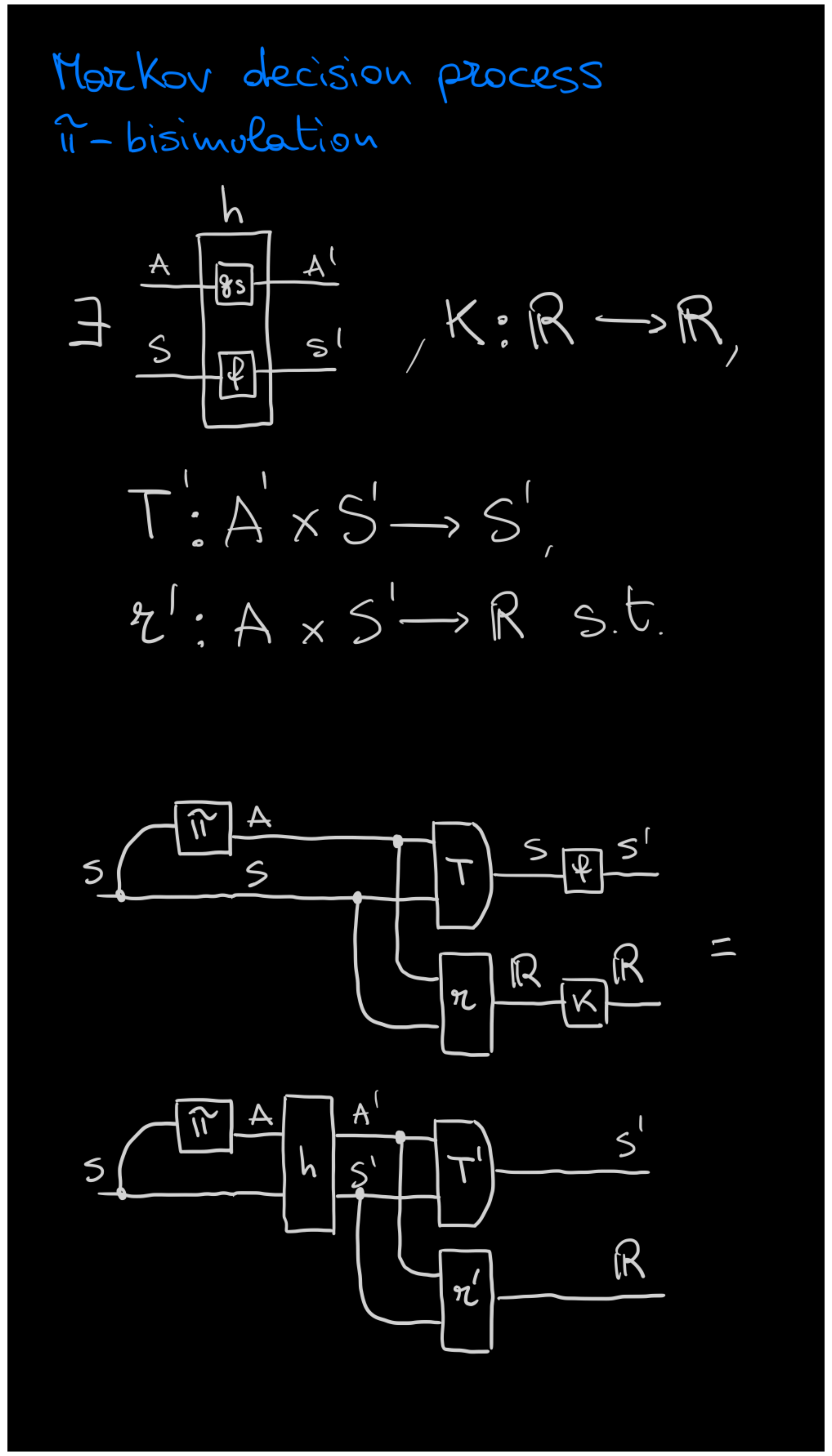
$$\forall C \in \mathcal{S}_{E^\pi}, \mathcal{P}_s^\pi(C) := \sum_a \pi(a|s) \sum_{s' \in C} P(s,a)(s')$$

Definition 3. *Given an MDP $\mathcal{M}$, an equivalence relation $E^\pi \subseteq \mathcal{S} \times \mathcal{S}$ is a π-**bisimulation relation** if whenever $(s,t) \in E^\pi$ the following properties hold:*

1. $\mathcal{R}_s^\pi = \mathcal{R}_t^\pi$
2. $\forall C \in \mathcal{S}_{E^\pi}, \mathcal{P}_s^\pi(C) = P_t^\pi(C)$

- cf.

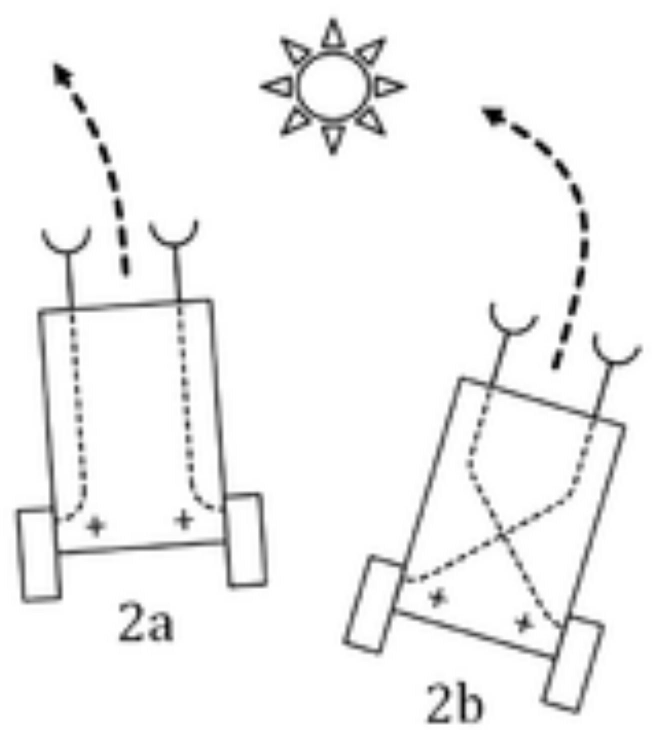


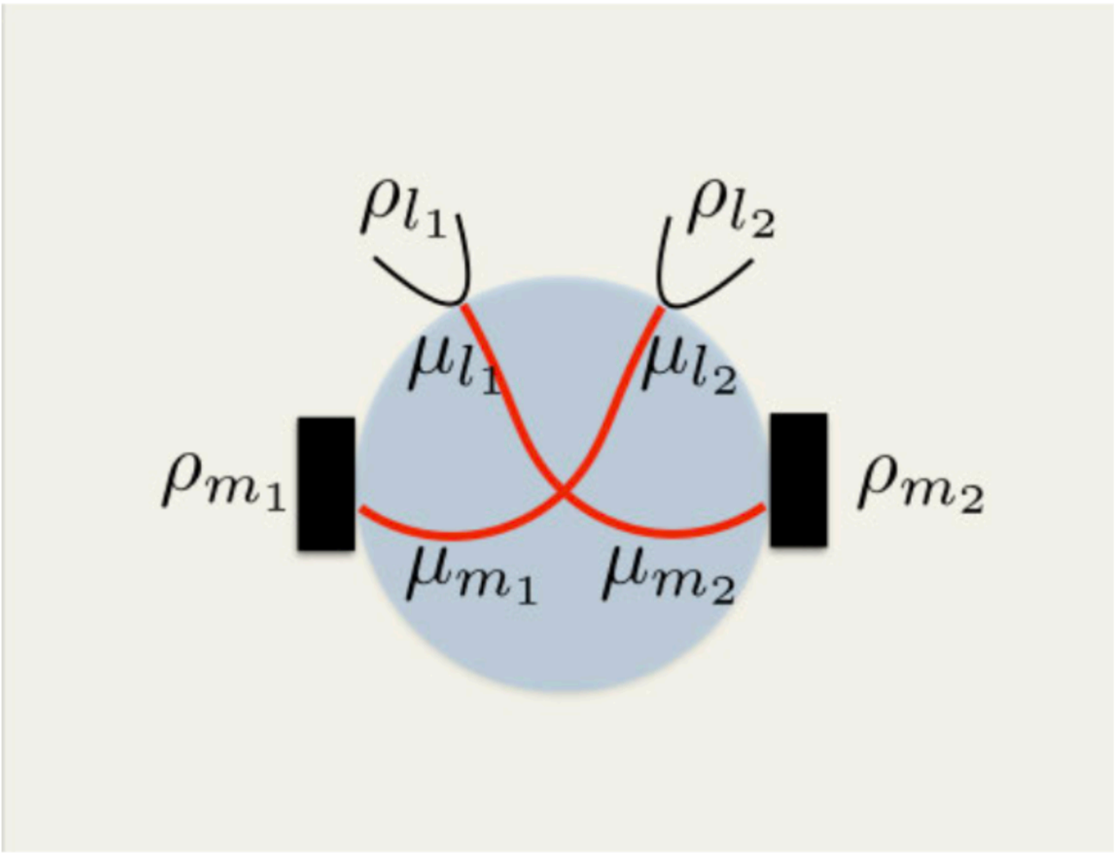


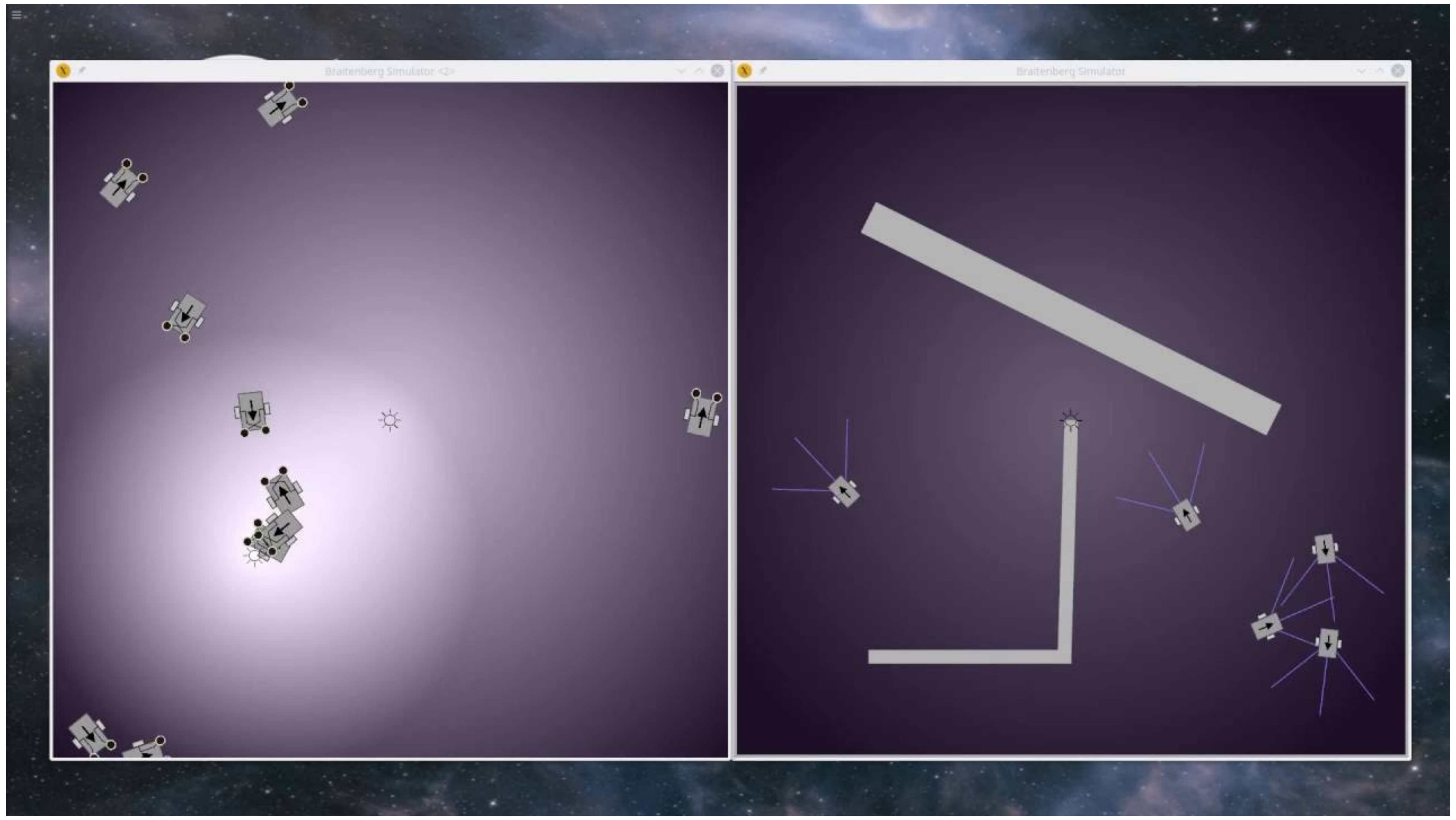
Braitenberg vehicles

And their beliefs (tentative)

- Taxis in terms of an MDP
- Question: can Braitenberg vehicles be interpreted as a bisimulation of an MDP?
- Structure:
  - Reward: chemical/light/... concentration
  - Transitions: navigation in space



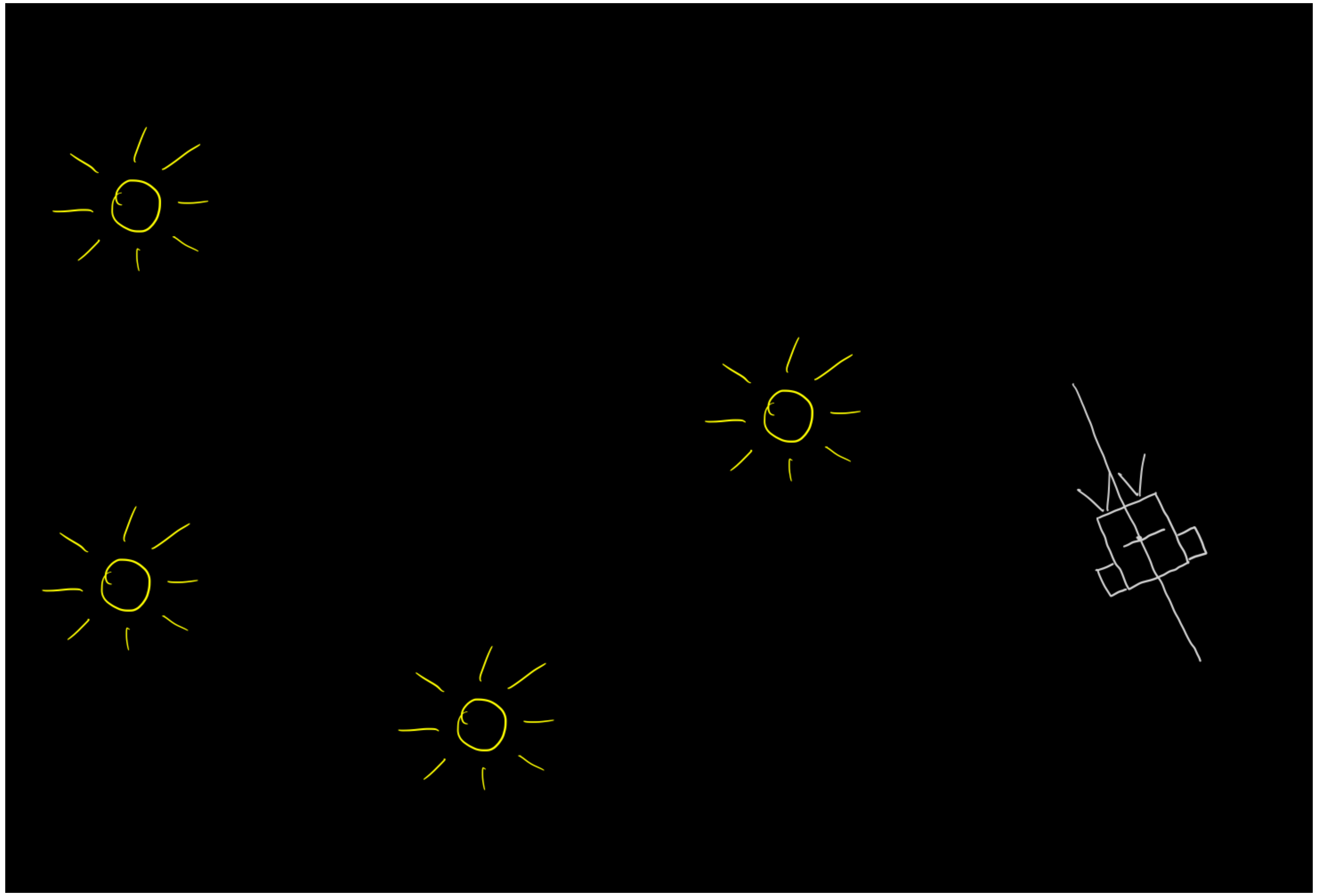




Simplified vehicles

Version 1

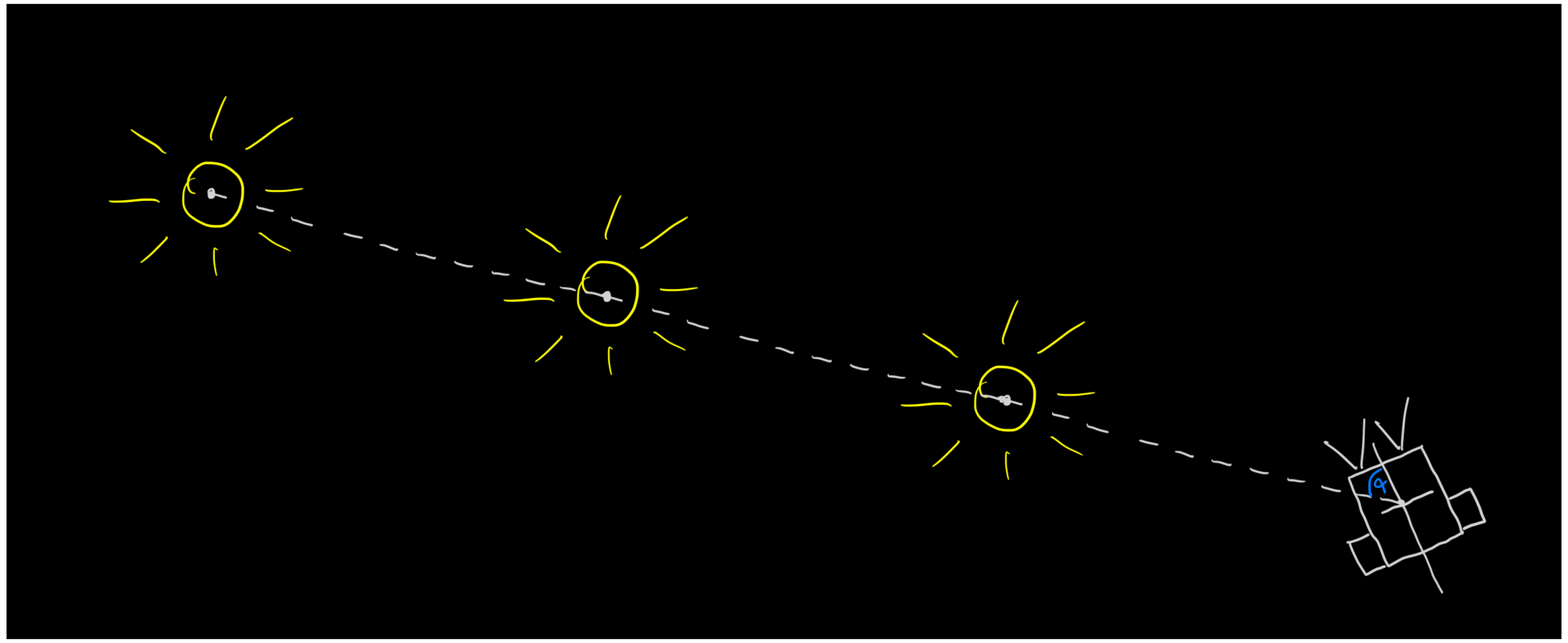
- Only gets one bit of sensory information per sensor (light/no light, chemical/no chemical)
- Only emits one bit of motor information per motor (full speed/no move)
- Cannot distinguish angle or distance to source



Simplified vehicles

Version 2

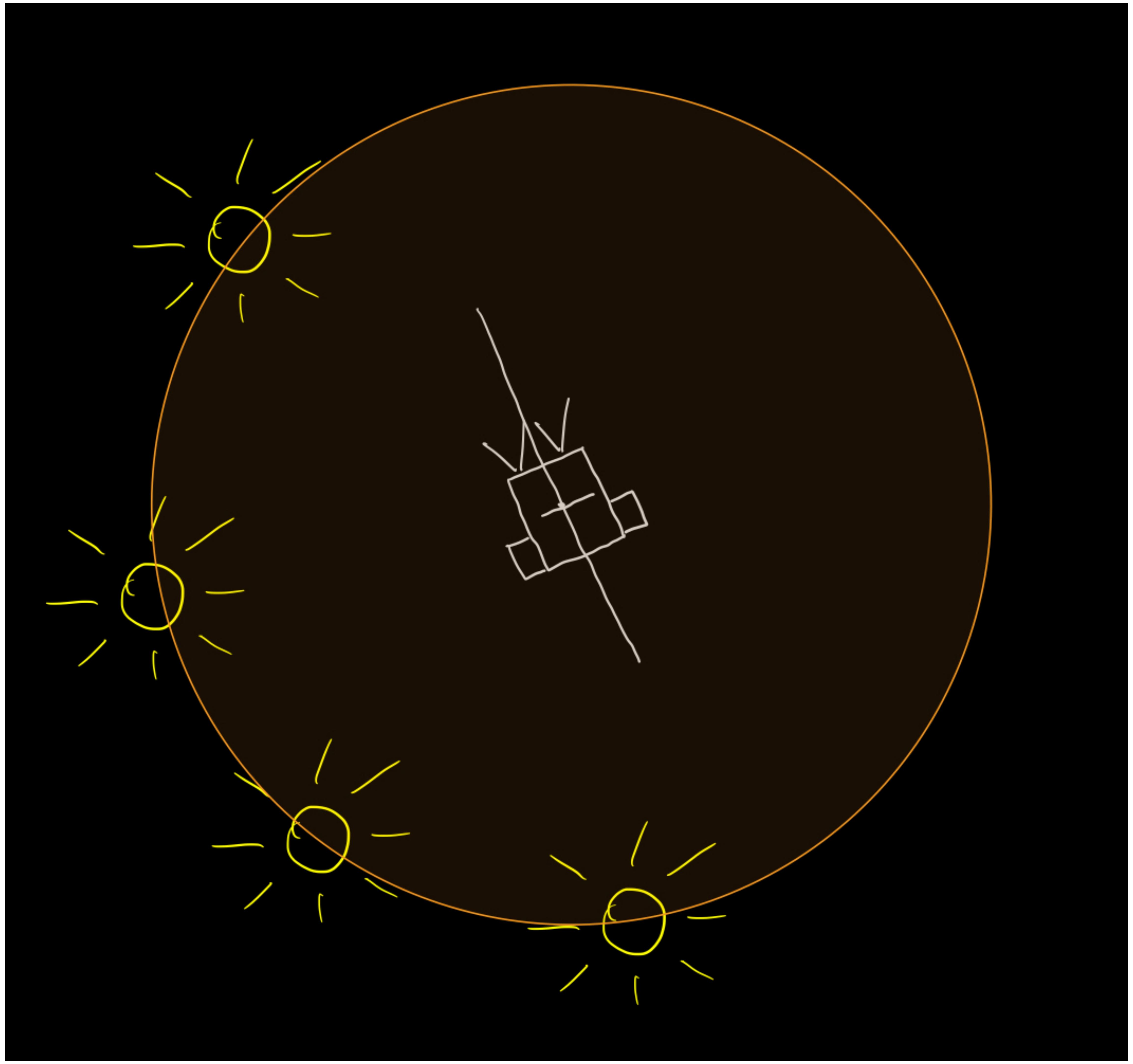
- Only emits one bit of motor information per motor (full speed/no move)
- Cannot distinguish distance to source



Simplified vehicles

Version 3

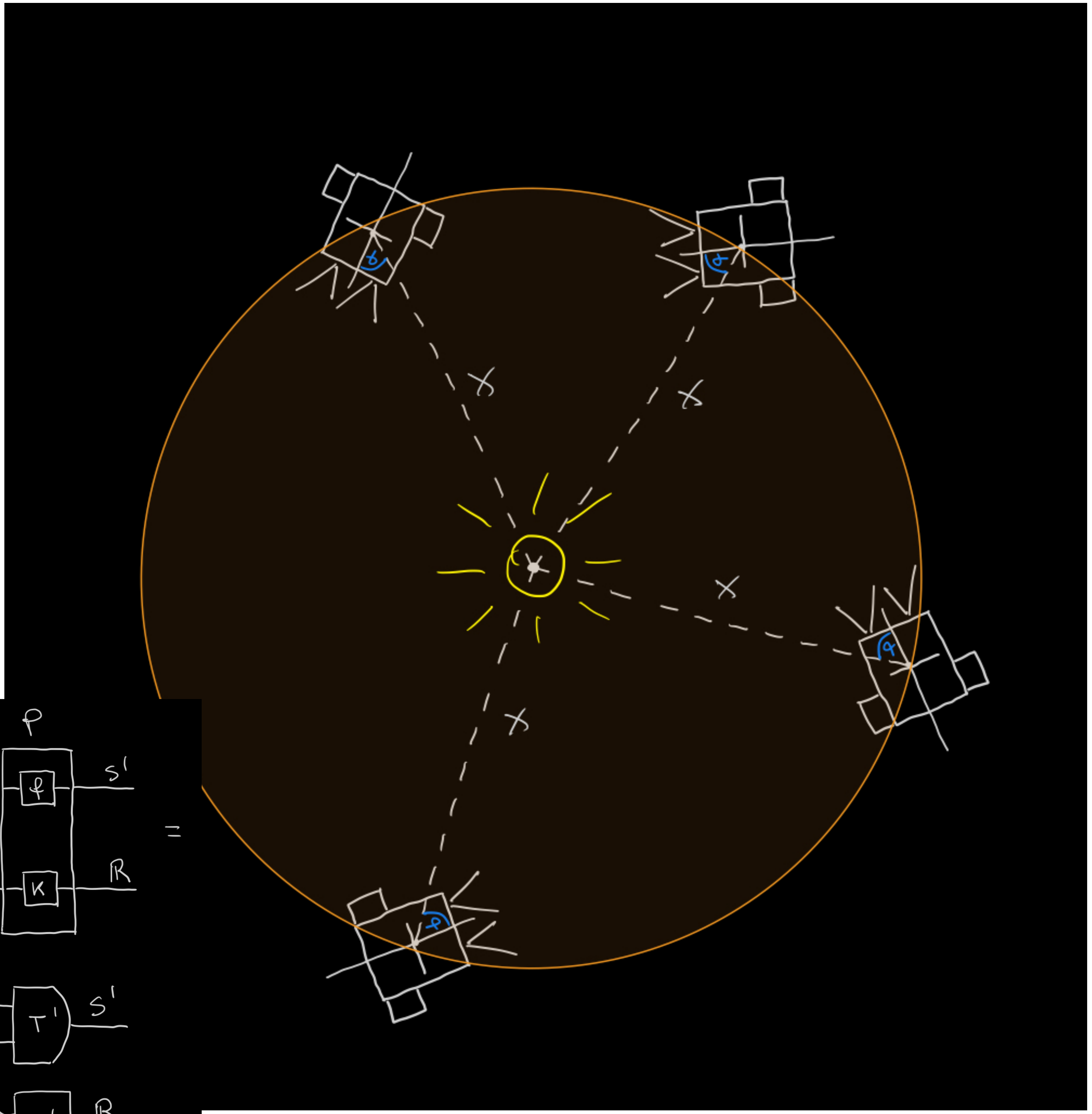
- Only gets one bit of sensory information per sensor (light/no light, chemical/no chemical)
- Cannot distinguish angle to source

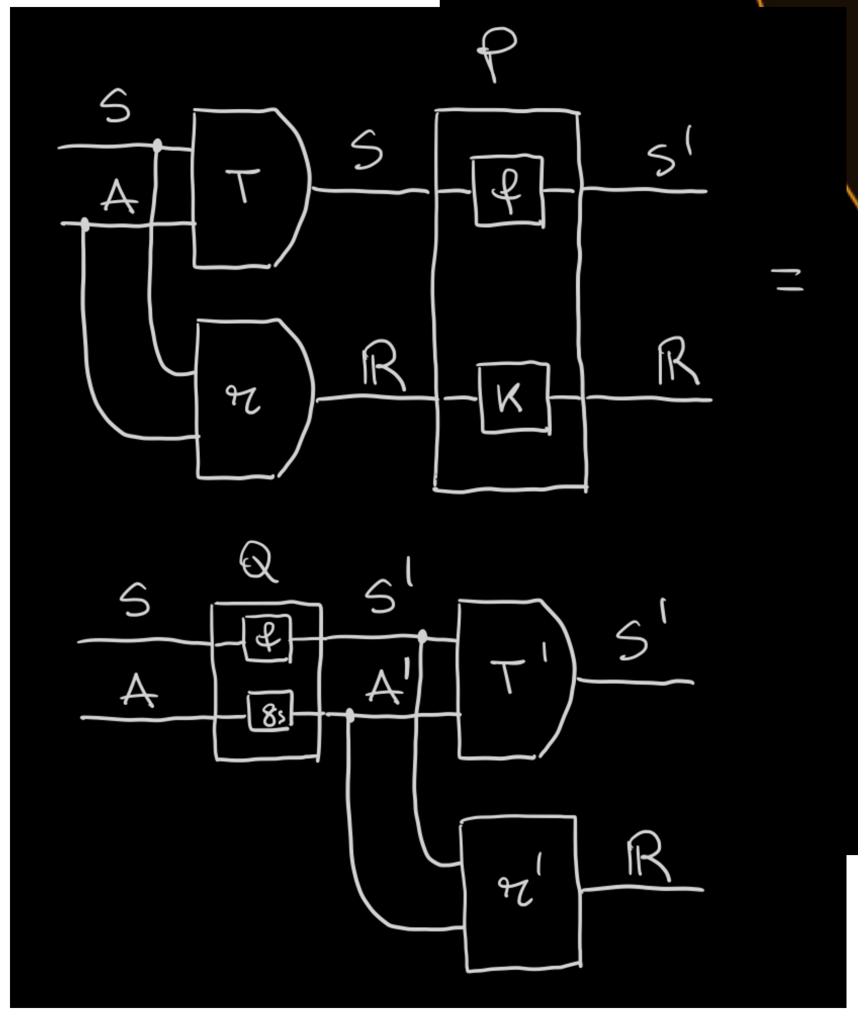


Standard vehicles

Properties

- Given same distance to source
- ...and same angle between front-facing direction and direction to source
- ...there is an invariance to rotations around sources
- WIP - bisimulation equivalence with
  - Distance to source ~ reward
  - State-space coarse-graining (states: pairs of distance and angle)
  - Actions are the same





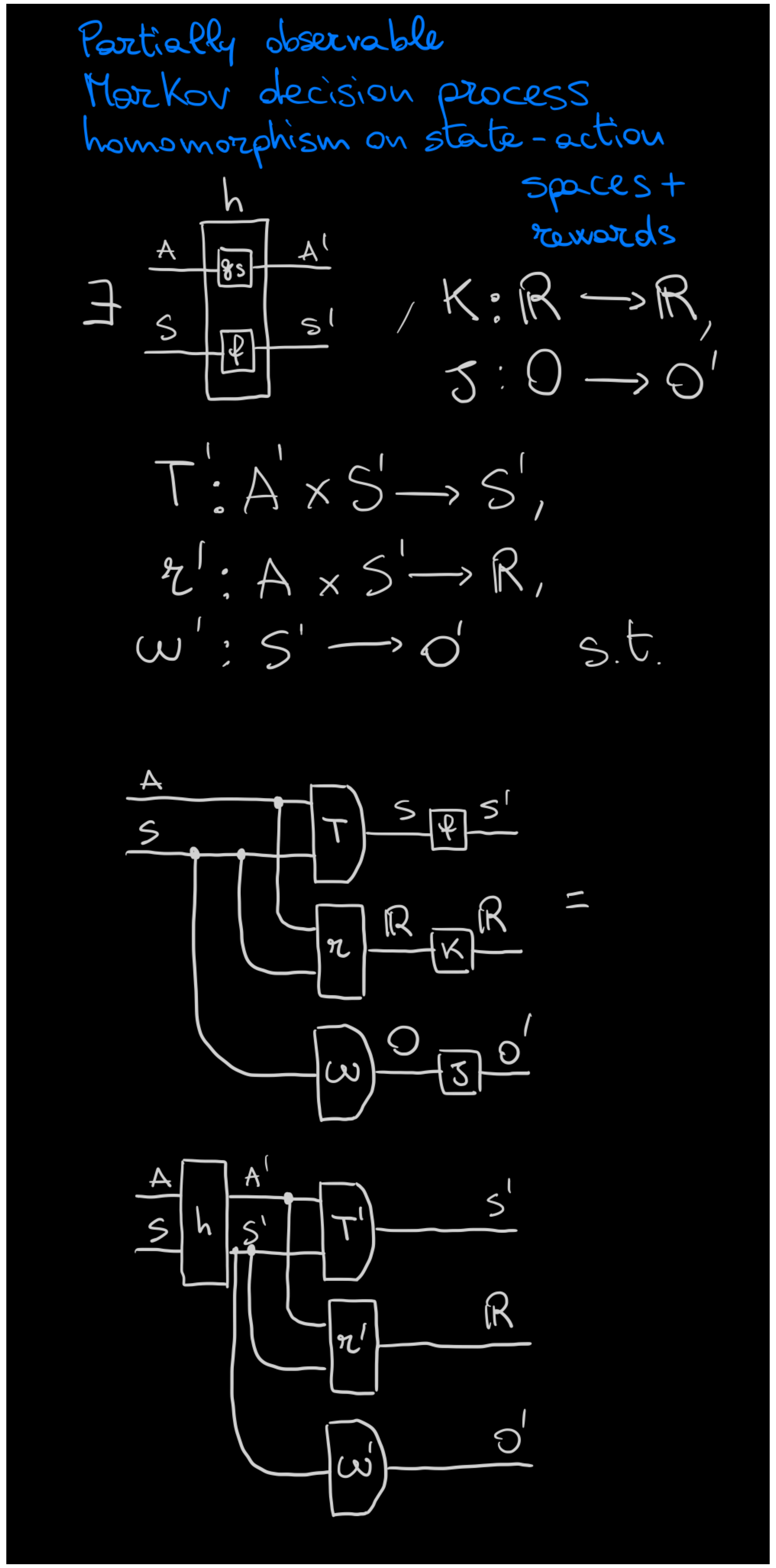
Discussion points

- What about partially observable systems?
- Why the simplest beliefs?
- What about continuous-time systems?
- Just old ideas (lumpability, state aggregation, dynamical consistency, epsilon machines, etc.)?
- Relations to what Nathaniel presented?

- We have some defs
- Doesn't work for agents that "can do more"
- We have some defs
- Sure, bisimulations are also old! Milner was working on similar things in the '70s, new applications for old ideas?
- Probably, still unclear

Summary

- Started from agents “à la Braitenberg”: simple internal structure but complex behaviour
- Interested in understanding what beliefs/goals can be attributed to these agents
- Formulated problems as MDPs to get a cheap notion of goals (reward/value)
- Systematically looked at compressions of MDPs, going through some examples
- Conjectured ways to look at Braitenberg vehicles’ beliefs



Or build belief MDP and apply previous ideas

Implementations in ML

- Task relevant vs. task irrelevant information
- Approximations with various pseudo-metrics
- Theorems to show that these pseudo-metrics are well-behaved