Grounding spatial perception:
A Sensorimotor Contingencies perspective

Alban Laflaquière
AI Lab, SoftBank Robotics Europe
Paris, France

LIRIS - 14th January 2020, Lyon

SoftBank
Robotics

SoftBank Robotics Europe:

Worldwide leader in humanoid robots

AI Lab

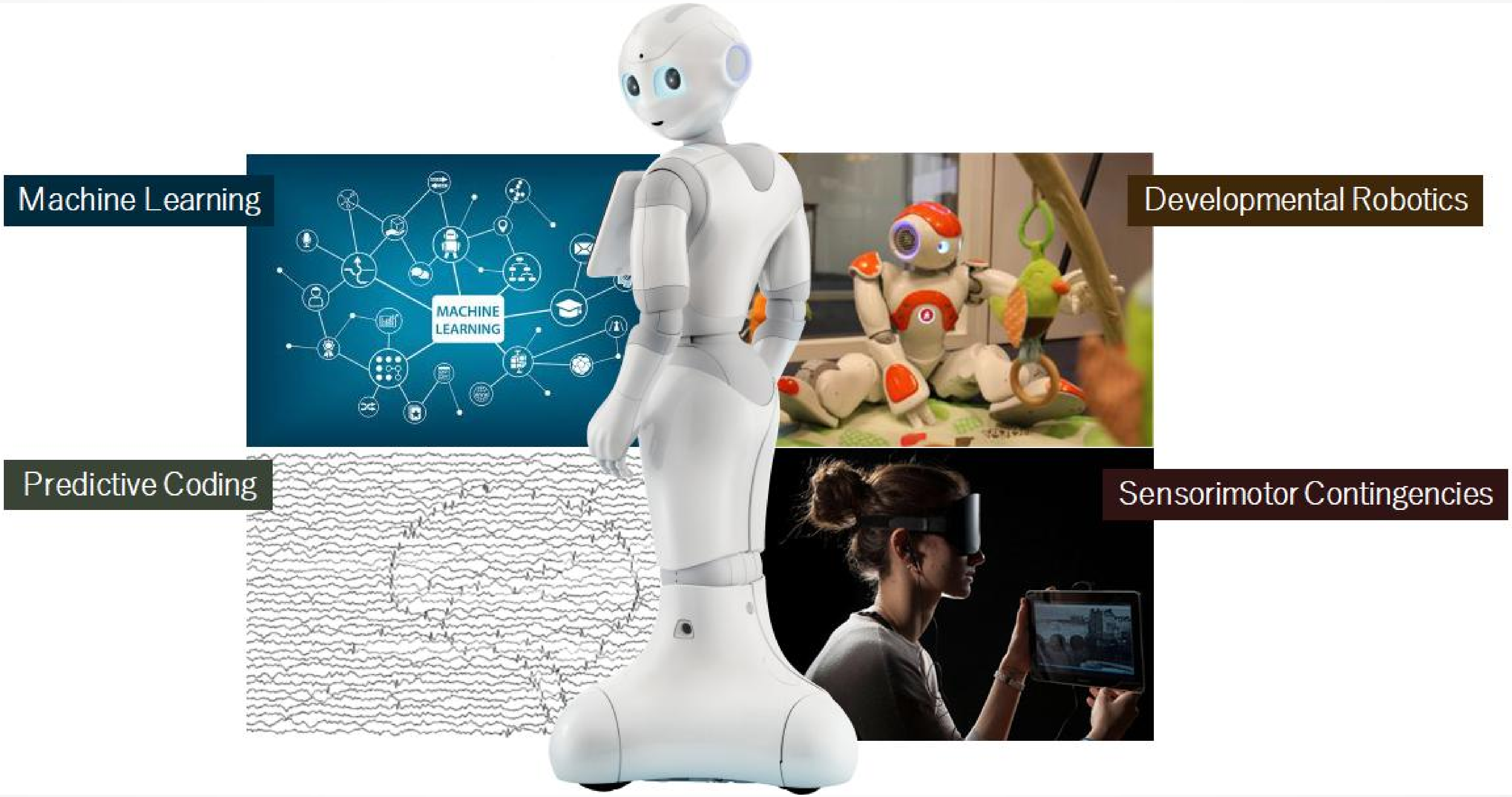
AI Lab:

Fundamental research unit in
Developmental Robotics & Machine Learning

Do we have the correct perspective on perception?

VS

How can a robot autonomously learn to perceive and interact with its environment?



Machine Learning

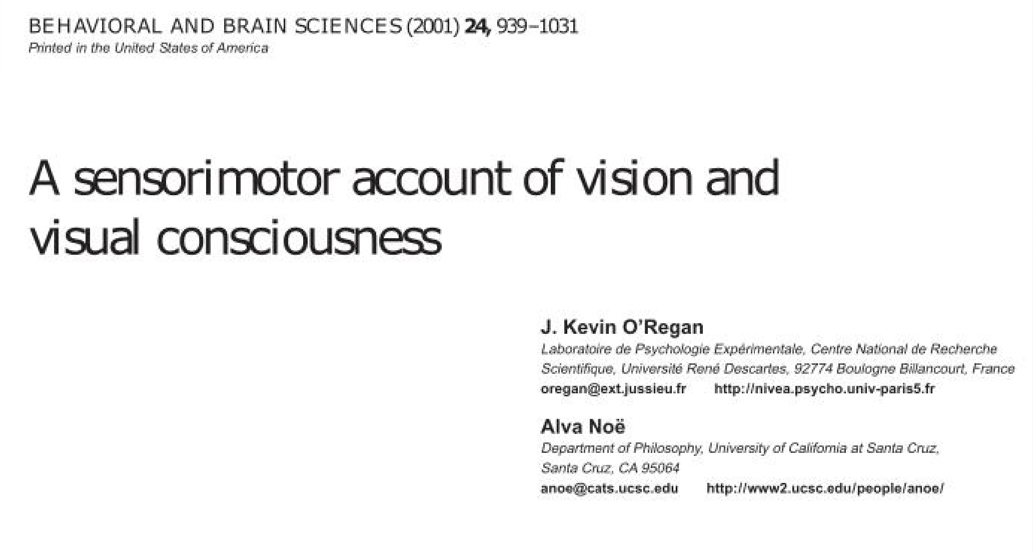
Developmental Robotics

Predictive Coding

Sensorimotor Contingencies

The Sensori-Motor Contingencies Theory

- Perception = “***mode of exploration of the world, mediated by the knowledge of sensorimotor contingencies***”
(SM contingencies = rules governing the sensory changes produced by motor actions)
- Different perceptions = different structures in the contingencies
- Brain: identify contingencies and use them for thoughts and planning

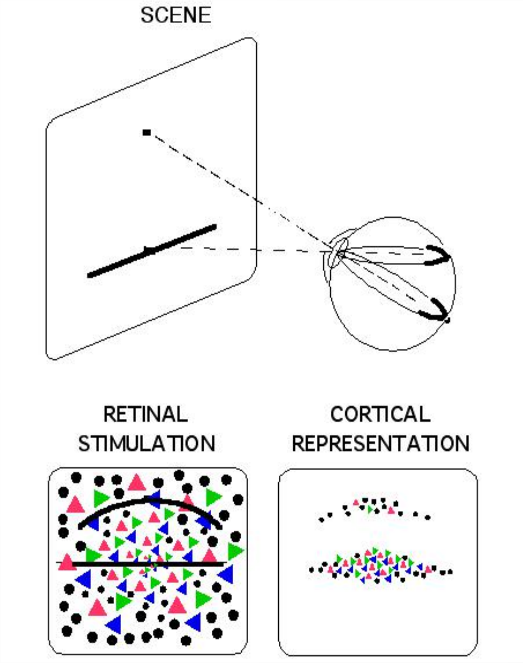


BEHAVIORAL AND BRAIN SCIENCES (2001) **24**, 939–1031
Printed in the United States of America

A sensorimotor account of vision and visual consciousness

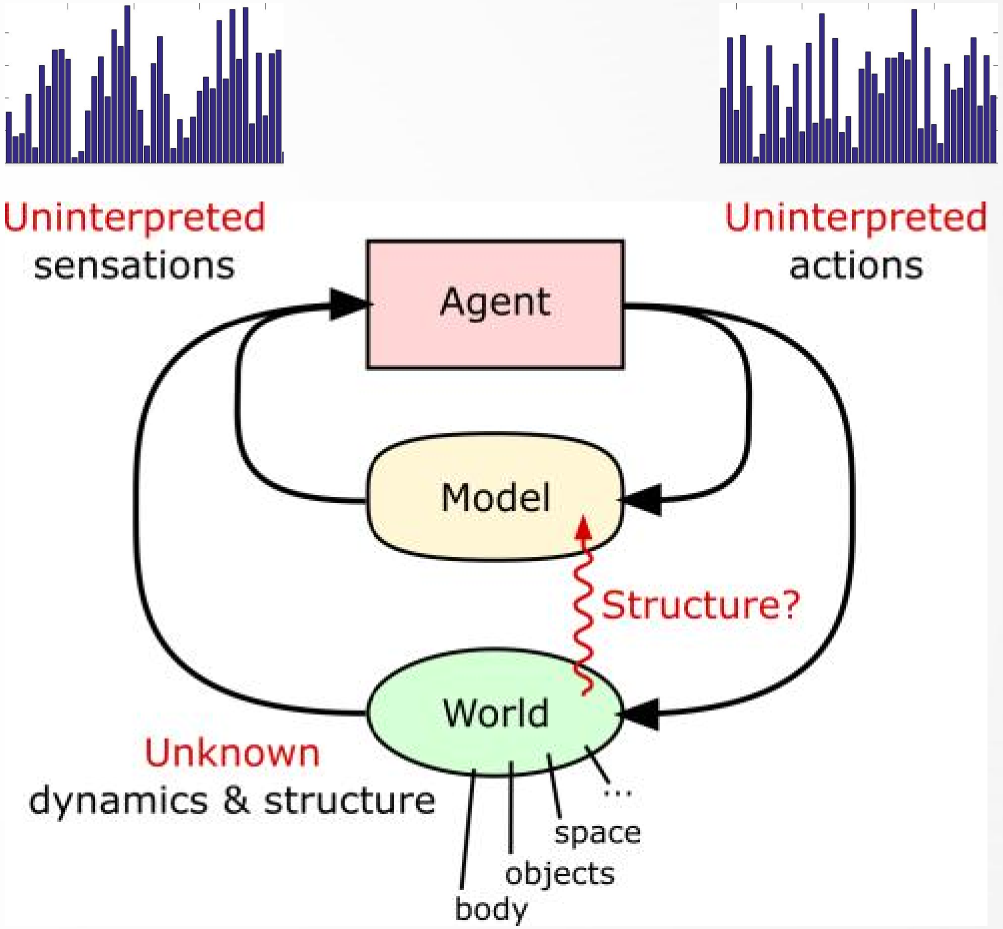
J. Kevin O'Regan
Laboratoire de Psychologie Expérimentale, Centre National de Recherche Scientifique, Université René Descartes, 92774 Boulogne Billancourt, France
oregan@ext.jussieu.fr   http://nivea.psycho.univ-paris5.fr

Alva Noë
Department of Philosophy, University of California at Santa Cruz, Santa Cruz, CA 95064
anoe@cats.ucsc.edu   http://www2.ucsc.edu/people/anoe/



Some lessons from the SMCT

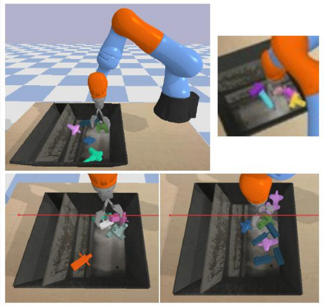
- Siuatedness
- Sensorimotor transitions as building block
- Forward model(s)
- Sensorimotor invariants
  (abstract & code-independent)

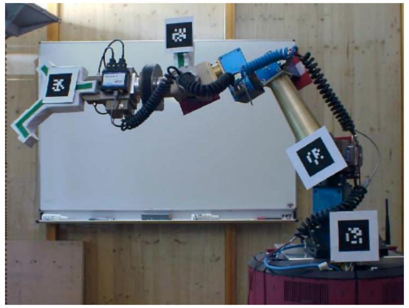


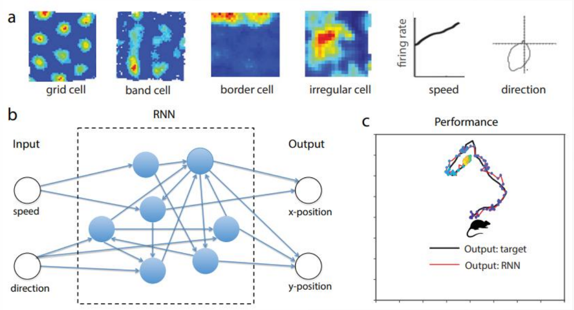
Problem: Grounding spatial perception?

Existing ways to deal with spatial tasks:

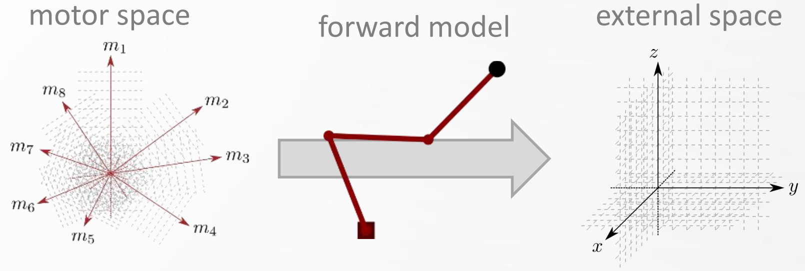
- “model-free” approaches: learn a policy or forward model without particular care for spatial structure
- “model-based” approaches: spatial structure provided a priori by humans
- dedicated approaches: strong inductive biases, supervised setting







Can spatial structure be discovered in a bottom-up fashion?
→ can an agent learn the xyz frame necessary to learn a forward model?

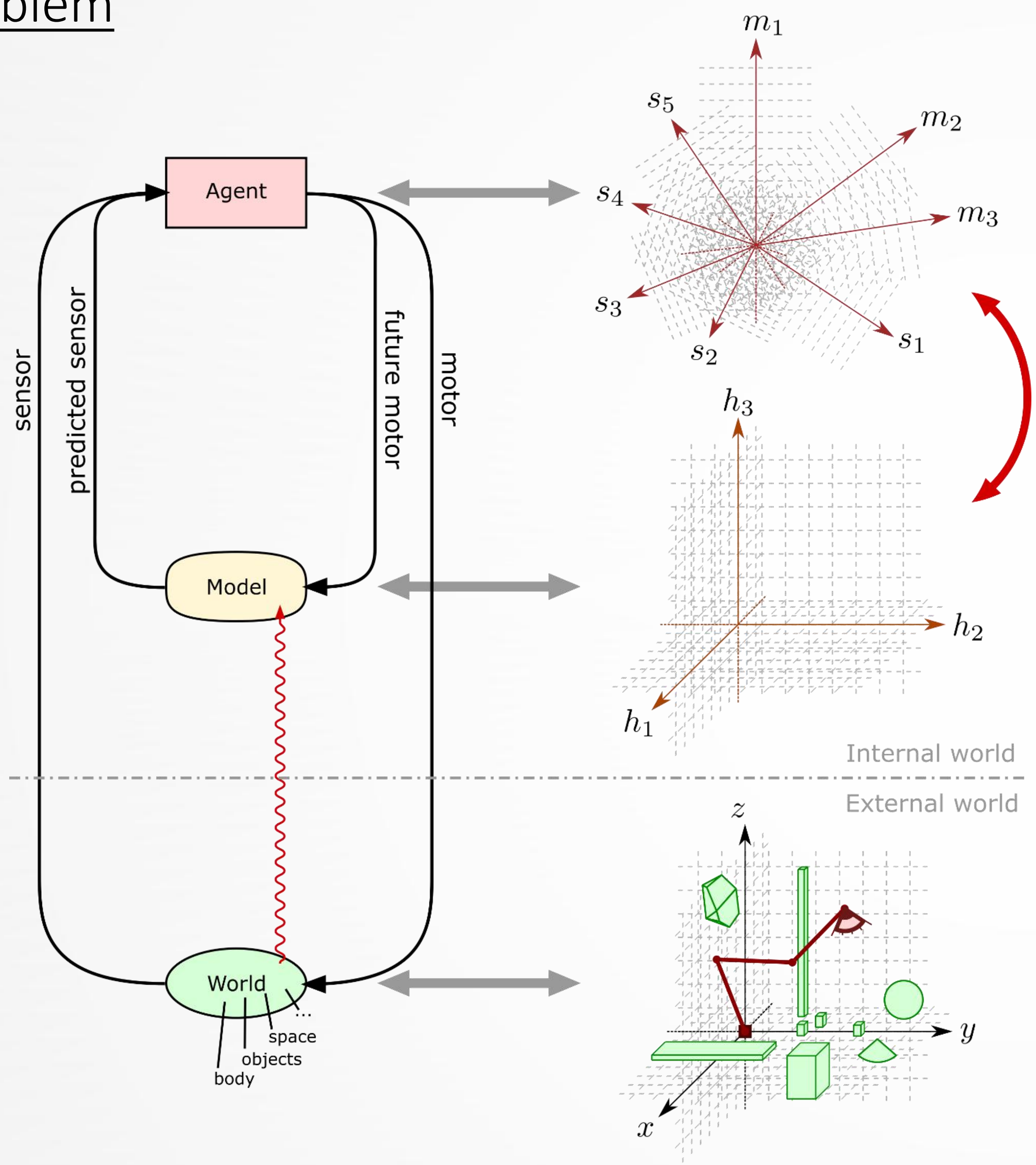

motor space

forward model

external space

“Deep reinforcement learning for vision-based robotic grasping: A simulated comparative evaluation of off-policy methods”, Quillen, 2018
“Body schema learning for robotic manipulators from visual self-perception”, Sturm, 2009
“Emergence of grid-like representations by training recurrent neural networks to perform spatial localization”, Cueva, 2018

Problem



"Internal world":

- raw sensorimotor states
- high dimension
- own metric

"External world":

- body and environment
- immersed in space
  - □ low dimension
  - □ Euclidean metric
  - □ shared by agent and environment
  - □ content-independent

→ Can the agent build an internal representation of space, with the same properties?

→ Is sensorimotor prediction a sufficient drive?

Intuition

H.Poincaré (1854-1912)

- The sensorimotor space in which the agent gathers experiences is fundamentally different from the external space in which it is embedded:

  “Thereby, the representative space, in the triple form visual, tacile and motor, is inherently different from the geometric space.”

- Sensorimotor states in themselves aren't informative about space; one need to look at the way sensations change:

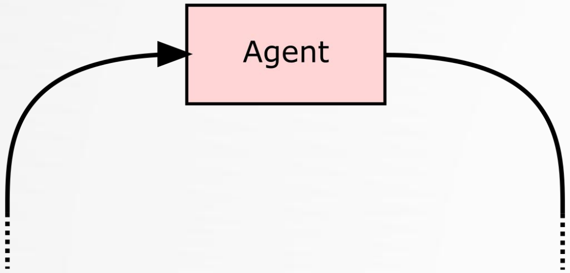
  “None of our sensations, considered separately, could have lead us to the idea of space; it comes to us by the study of the laws according to these sensations come one after antoher.”

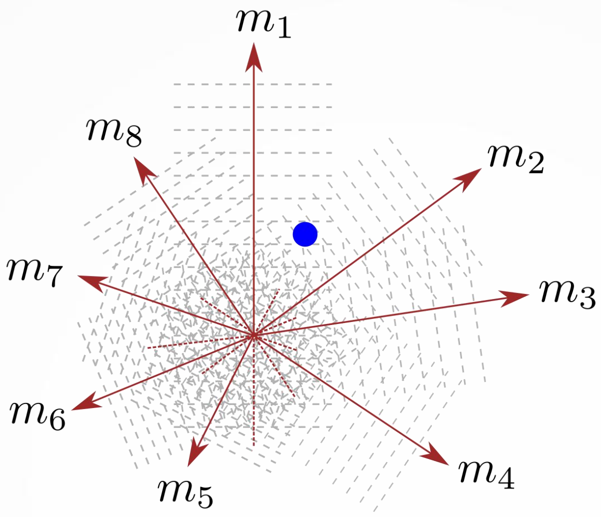
- Spatial changes (displacements) have the particular property of being compensable:

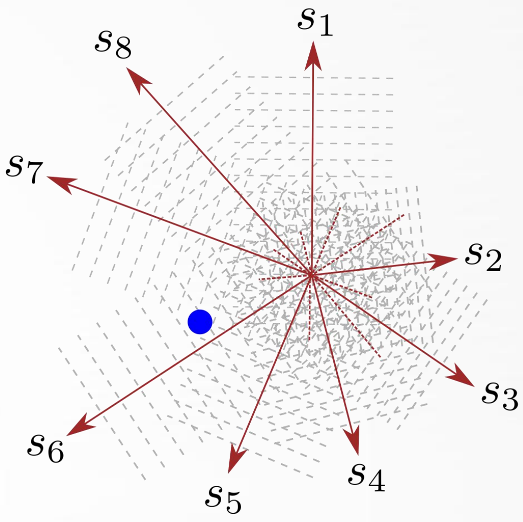
  “What characterizes a change of position [of an object], is that it can be compensated [by a motor command].”

“L’espace et la géométrie”, H.Poincaré, 1895

Sensory/motor states are not "spatial"

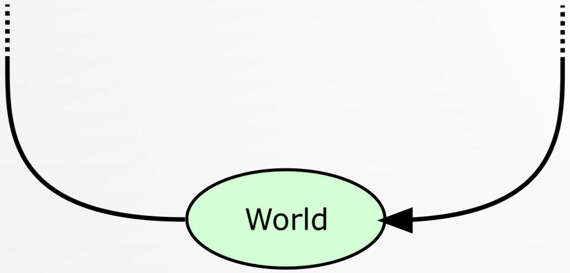


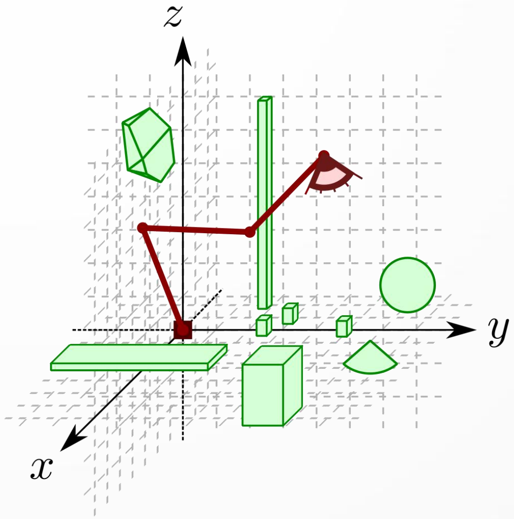




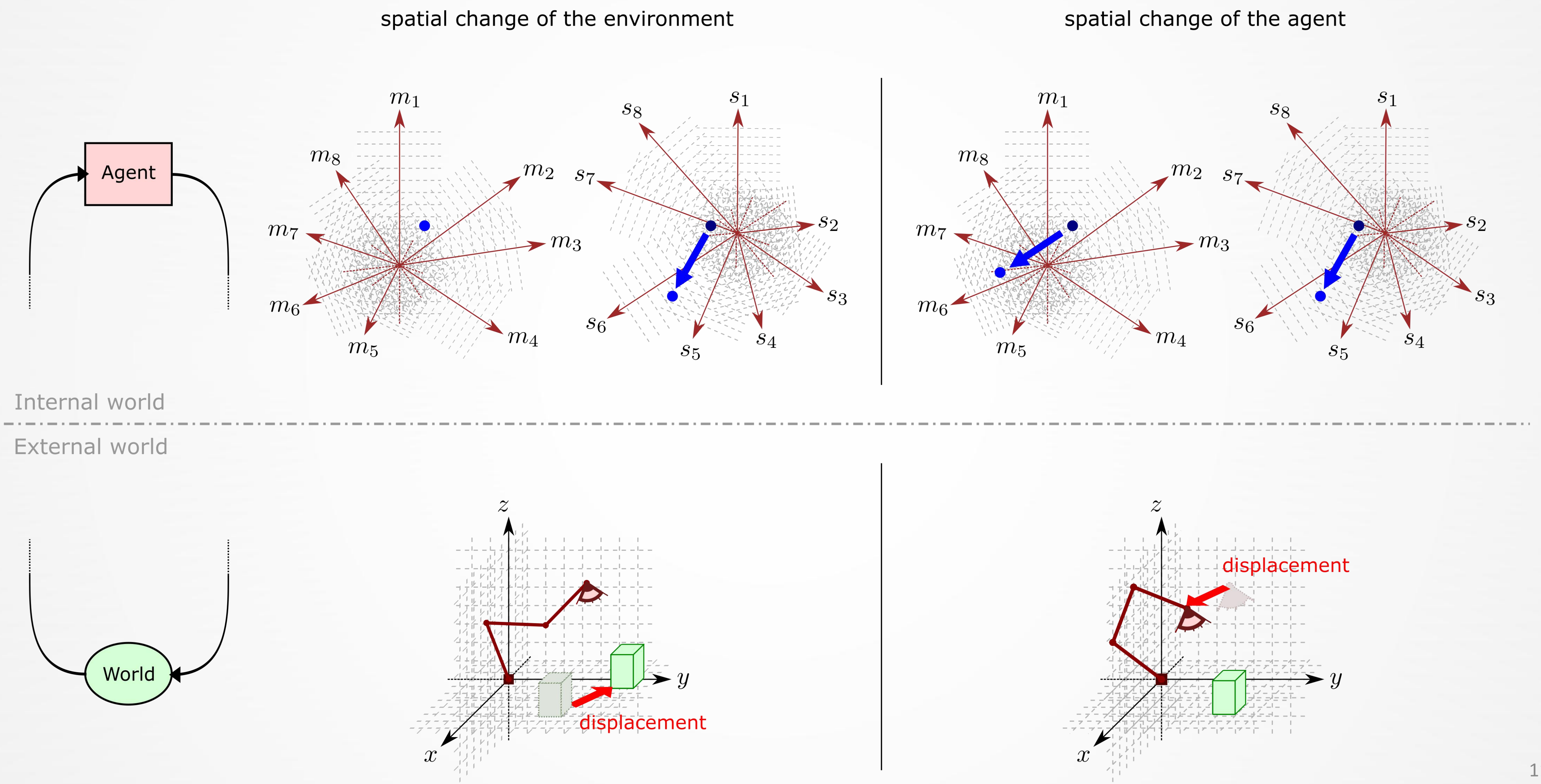
Internal world

External world





Sensory/motor changes can be "spatial"



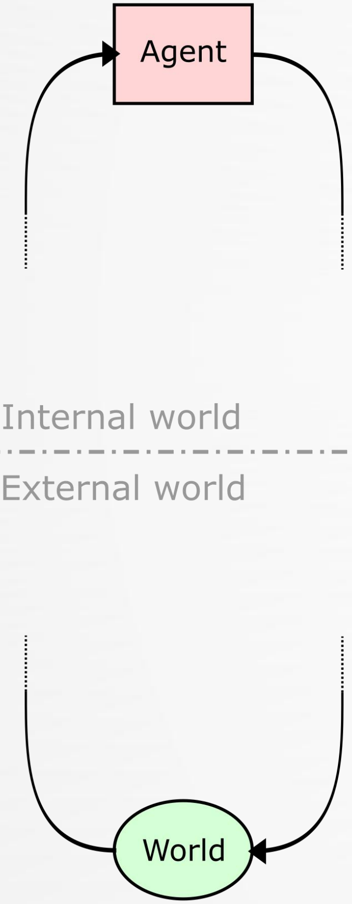
spatial change of the environment

spatial change of the agent

Internal world

External world

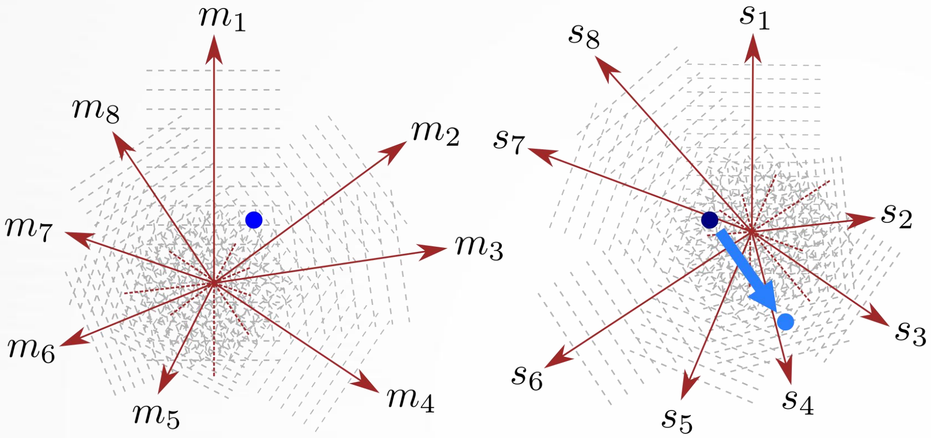
Not all sensory/motor changes are "spatial"

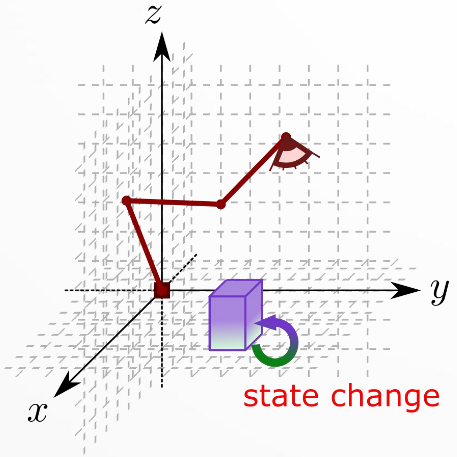


Internal world

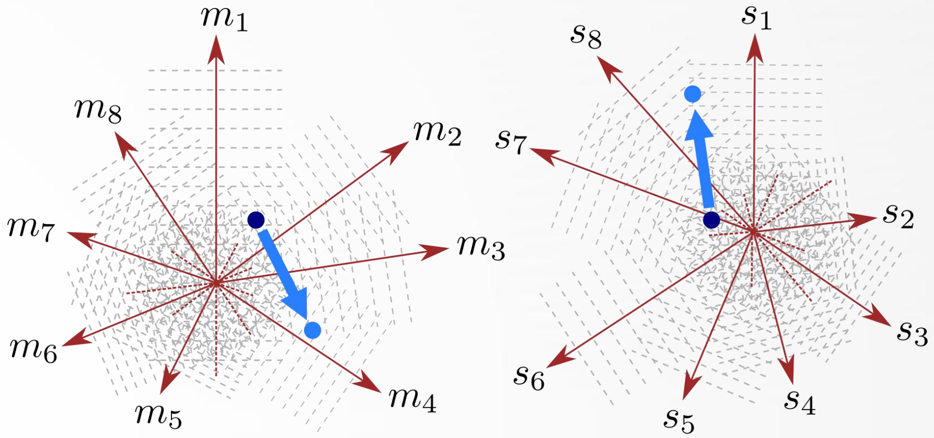
External world

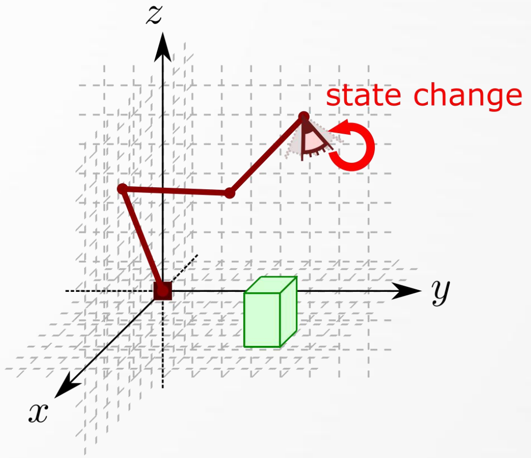
state change of the environment





state change of the agent

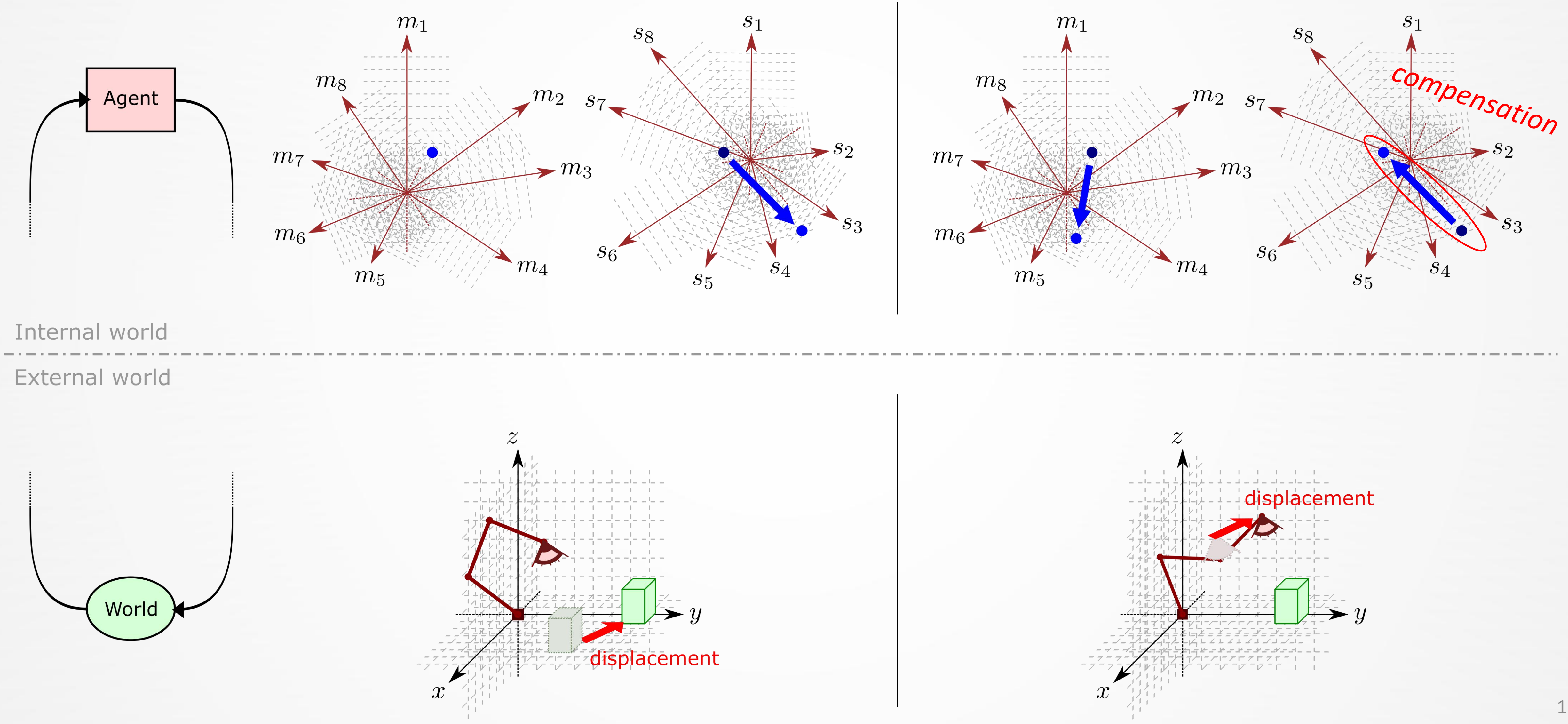




Spatial sensory/motor changes are “compensable”

spatial change of the environment

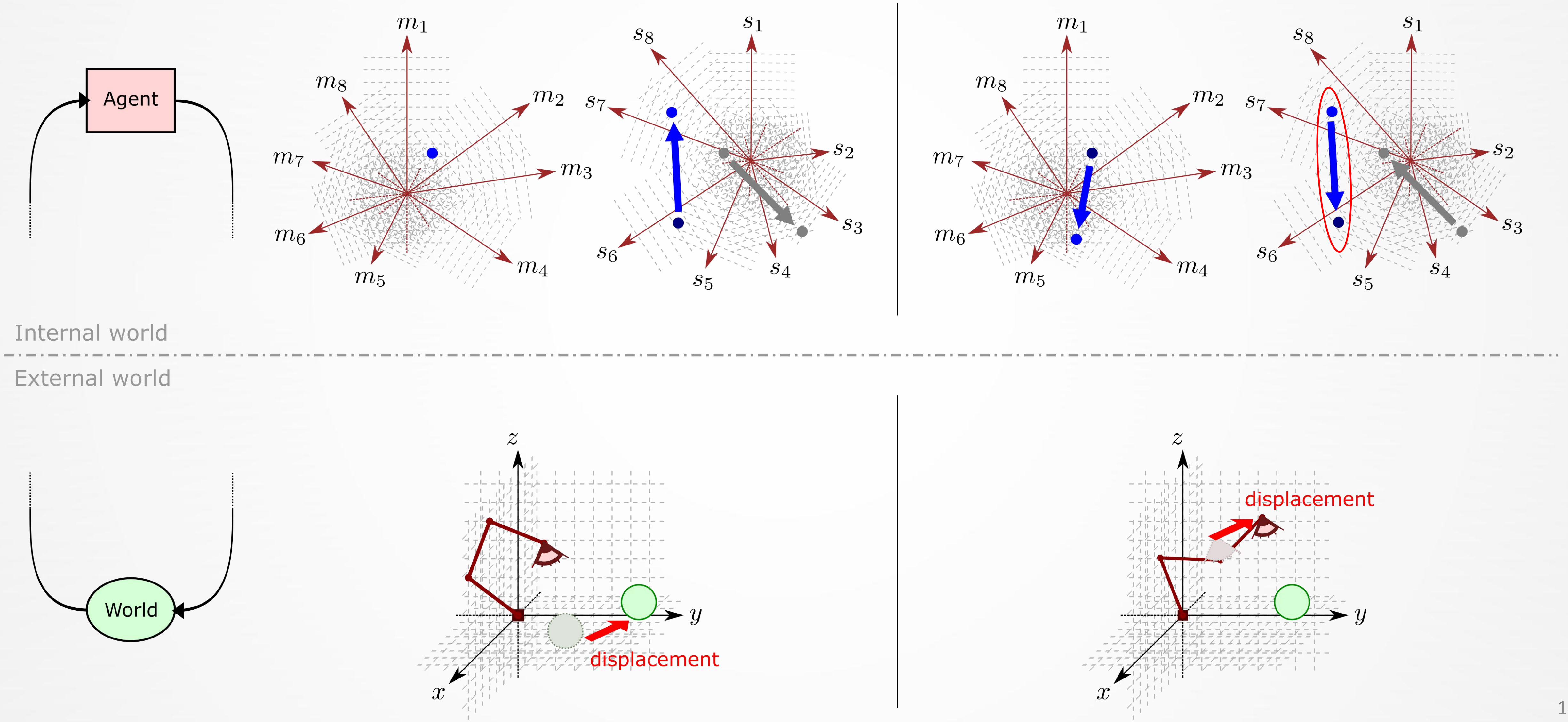
spatial change of the agent



“Compensability” is content-independent on the motor side

spatial change of the environment

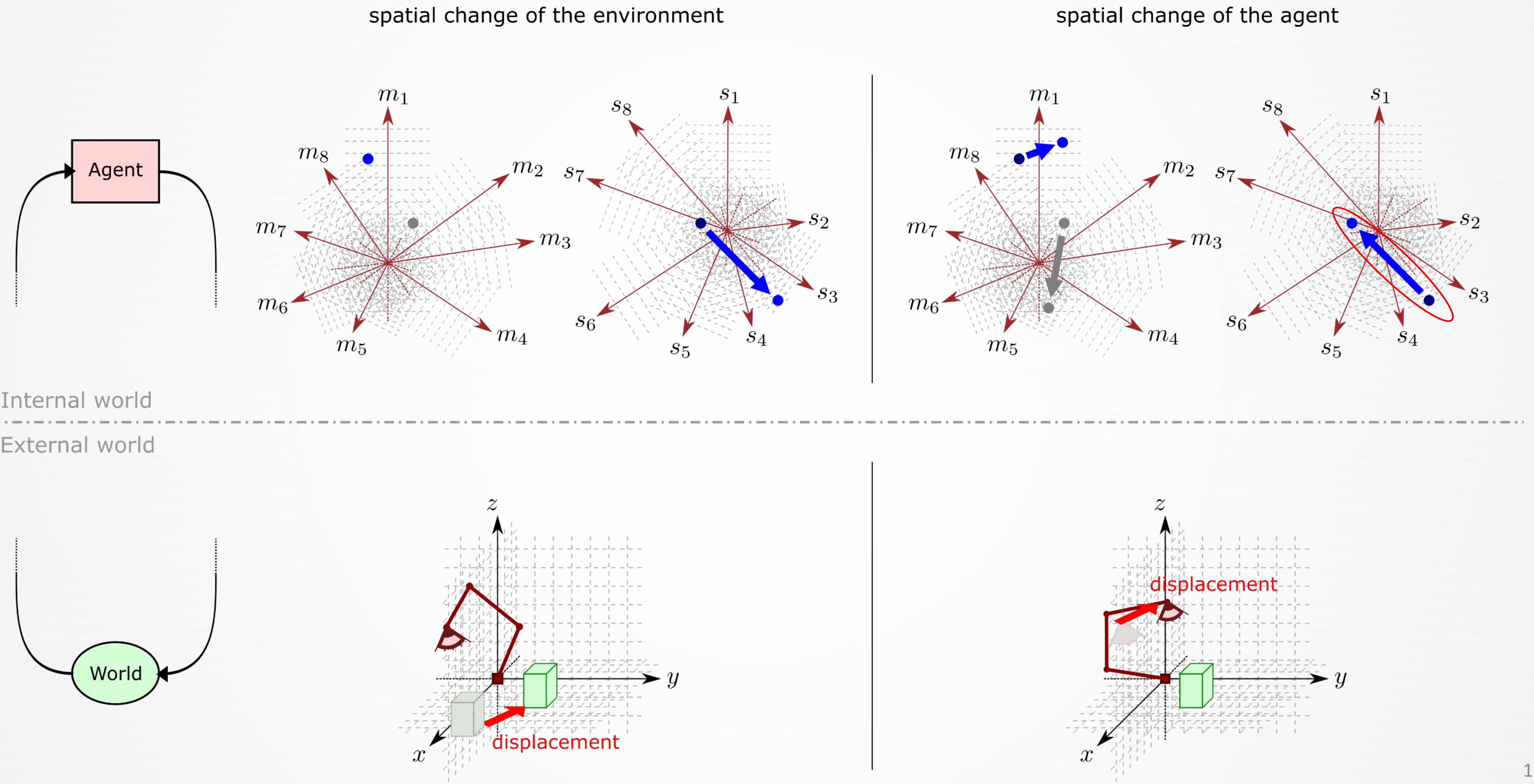
spatial change of the agent



Internal world

External world

“Compensability” holds everywhere



spatial change of the environment

spatial change of the agent

Agent

Internal world

External world

World

Problem

“Internal world”:

- raw sensorimotor states
- high dimension
- own metric

Compensable transitions

Displacements

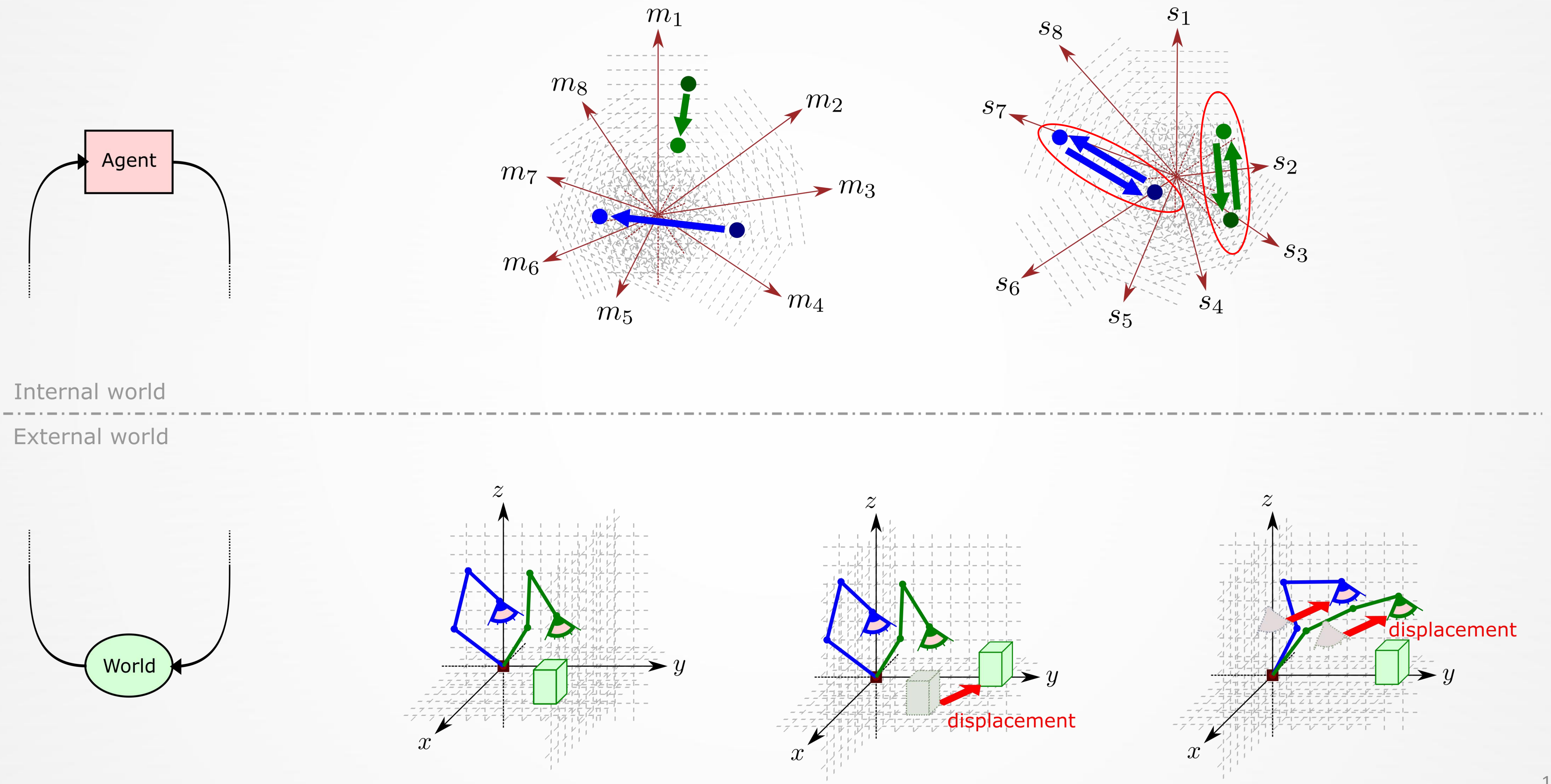
“External world”:

- body and environment
- immersed in space
  - □ low dimension
  - □ Euclidean metric
  - □ shared by agent and environment
  - □ content-independent

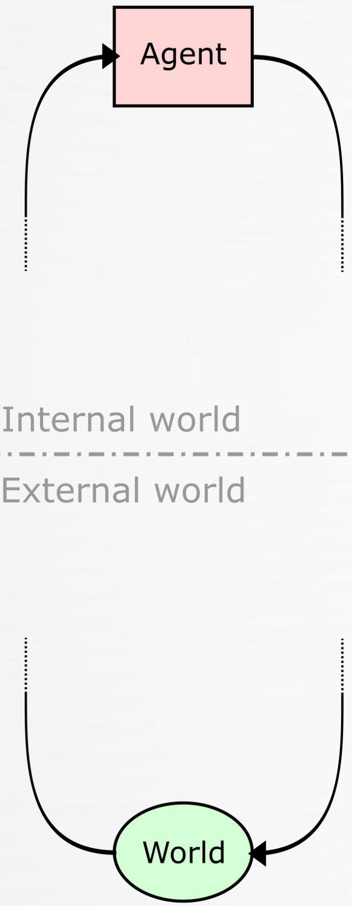
⟶ Can the agent build an internal representation of space, with the same properties? (topology, metric,…)

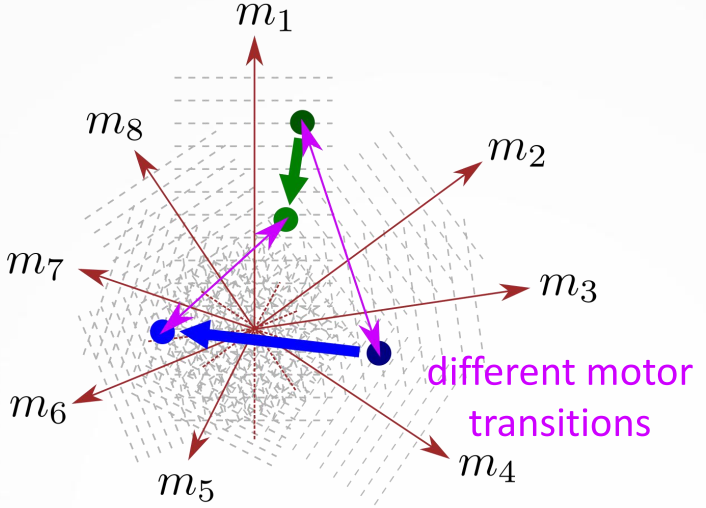
⟶ Is sensorimotor prediction a sufficient drive?

Capturing the Euclidean metric

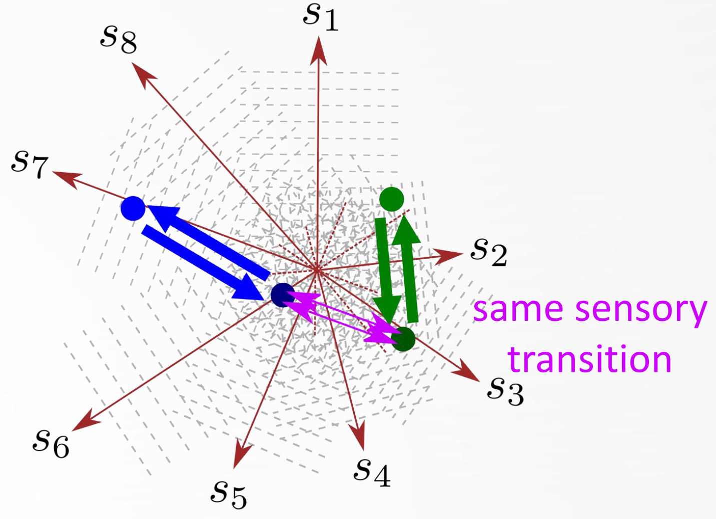


Capturing the Euclidean metric





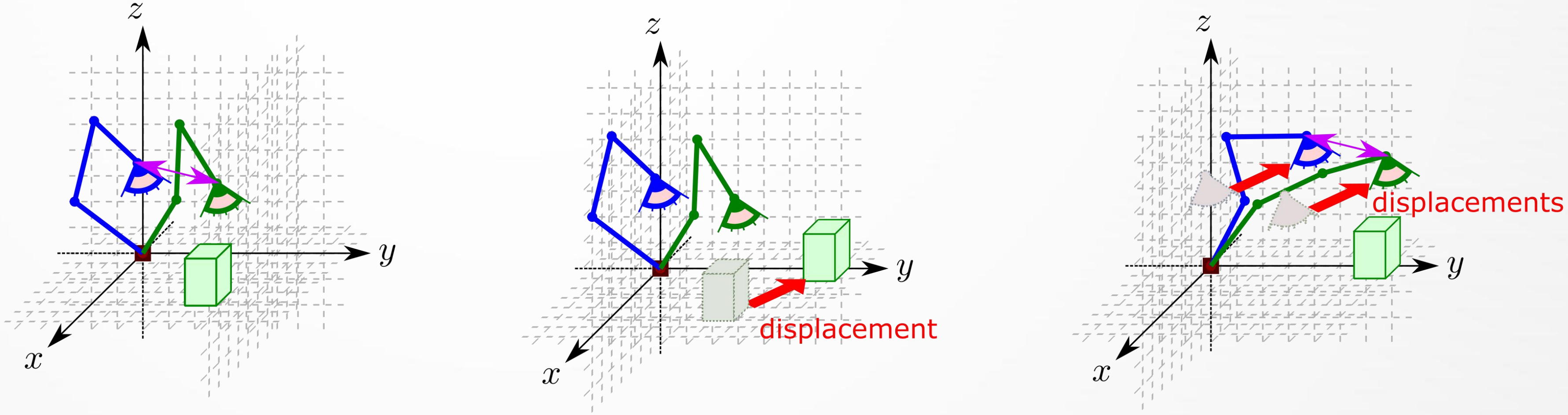
different motor transitions



same sensory transition

Internal world

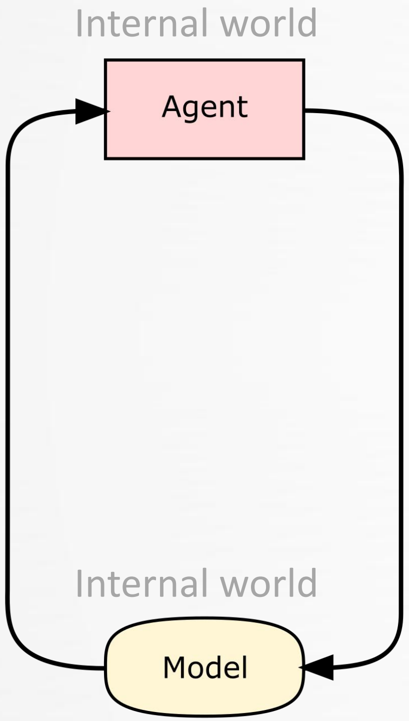
External world



displacement

displacements

Capturing the Euclidean metric

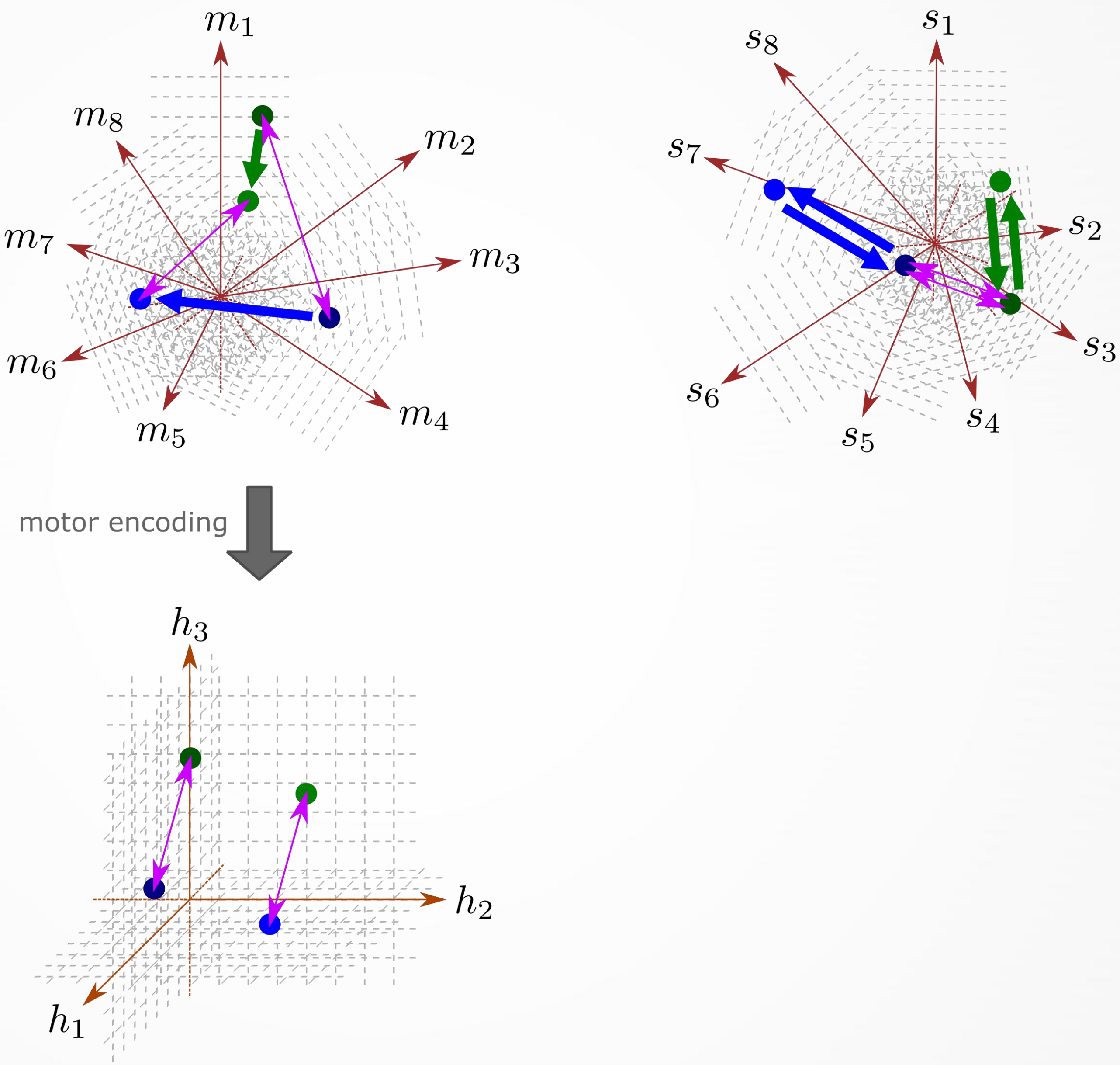


Internal world

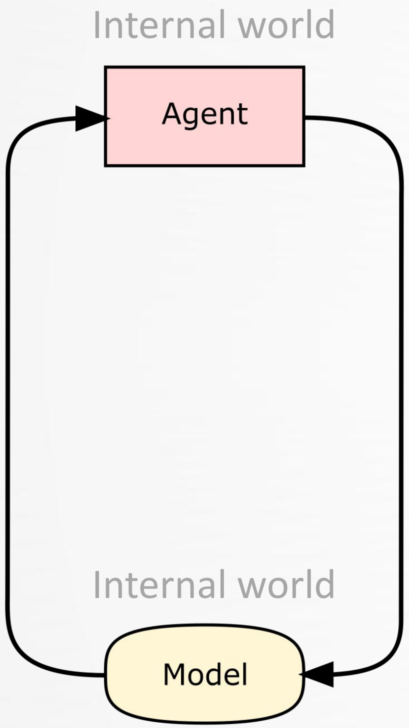
Agent

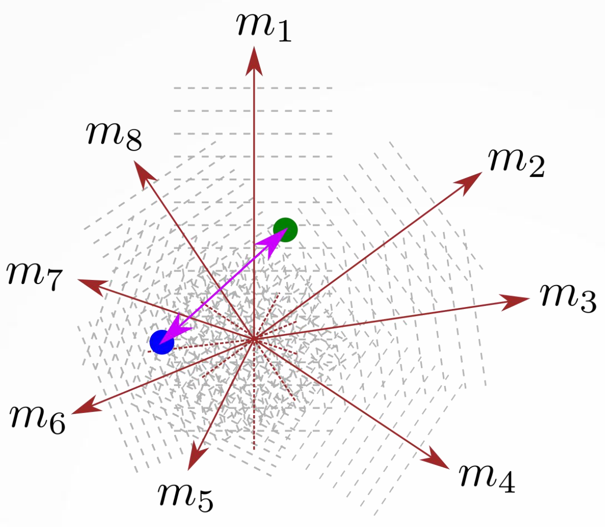
Internal world

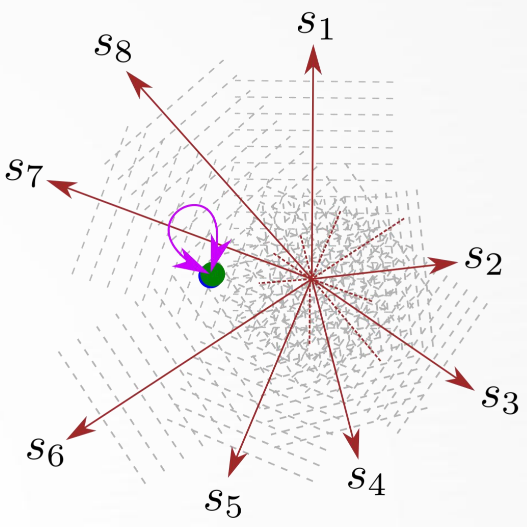
Model

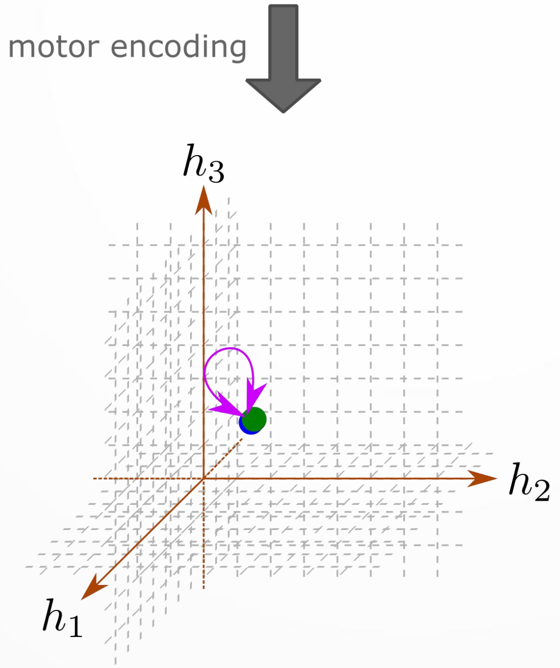


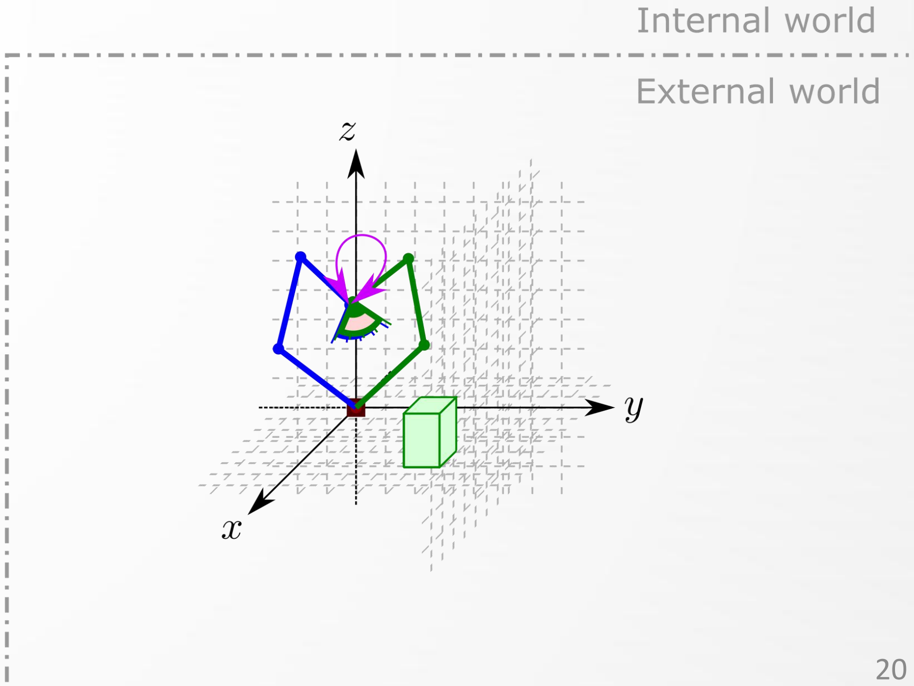
Capturing the topology of space





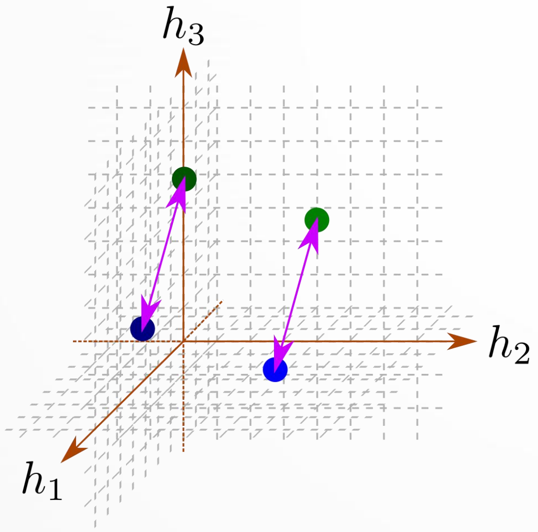


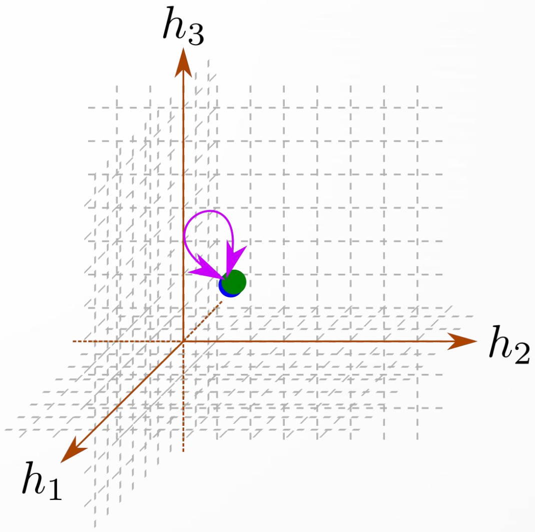




Capturing the structure of the external space

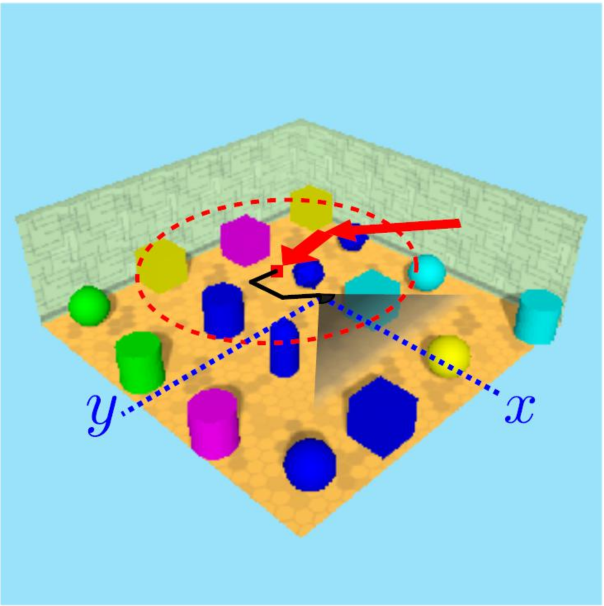
- The external space has a low-dimensional Euclidean structure
- This structure induces invariants in sensorimotor transitions
- Capturing these invariants should be beneficial for sensorimotor prediction
- They can be grounded in the motor space, by encoding motor states such that:
  - □ motor transitions corresponding to the same external displacements are encoded by the same internal change
  - □ motor states corresponding to the same position are encoded into a single point

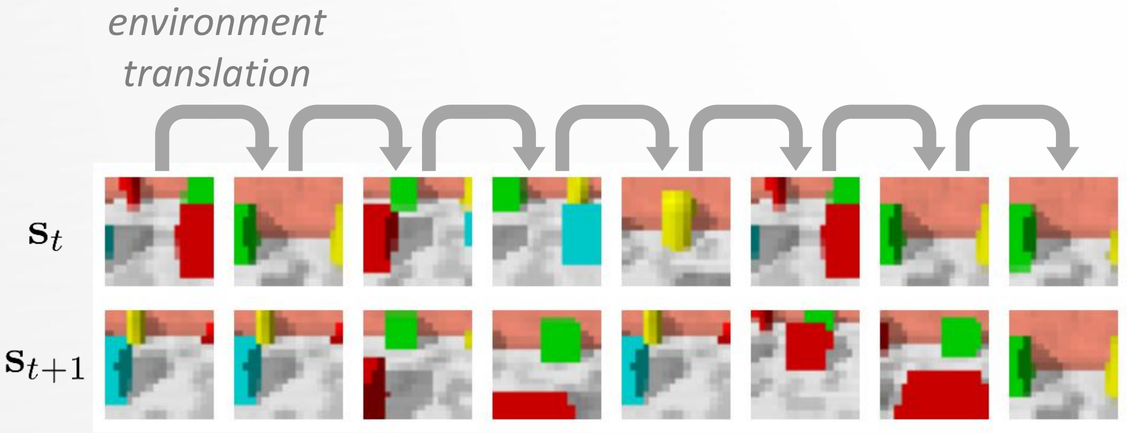


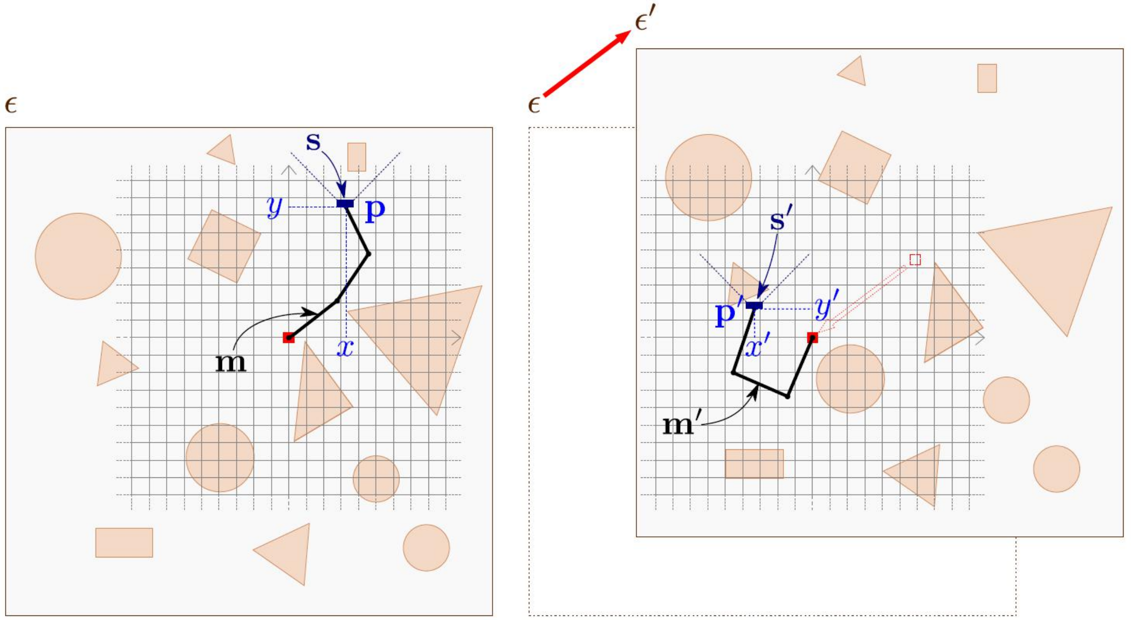


Experiment: simulation

- Robot arm
  - □ 3 degrees of freedom
  - □ camera on the tip (16 x 16 RGB), with fixed orientation
  - □ base fixed in space
- Environment
  - □ room filled with objects
  - □ can translate with respect to the agent
- Data:
  - □ transitions $(\mathbf{m}_t, \mathbf{s}_t) \rightarrow (\mathbf{m}_{t+1}, \mathbf{s}_{t+1})$
  - □ environment translations in-between SM transitions





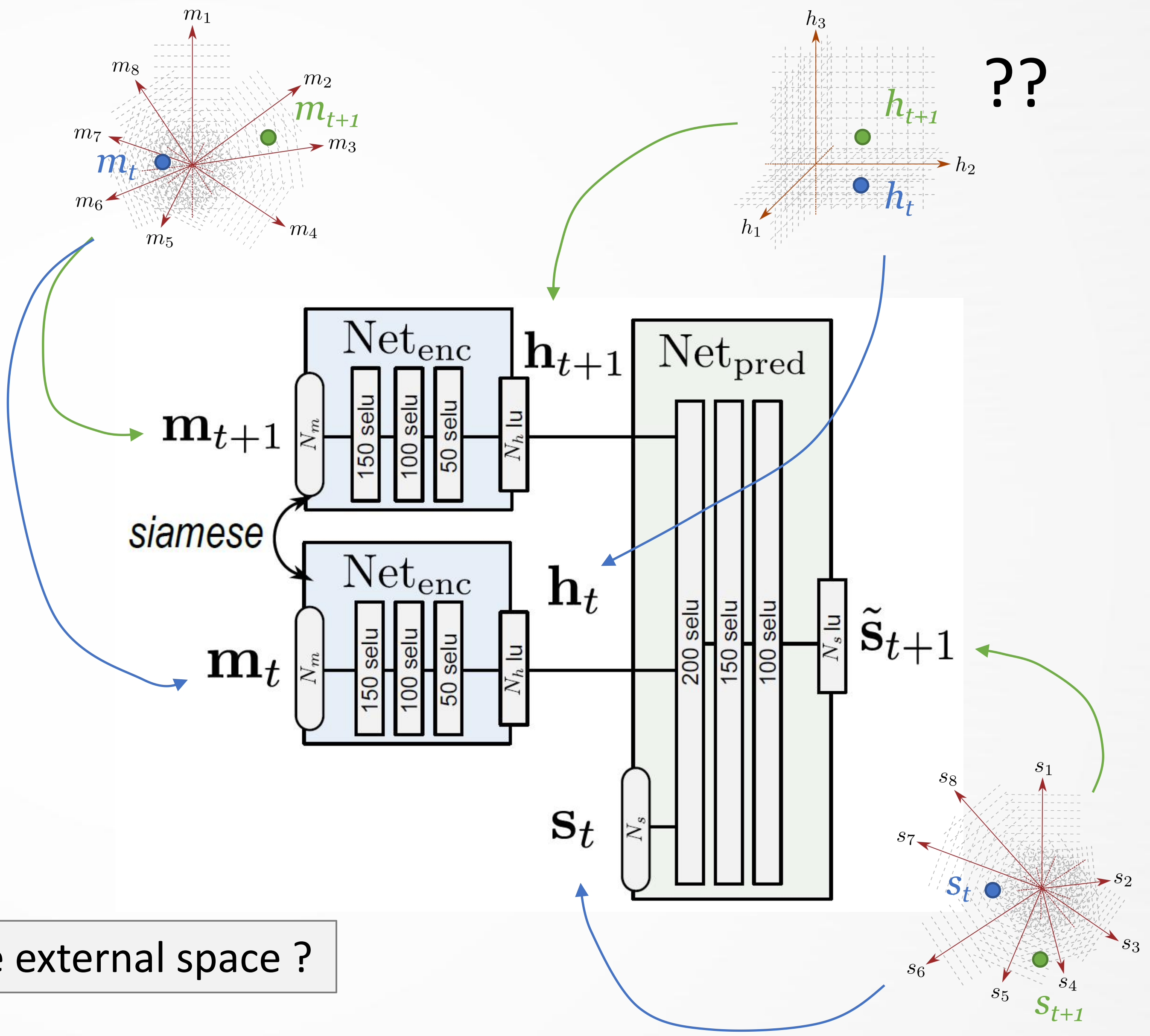


Experiment: predictive model

- Sensorimotor predictive model:
  $$f(\mathbf{m}_t, \mathbf{s}_t, \mathbf{m}_{t+1}) = \tilde{\mathbf{s}}_{t+1}$$

- Motor states are encoded using a shared encoding module

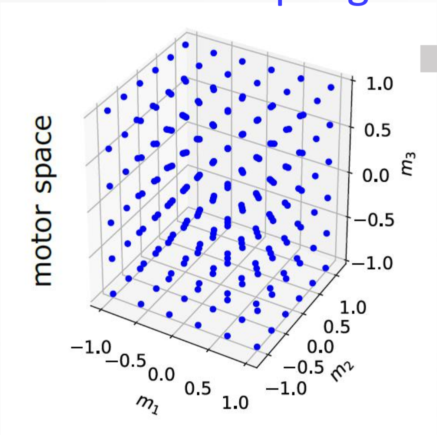
- Unconstrained encoding h

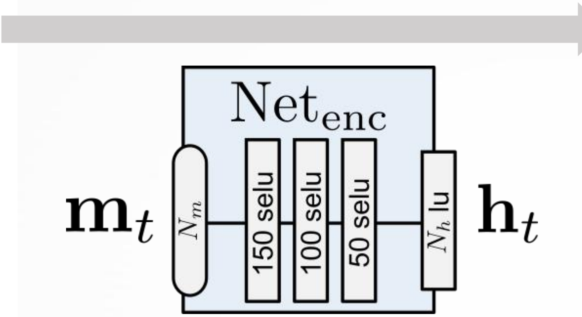


Will h capture the Euclidean structure of the external space ?

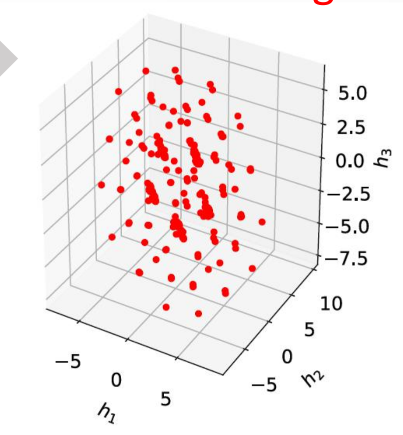
Results: motor encoding

motor sampling

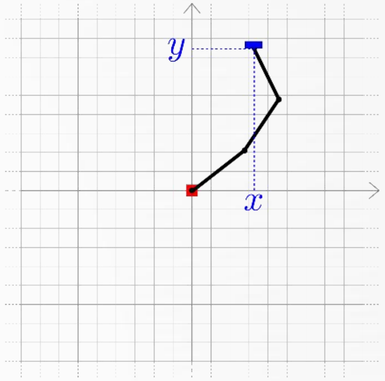




motor encoding

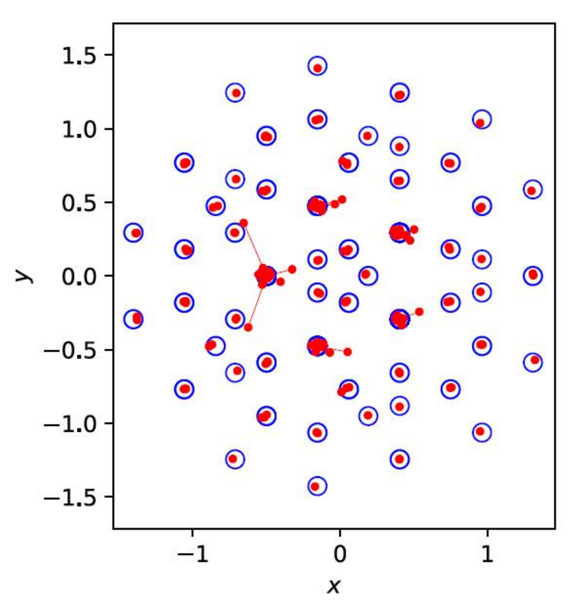


true sensor position



Test of the encoding module:

- regularly sample the motor space
- encode the motor samples
- compare the encoding with the true sensor position (affine transformation)

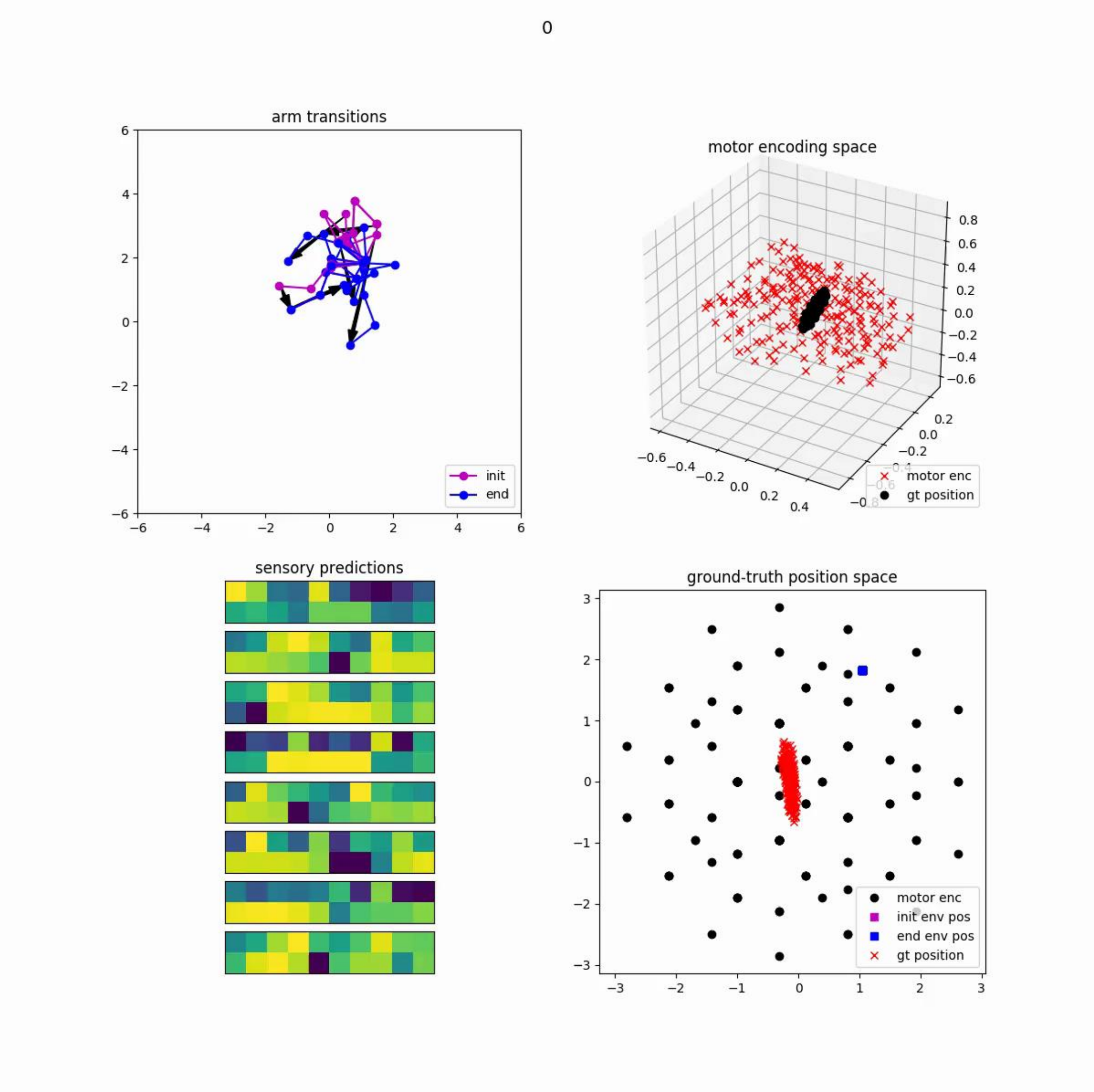


motor encoding

true sensor position

The structure of the encoding matches the one of the external space!

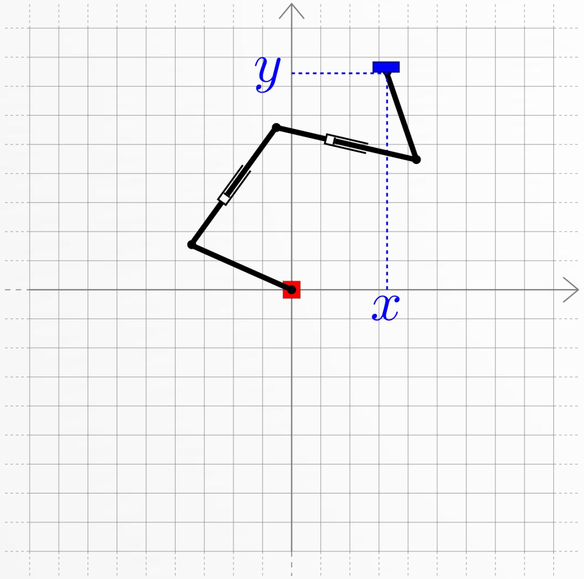
Results: motor encoding

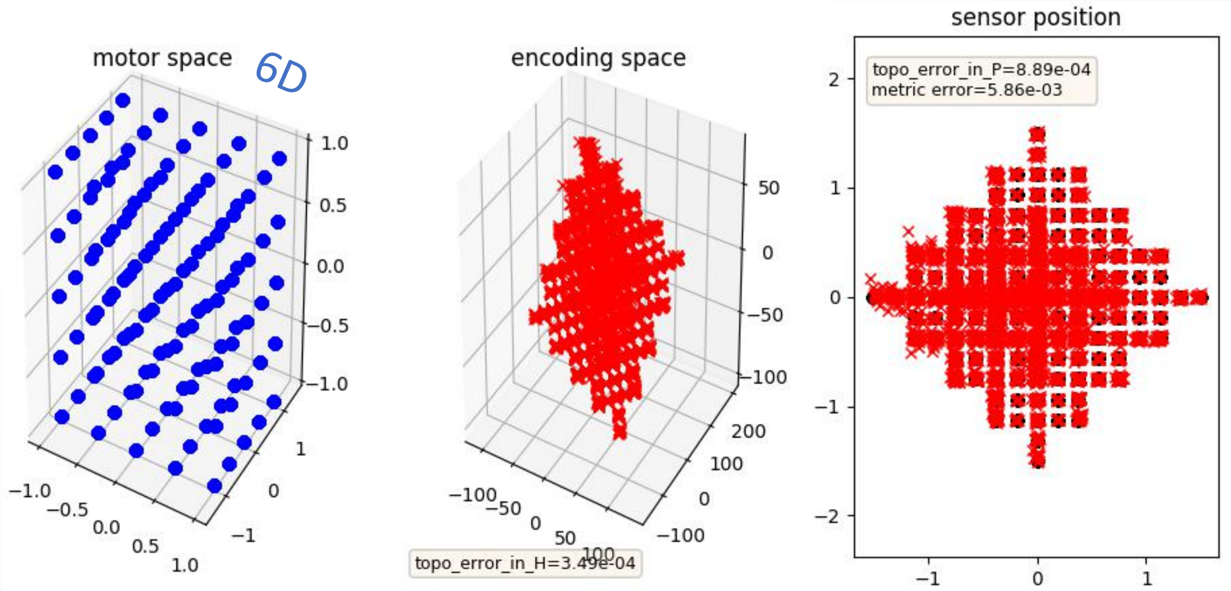


Results: motor encoding

The structure of the encoding matches the one of the external space

- robust to the body complexity
- robust to the type of sensor
- robust to the dimension of the encoding space
- robust to the complexity of the neural network

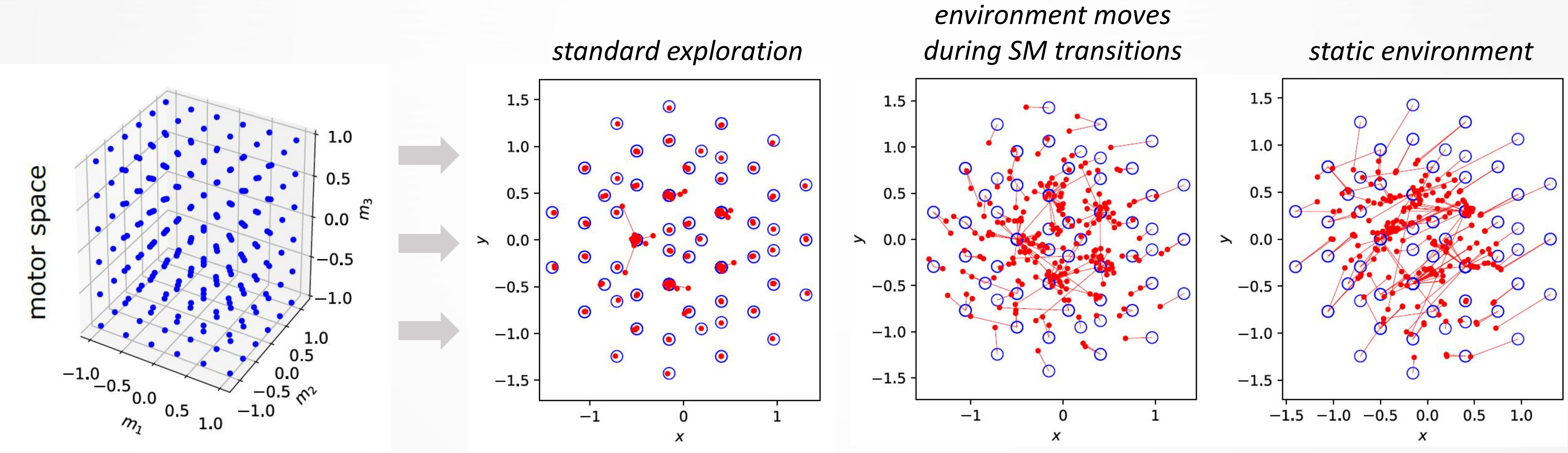




Results: motor encoding

For the spatial structure to emerge:

- The agent needs to experience consistent SM transitions (predictability)
- The agent needs to experience different positions of the environment



standard exploration

environment moves during SM transitions

static environment

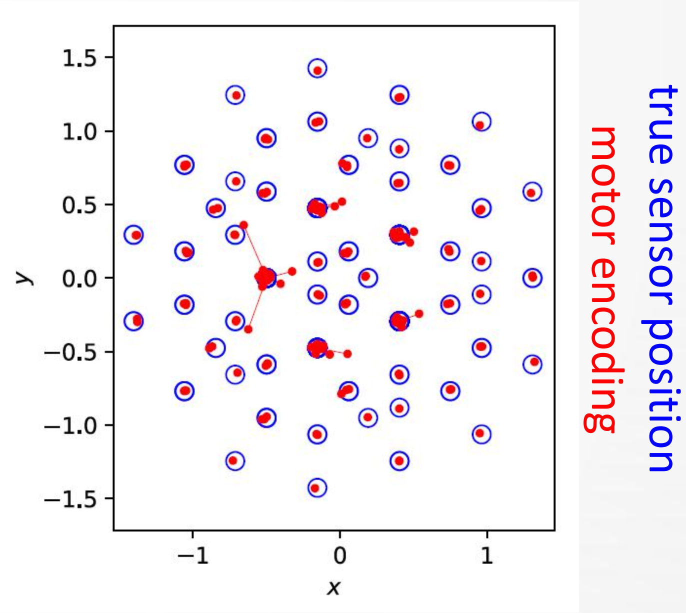
Conclusion

Take home messages:

- it is possible to go from scratch to a space akin to “xyz”
- spatial structure shapes SM experiences
  → displacements appear as compensable transitions
- capturing this structure is advantageous for SM prediction
- spatial structure gets captured in the motor encoding when optimizing for sensorimotor prediction

Compensable transitions

Displacements



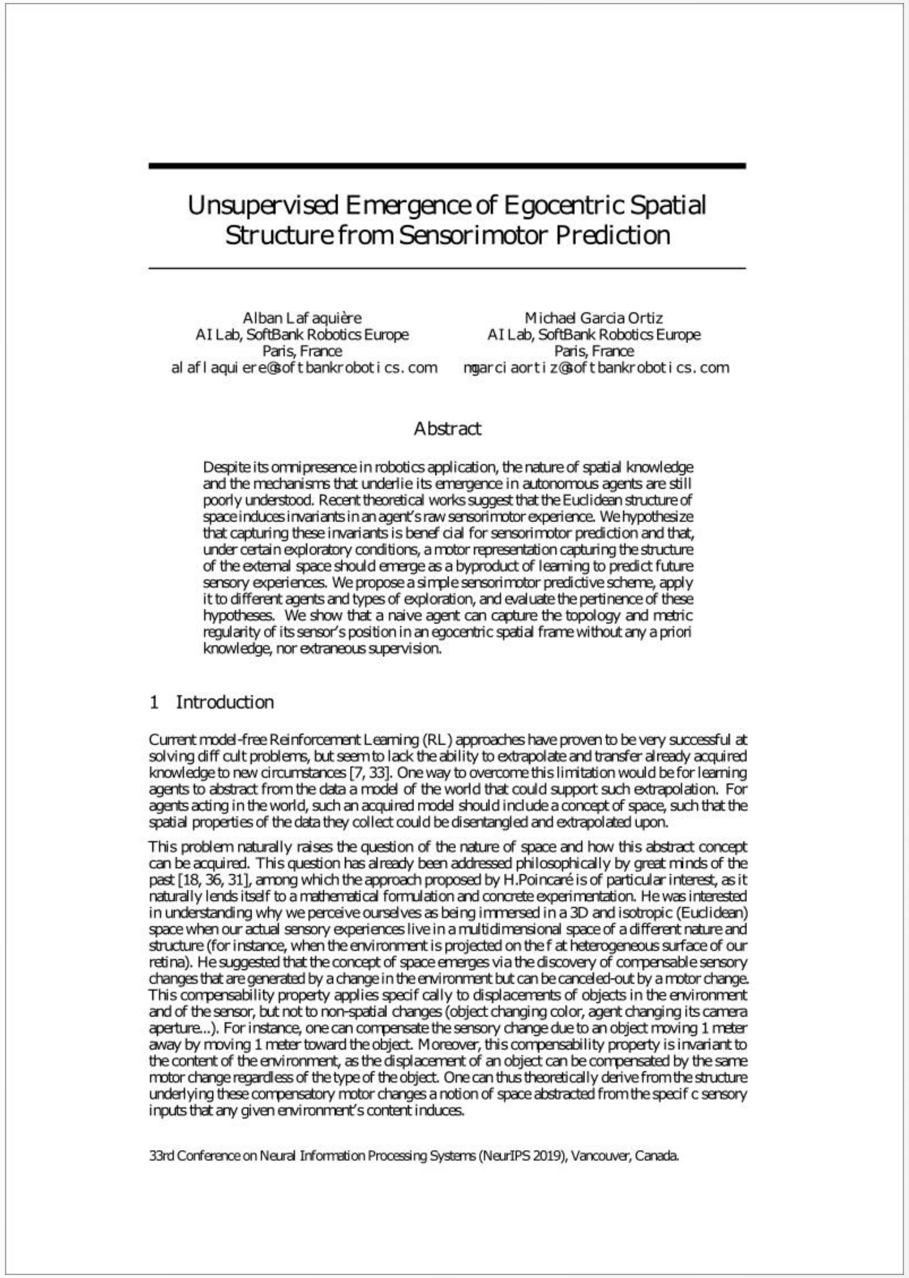
Important points:

- action is required for a notion of space
- displacements of the environment also
- spatial knowledge grounded in the motor space
- spatial knowledge based on egocentric frame

Thank you!

https://github.com/alaflaquiere/learn-spatial-structure
https://arxiv.org/abs/1906.01401

https://github.com/alaflaquiere
alaflaquiere@softbankrobotics.com



Unsupervised Emergence of Egocentric Spatial Structure from Sensorimotor Prediction

Alban Laflaquière
AI Lab, SoftBank Robotics Europe
Paris, France
alaflaquiere@softbankrobotics.com

Michael Garcia Ortiz
AI Lab, SoftBank Robotics Europe
Paris, France
mgarciaortiz@softbankrobotics.com

Abstract

Despite its omnipresence in robotics application, the nature of spatial knowledge and the mechanisms that underlie its emergence in autonomous agents are still poorly understood. Recent theoretical works suggest that the Euclidean structure of space induces invariants in an agent's raw sensorimotor experience. We hypothesize that capturing these invariants is beneficial for sensorimotor prediction and that, under certain exploratory conditions, a motor representation capturing the structure of the external space should emerge as a byproduct of learning to predict future sensory experiences. We propose a simple sensorimotor predictive scheme, apply it to different agents and types of exploration, and evaluate the pertinence of these hypotheses. We show that a naive agent can capture the topology and metric regularity of its sensor's position in an egocentric spatial frame without any a priori knowledge, nor extraneous supervision.

1 Introduction

Current model-free Reinforcement Learning (RL) approaches have proven to be very successful at solving difficult problems, but seem to lack the ability to extrapolate and transfer already acquired knowledge to new circumstances [7, 33]. One way to overcome this limitation would be for learning agents to abstract from the data a model of the world that could support such extrapolation. For agents acting in the world, such an acquired model should include a concept of space, such that the spatial properties of the data they collect could be disentangled and extrapolated upon.

This problem naturally raises the question of the nature of space and how this abstract concept can be acquired. This question has already been addressed philosophically by great minds of the past [18, 36, 31], among which the approach proposed by H.Poincaré is of particular interest, as it naturally lends itself to a mathematical formulation and concrete experimentation. He was interested in understanding why we perceive ourselves as being immersed in a 3D and isotropic (Euclidean) space when our actual sensory experiences live in a multidimensional space of a different nature and structure (for instance, when the environment is projected on the flat heterogeneous surface of our retina). He suggested that the concept of space emerges via the discovery of compensable sensory changes that are generated by a change in the environment but can be canceled-out by a motor change. This compensability property applies specifically to displacements of objects in the environment and of the sensor, but not to non-spatial changes (object changing color, agent changing its camera aperture...). For instance, one can compensate the sensory change due to an object moving 1 meter away by moving 1 meter toward the object. Moreover, this compensability property is invariant to the content of the environment, as the displacement of an object can be compensated by the same motor change regardless of the type of the object. One can thus theoretically derive from the structure underlying these compensatory motor changes a notion of space abstracted from the specific sensory inputs that any given environment's content induces.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.