“What is it like to be a Braitenberg vehicle?”?

Manuel Baltieri - October 4th 2025, Models of Consciousness 6, Sapporo, Japan

ARAYA RESEARCH

Plan

Work in progress + some rants + looking for feedback/collaborators

- “Objective” to subjective models
- Embodiment
- Symmetries to go subjective
- Braitenberg vehicles and their *phenomenology*
- Qualia?

Models

- Internal model principle
- Good regulator theorem
- Interfaces and causal states
- Beliefs and synchronisation

A Bayesian Interpretation of the Internal Model Principle

Manuel Baltieri, Araya Inc.[†]
Martin Biehl, Cross Labs, Cross Compass Ltd.
Matteo Capucci, University of Strathclyde and Independent Researcher
Nathaniel Virgo, University of Hertfordshire and
Earth-Life Science Institute, Institute of Science Tokyo

Abstr
the theo
more ge
The cen
tions, if
external
contains
be used
on the r
especiall
often su
agent) n
perform
the Bay
any forr
treatmer
model p
concepts
formal c
model p
related t
be seen
result is
categori
can be i
world ir

In a
good
Artif
perfo
Cona
its re
intui
agen
obse
ment
notic
more
orem
a cha
tial r
syste
regar
in a
its ov
way.
appa

The wo
develo

A "Good Regulator Theorem" for Embodied Agents

Nathaniel Virgo[1,2], Martin Biehl[3], Manuel Baltieri[4], and Matteo Capucci[5]
[1]University of Hertfordshire, UK [2]Earth-Life Science Institute (ELSI), Institute of Science Tokyo, Japan
[3]Cross Labs, Cross Compass Ltd., Japan [4]Araya, Inc., Japan [5]Independent researcher

AI in a vat: Fundamental limits of efficient world modelling for agent sandboxing and interpretability

Fernando E. Rosas[1-5], **Alexander Boyd**[1,6], **Manuel Baltieri**[7,1]
f.rosas@sussex.ac.uk, alecboy@gmail.com, manuel_baltieri@araya.org

ARAYA RESEARCH

A coalgebraic perspective on predictive processing

Manuel Baltieri [1,2,†], Filippo Torresan [1,2], Tomoya Nakai [1]

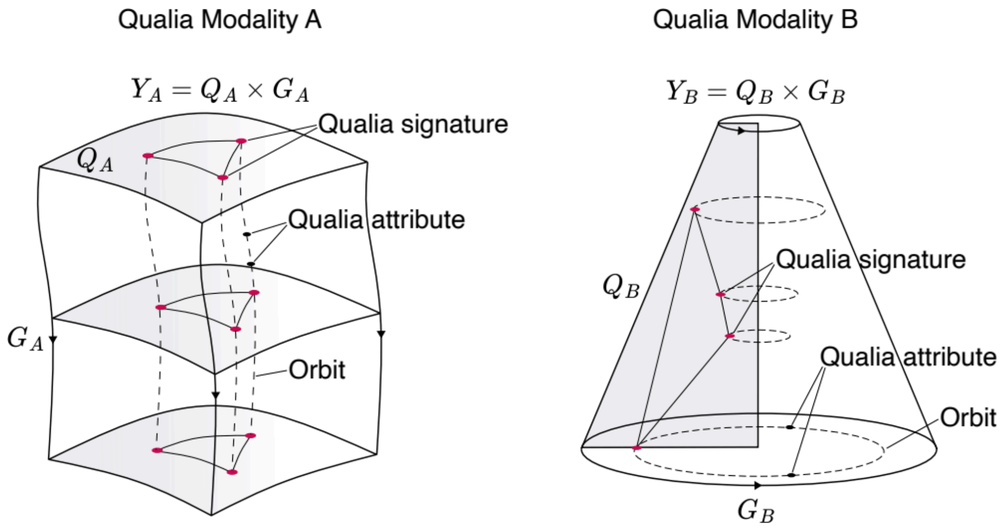
[1] Araya Inc., Tokyo, Japan
[2] University of Sussex, Brighton, UK

Predictive processing and active inference posit that the brain is a system performing Bayesian inference on the environment. By virtue of this, a prominent interpretation of predictive processing states that the generative model (a POMDP) encoded by the brain synchronises with the generative process (another POMDP) representing the environment while trying to explain what hidden properties of the world generated its sensory input. In this view, the brain is thought to become a copy of the environment. This claim has however been disputed, stressing the fact that a structural copy, or isomorphism as it is at times invoked to be, is not an accurate description of this process since the environment is necessarily more complex than the brain, and what matters is not the capacity to exactly recapitulate the veridical causal structure of the world. In this work, we make parts of this counterargument formal by using ideas from the theory of coalgebras, an abstract mathematical framework for dynamical systems that brings together work from automata theory, concurrency theory, probabilistic processes and other fields. To do so, we cast

This week

Ryota's proposal

- **Qualia modality**: Defined by the algebraic structure of the symmetry group G itself. The non-isomorphic nature of the groups that govern different senses (e.g., $SO(2)$ for hue versus $\mathbb{R}$ for pitch) accounts for the fundamental differences between sensory worlds.
- **Qualia signature**: Defined by a point in the bundle's base space, the quotient space ($Q = Y/G$). This represents the G-invariant identity of a percept (e.g., the signature 'cat').
- **Qualia attribute**: Defined by a point on a specific orbit or fiber $\mathcal{O} \subset Y$. This represents the continuous, G-variant parameters of a percept (e.g., the specific location of the 'cat').



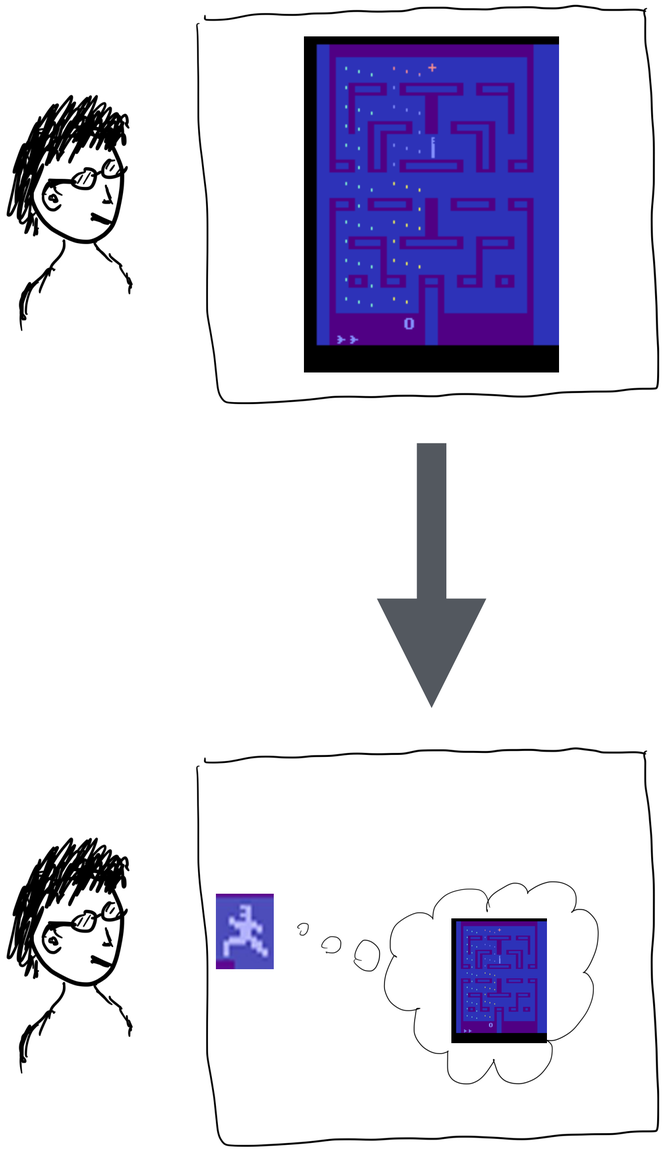
Change of perspective

Some good ideas

Attempt to describe some aspects of phenomenology.

Moving from “objective” (imposed externally) description to subjective world (?).

Q: “What is it like to be a bat?”
A: "I can’t know, because my experience is invariant to things a bat is not”



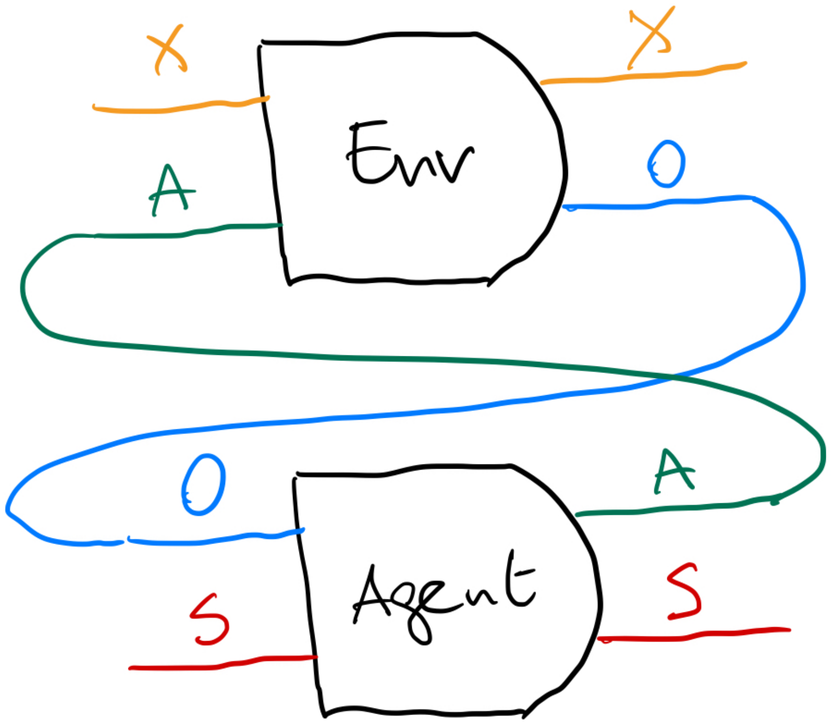
Embodiment and situatedness

Subjective + open + time-dependent

- Agents are open systems
- Agents are time-dependent

We want a model that is

1. subjective,
2. open, and
3. time-dependent.



What an agent senses

Interface

- a collection $\{p(o_{:t} \mid a_{:}), t \in \mathbb{N}\}$.

World model

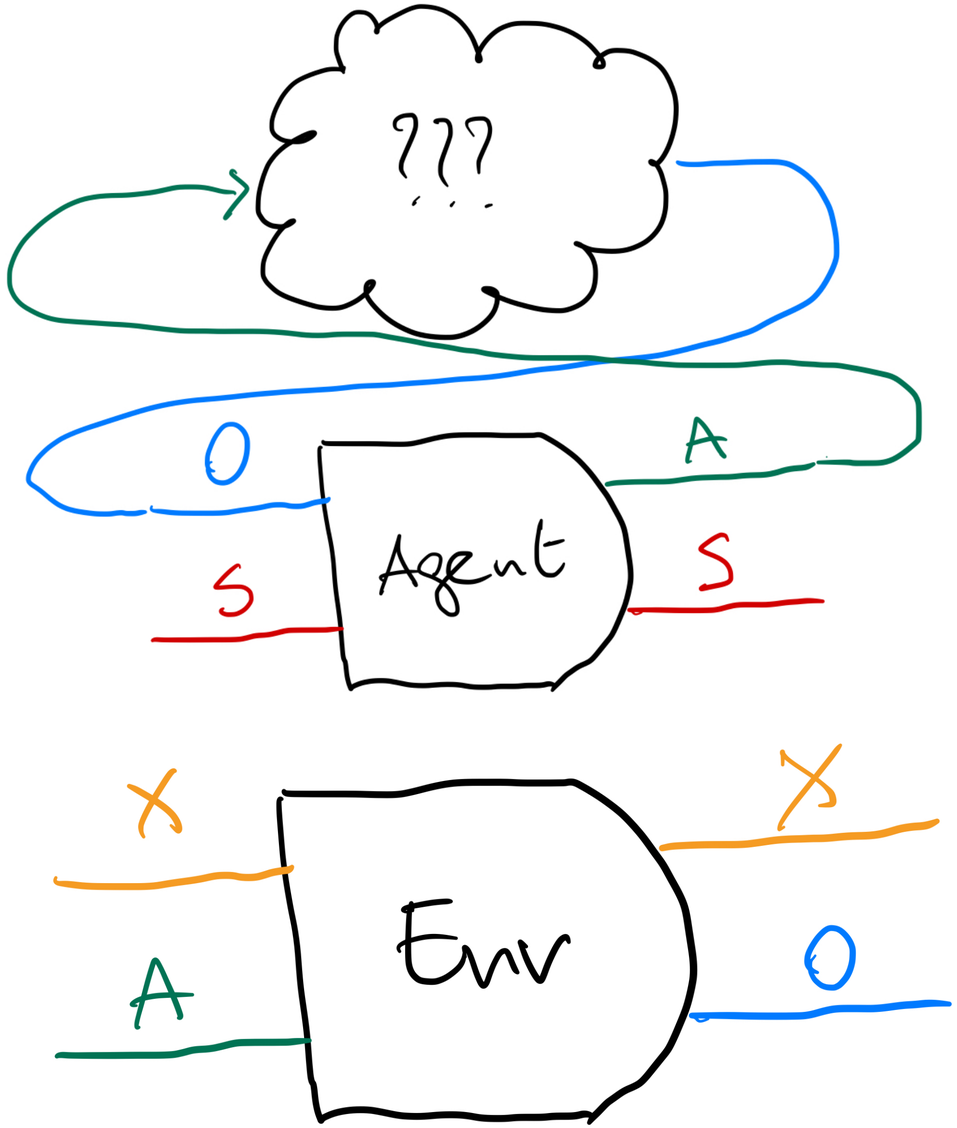
- a collection $\{p(s_{:t} \mid o_{:}, a_{:}), t \in \mathbb{N}\}$ that obeys,

  $\forall t \in \mathbb{N}$

  (1) $p(o_t \mid s_{:t}, a_{:t}, o_{:t-1}) = p(o_t \mid s_t, a_t)$, and

  (2) $p(s_{:t} \mid o_{:t-1}, a_{:t-1}, a_{t:}) = p(s_{:t} \mid o_{:t-1}, a_{:t-1})$

  and generates a given interface.



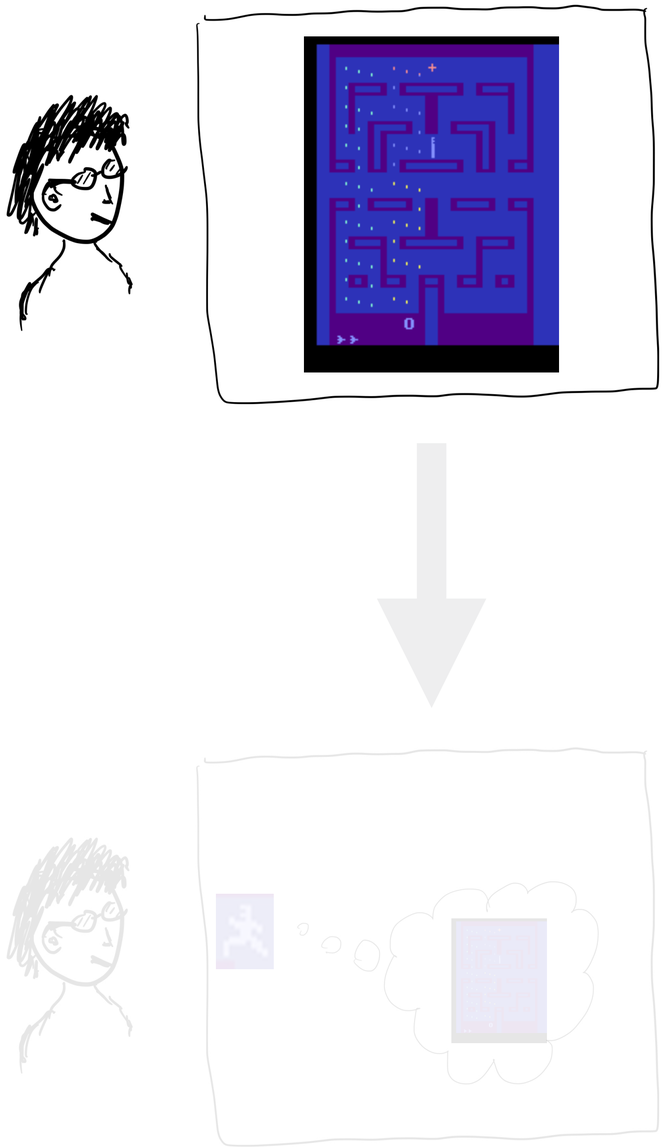
POMDPs as world models

“Objective” descriptions of the world

A POMDP (S, A, O, T, M) is given by

- State space, S
- Action space, A
- Observations, O
- Transition map, $T : S \times A \to P(S)$, often just written as $T(s' \mid s, a)$
- Observation map, $M : S \to P(O)$, often just written as $M(o \mid s, a)$

where $P(X)$ is the set of distributions on X.



Bisimulation equivalences of POMDPs

Moving from objective to subjective

Given a POMDP, an equivalence relation $R \subseteq S \times S$ is a bisimulation equivalence if the following conditions hold for $C \in S/R$:

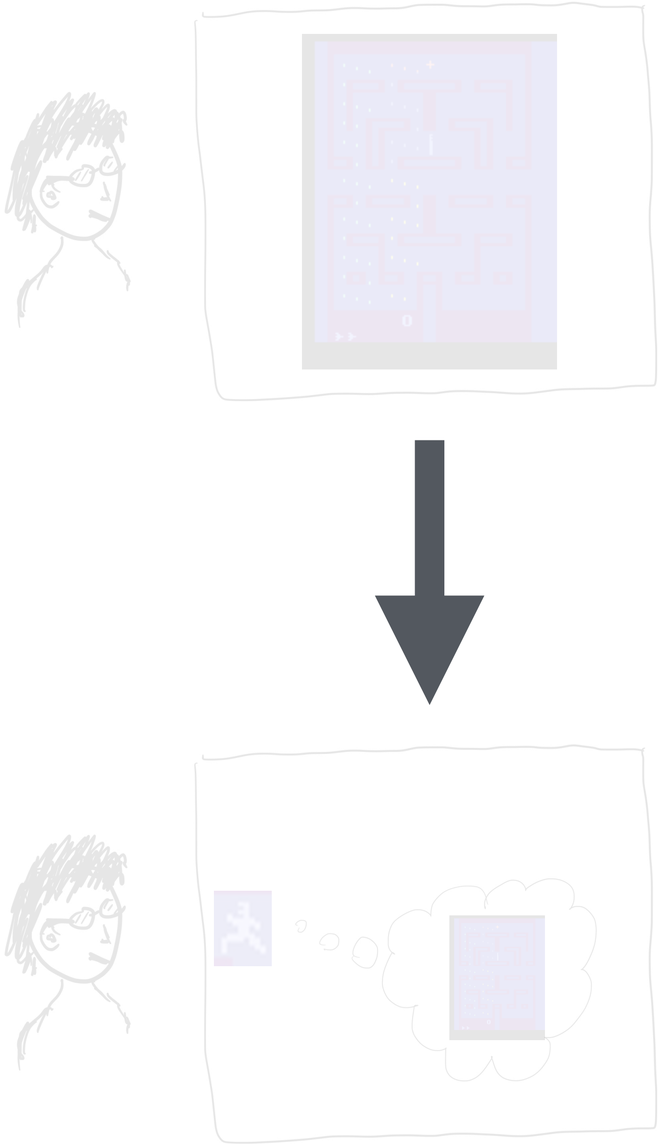
$$\sum_{s' \in C} T(s' \mid s, a) = \sum_{s' \in C} T(s' \mid \bar{s}, a)$$

(transitions)

$$M(o \mid s, a) = M(o \mid \bar{s}, a)$$

(observations)

and we say that $s \sim_R \bar{s}$ when $(s, \bar{s}) \in R$.



Quotient POMDP induced by R_G

Subjective description of the world

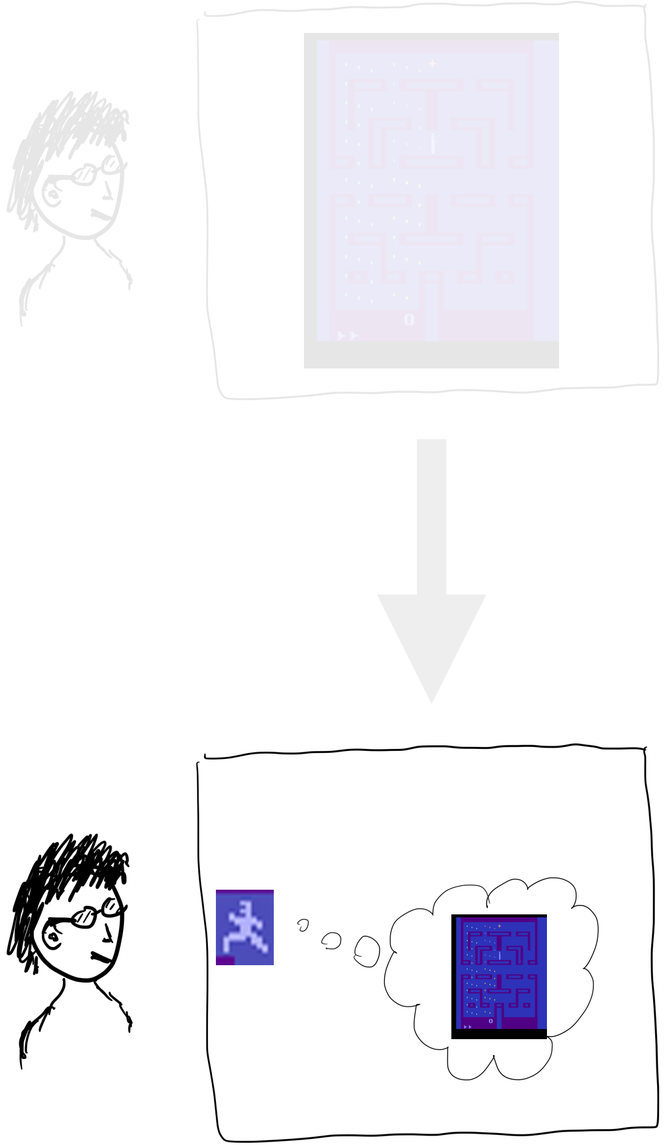
Given a G-equivariant POMDP $(S, A, O, T, M, G, \alpha_S, \alpha_A, \alpha_O)$ is the tuple $(S/R_G, A, O/R_G, \bar{T}, \bar{M})$, given by

- State space, orbit space S/R_G
- Action space, A
- Observation space: O/R_G
- Transition kernel $\bar{T} : (S/R_G) \times A \to P(S/R_G)$:

$$\bar{T}\left([s'] \mid [s], a\right) := \sum_{u \in [s']} T(u \mid s, a), \quad [s], [s'] \in S/R_G, s \in [s], a \in A$$

- Observation kernel $\bar{\Omega} : S/R_G \to P(O/R_G)$:

$$\bar{\Omega}([o] \mid [s]) := \sum_{v \in [o]} \Omega(v \mid s), \quad [s] \in S/R_G, [o] \in O/G, s \in [s]$$



Umwelt

Subjective world of an agent

Theory Biosci. (2015) 134:105–116
DOI 10.1007/s12064-015-0217-3

ORIGINAL PAPER

The *Umwelt* of an embodied agent—a measure-theoretic definition

Nihat Ay[1,2,3] · Wolfgang Löhr[4]

Published online: 9 December 2015

Abstract We consider a general model of the sensorimotor loop of an agent interacting with the world. This formalises Uexküll's notion of a *function-circle*. Here, we assume a particular causal structure, mechanistically described in terms of Markov kernels. In this generality, we define two σ-algebras of events in the world that describe two respective perspectives: (1) the perspective of an external observer, (2) the intrinsic perspective of the agent. Not all aspects of the world, seen from the external perspective, are accessible to the agent. This is expressed by the fact that the second σ-algebra is a subalgebra of the first one. We propose the smaller one as formalisation of Uexküll's *Umwelt* concept. We show that, under continuity and compactness assumptions, the global dynamics of the world can be simplified without changing the internal process. This simplification can serve as a minimal world model that the system must have in order to be consistent with the internal process.

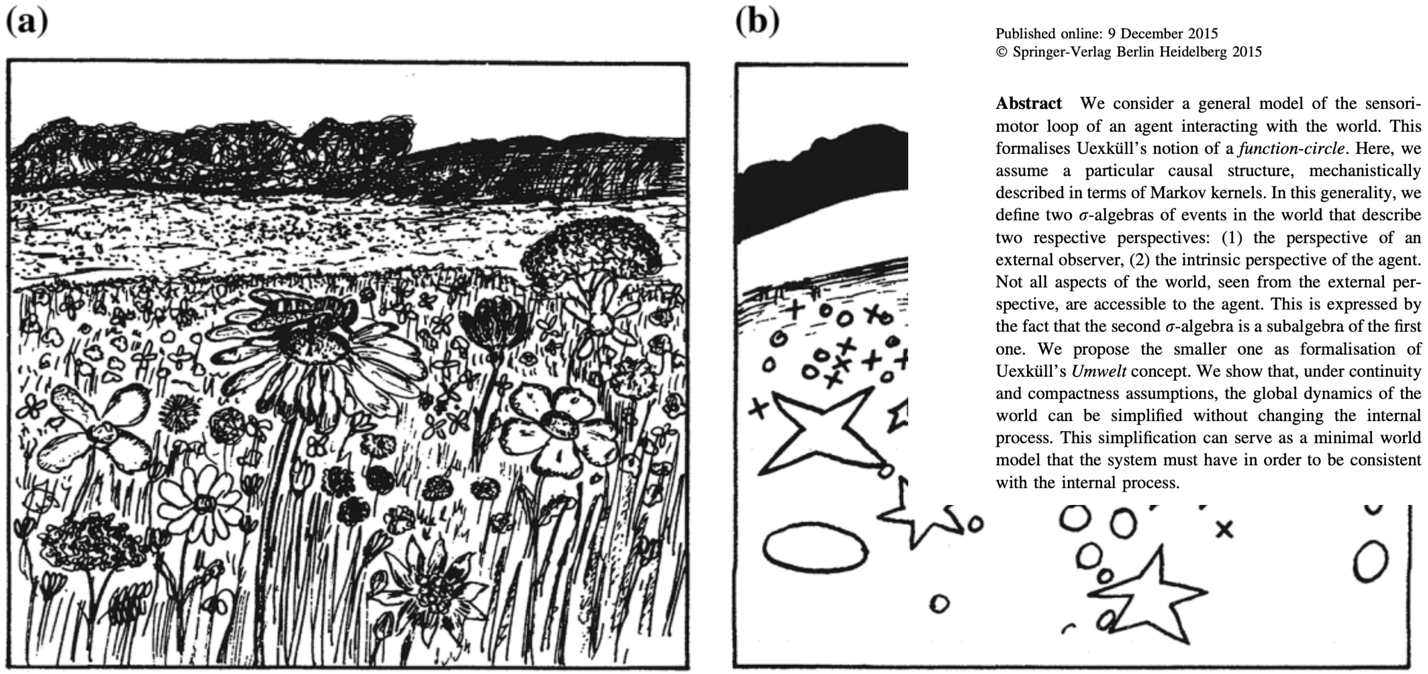


Fig. 1 The *Umwelt* of a bee as illustrated in Von Uexküll (1934). **a** The environment of a bee how we perceive it as an external observer. **b** The same bee perceives only particular aspects of the same world, which constitute its *Umwelt*

Introduction: the intrinsic view of embodied agents

Uexküll's *function-circle* and the sensorimotor loop

A key observation based on many case studies within the field of embodied intelligence implies that quite simple control mechanisms can lead to very complex behaviours (Pfeifer and Bongard 2007). This gap between simplicity and complexity related to the same thing, the agent's behaviour, is the result of two different frames of reference in the description. Here, the intrinsic view of the agent, which provides the basis for its control, can greatly differ from the (extrinsic) view of an external observer. This important understanding is not new. In the first half of the last century, Uexküll has conceptualised this understanding by his notion of *Umwelt*, which summarises all aspects of the world that have an effect on the agent and can be affected by the agent (see Von Uexküll 2014). Furthermore, he has convincingly exemplified this notion in terms

Bisimulation (equivalences)

Published as a conference paper at ICLR 2021

Learning Invariant Representations for Reinforcement L

Amy Zhang[*12] **Rowan**
[1]McGill University
[2]Facebook AI Research
[3]University of California,
[4]OATML group, Universi

We study how repr
rich observations,
pixel-reconstructic
downstream contr
quantify behaviora
pose using to learn
information from
latent space equal
tiveness of our me
visual MuJoCo ta
and natural videos,
highway driving t
time of day. Final
bisimulation metri

1 Introduction

Learning control from ima
applications. While deep re

Learning Causal State Representations of Partially Observable Environments

Amy Zhang
McGill
Faceboo
Zacha
Carnegi
Luis F
Faceboo
Kamy
Purdue
Anima
Californ
Laure
Univers
Joelle
McGill
Faceboo
Tomm
HK3 La

Editor:

Int
abs
are
obs

AMY.X.ZHANG@MAIL.MCGILL.CA

Equivalence Relations in Fully and Partially Observable Markov Decision Processes

Pablo Samuel Castro, Prakash Panangaden, Doina Precup
School of Computer Science
McGill University
{pcastr,prakash,dprecup}@cs.mcgill.ca

Abstract

We explore equivalence relations between states in Markov Decision Processes and Partially Observable Markov Decision Processes. We focus on two different equivalence notions: bisimulation [Givan *et al.*, 2003] and a notion of trace equivalence, under which states are considered equivalent if they generate the same conditional probability distributions over observation sequences (where the conditioning is on action sequences). We show that the relationship between these two equivalence notions changes depending on the amount and nature of the partial observability. We also present an alternate characterization of bisimulation based on trajectory equivalence.

1 Introduction

Probabilistic systems are very useful modeling tools in many fields of science and engineering. In order to understand the behavior of existing models, or to provide compact models, notions of equivalence between states in such systems are necessary. Equivalence relations have to be defined in such a way that important properties are preserved, i.e., the *long-term behavior* of equivalent states should be the same. However, there are different ways in which "long-term behavior" could be defined, leading to different equivalence notions. In this paper, we focus on two equivalence relations which have been explored in depth in the process algebra literature generated several pieces of follow-up work and extensions (e.g. Dean & Givan [1997], Ferns et al. [2004], Taylor et al. [2009]). Comparatively little work has focused on bisimulation for POMDPs, except for a basic definition of a bisimulation notion for POMDP states [Pineau, 2004] (though the terminology of "bisimulation" is not used there). To our knowledge, trace equivalence has not really been explored in either MDPs or POMDPs. However, using traces holds the potential of offering a more efficient and natural way of computing and approximating state equivalence through sampling methods (rather than the global, model-based process used typically to compute bisimulation). Moreover, in POMDPs, trace equivalence is intimately related to predictive state representations (PSRs) [Litman *et al.*, 2002] as well as lossless compression [Poupart and Boutilier, 2003]. As we will discuss in more detail later, this link opens up other potential avenues for checking trace equivalence efficiently.

In this paper we investigate the relationship between bisimulation and trace equivalence, focusing on partially observable systems. We show that these two notions are not equivalent in MDPs, but they can be equivalent in POMDPs. We also present a different characterization of bisimulation in MDPs based on trace equivalence, which could potentially yield new algorithms for computing or approximating bisimulation.

The paper is organized as follows. In Sec. 2, we present the definitions and theoretical analysis of the relationship between bisimulation and trajectory (or trace) equivalence in MDPs. The analysis reveals the surprising fact that trajec-

ARAYA RESEARCH

A coalgebraic perspective on predictive processing

Manuel Baltieri [1,2,†], Filippo Torresan [1,2], Tomoya Nakai [1]

[1] Araya Inc., Tokyo, Japan
[2] University of Sussex, Brighton, UK

Predictive processing and active inference posit that the brain is a system performing Bayesian inference on the environment. By virtue of this, a prominent interpretation of predictive processing states that the generative model (a POMDP) encoded by the brain synchronises with the generative process (another POMDP) representing the environment while trying to explain what hidden properties of the world generated its sensory input. In this view, the brain is thought to become a copy of the environment. This claim has however been disputed, stressing the fact that a structural copy, or isomorphism as it is at times invoked to be, is not an accurate description of this process since the environment is necessarily more complex than the brain, and what matters is not the capacity to exactly recapitulate the veridical causal structure of the world. In this work, we make parts of this counterargument formal by using ideas from the theory of coalgebras, an abstract mathematical framework for dynamical systems that brings together work from automata theory, concurrency theory, probabilistic processes and other fields. To do so, we cast generative model and process, in the form of POMDPs, as coalgebras, and use maps between them to describe a form of consistency that goes beyond mere structural similarity, giving the necessary mathematical background to describe how different processes can be seen as behaviourally, rather than structurally, equivalent, i.e. how they can be seen as emitting the same observations, and thus minimise prediction error, over time without strict assumptions about structural similarity. In particular, we will introduce three standard notions of equivalence from the literature on coalgebras, evaluating them in the context of predictive processing and identifying the one closest to claims made by proponents of this framework.

… but toy models are great

On the utility of toy models for theories of consciousness

Larissa Albantakis[1,*]

[1] Department of Psychiatry, University of Wisconsin, Madison, WI, USA
* albantakis@wisc.edu

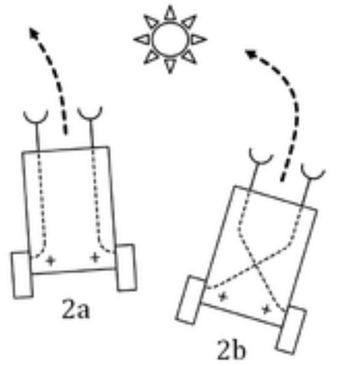
Abstract

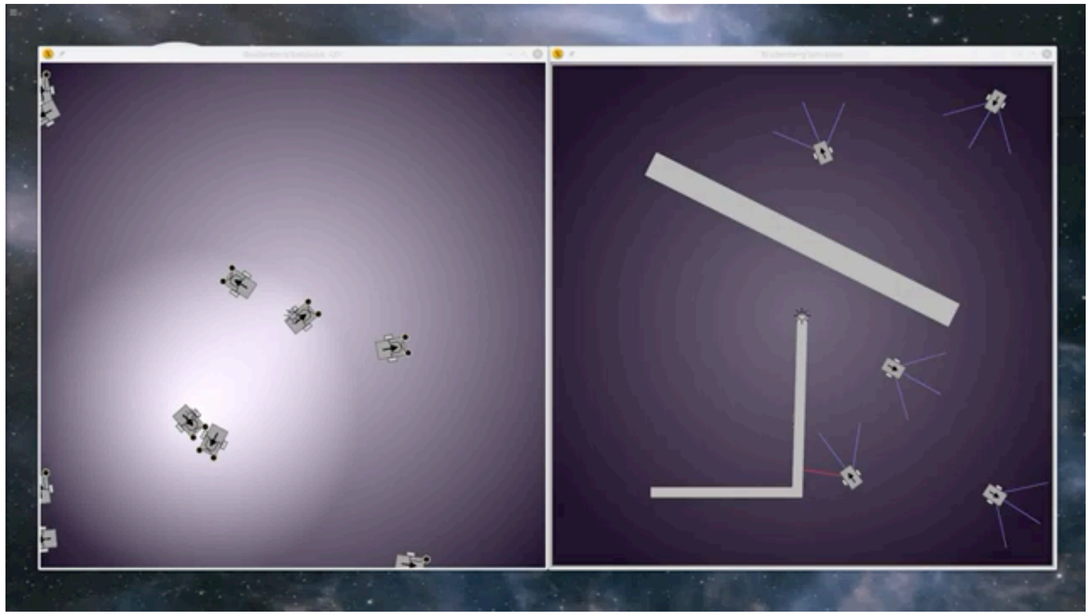
Toy models are highly idealized and deliberately simplified models that retain only the essential features of a system in order to explore specific theoretical questions. Long used in physics and other sciences, they have recently begun to play a more visible role in consciousness research. This chapter examines the potential utility of toy models for developing and evaluating scientific theories of consciousness in terms of their ability to clarify theoretical frameworks, test assumptions, and illuminate philosophical challenges. Drawing primarily on examples from Integrated Information Theory (IIT) and Global Workspace Theory (GWT), I show how these simplified systems could make abstract concepts more tangible, enabling researchers to probe the coherence, consistency, and implications of competing frameworks. In addition to supporting theory development, toy models can also address specific features of experience, as exemplified by the account of spatial extendedness and temporal flow provided by integrated information theory (IIT) and recent theory-independent structural approaches. Moreover, toy models bring philosophical debates into sharper focus, such as the distinction between functional and structural theories of consciousness. By bridging abstract claims and empirical inquiry, toy models provide essential insights into the challenges of building comprehensive theories of consciousness.

Braitenberg vehicles

Models of taxis

- Vehicles 2 and 3
- Agent with two sensors and two wheels
- Sensors and wheels connected by wires
- Implementation: (Left/right) Wheel rotational velocity = constant * (right/left) sensory reading





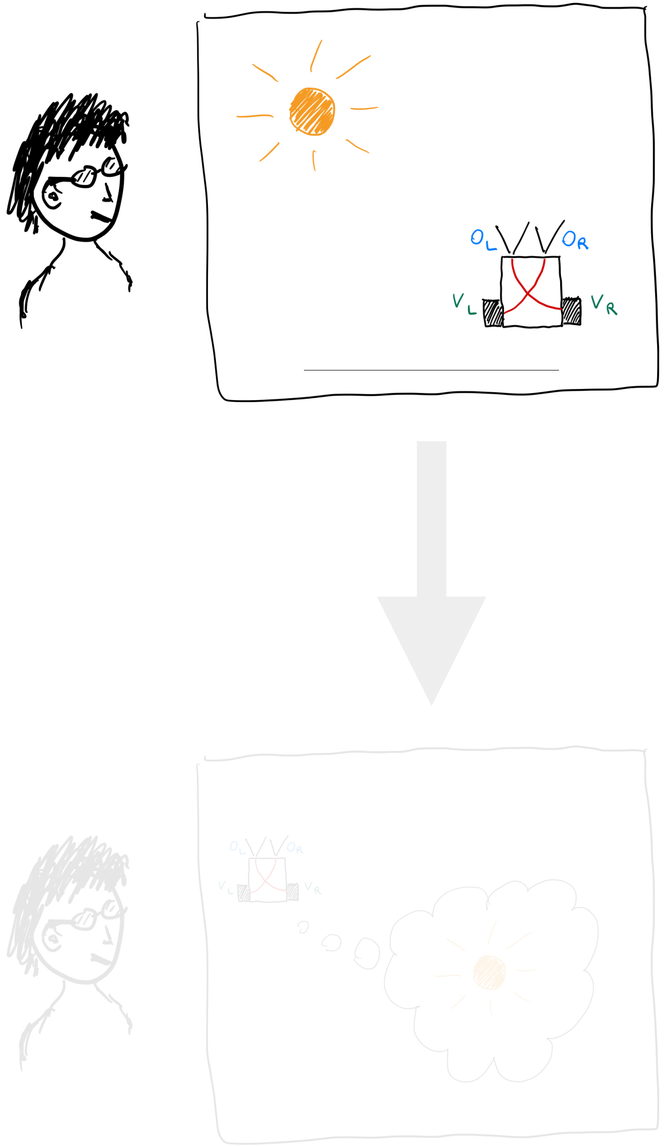
(Youtube) “Robot Simulator for Braitenberg Vehicles and Other Small Robots” - Lukas Stratmann

Braitenberg POMDP

"Objective" physics

- State space, $S := X_{\text{Centre}} \times \Theta \times X_{\text{Stim}} \times R_{\text{Body}}$
- Action space, $A := V_l \times V_r$
- Observation space, $O := O_l \times O_r$
- Transition dynamics, position $f_{\text{Centre}} : X_{\text{Centre}} \times V_l \times V_r \to X_{\text{Centre}}$ and angle $f_\Theta : \Theta \times V_l \times V_r \to \Theta$
- Observation map, deterministic, given by the composition of f_{Sensor} and f_{Stim}, with $f_{\text{Sensor}} : X_{\text{Centre}} \times \Theta \times X_{\text{Stim}} \times R_{\text{Body}} \to X_l \times X_r \times X_{\text{Stim}}$, and $f_{\text{Stim}} : X_l \times X_r \times X_{\text{Stim}} \to O_l \times O_r$.

Symbol	Description
$X_{\text{Centre}} \subseteq \mathbb{R}^2$	Centre of vehicle's body
$\Theta \subseteq (-\pi, \pi]$	Absolute heading of the vehicle
$\Theta^{\text{rel}} \subseteq (-\pi, \pi]$	Heading relative to the stimulus
$X_{\text{Stim}} \subseteq \mathbb{R}^2$	Position of the stimulus
$R_{\text{Body}} \subseteq \mathbb{R}_{\geq 0}$	Radius of the vehicle body
$V_l \subseteq \mathbb{R}, V_r \subseteq \mathbb{R}$	Left/right linear velocities of wheels
$O_l \subseteq \mathbb{R}, O_r \subseteq \mathbb{R}$	Left/right sensory readings
$R \subseteq \mathbb{R}_{\geq 0}$	Radius of polar coord. (X_{Centre})
$\Phi \subseteq (0, 2\pi]$	Angle of polar coord. (X_{Centre})



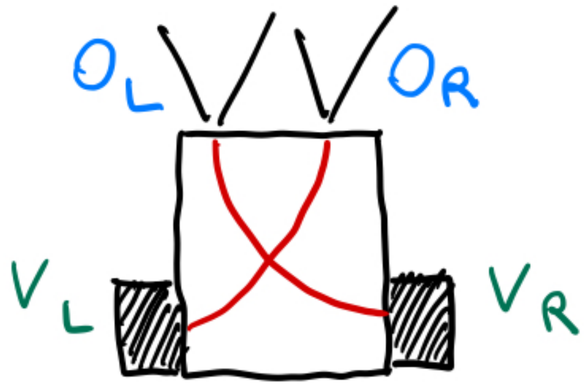
Policy

Trivial (but why?)

$$\pi : O \to A$$

$$o_r \mapsto v_l = k o_r, \quad o_l \mapsto v_r = k o_l$$

$$\frac{V_L}{O_R} = \text{constant} = \frac{V_R}{O_L}$$



Property 1: $\mathbf{SO(2)}$ symmetry

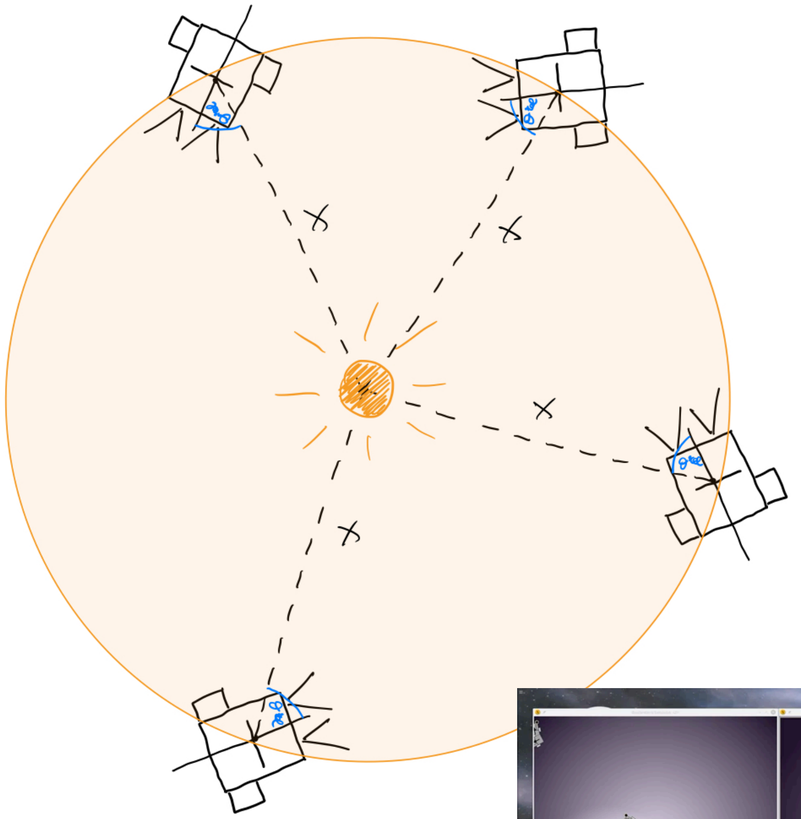
Invariance to rotations around source

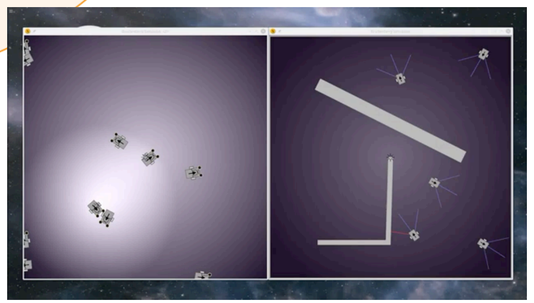
Given

- same distance to source
- same attitude (angle θ^{rel})

there is an invariance to rotations centred around the source.

Rotating the world around the light, doesn't change sensations.



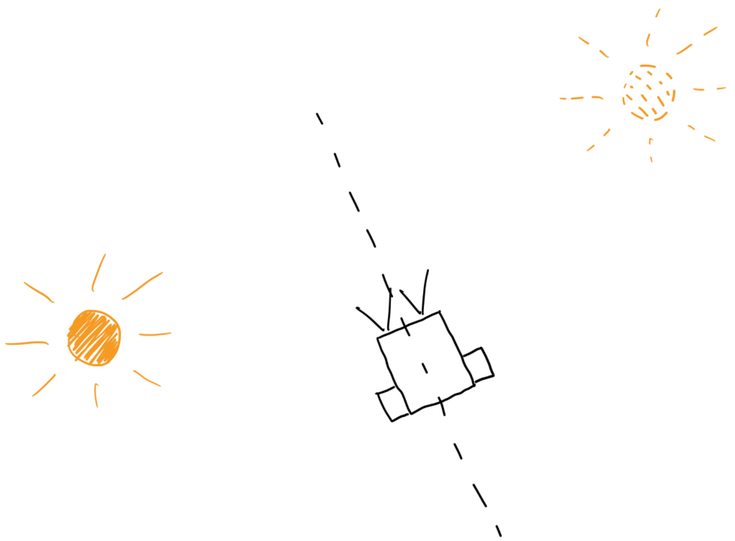


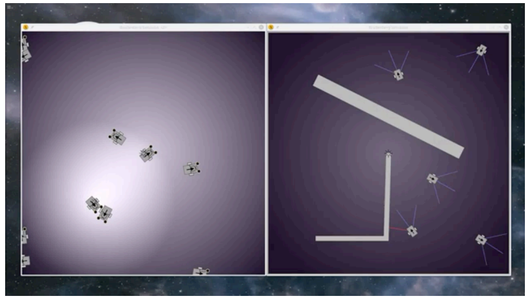
Property 2: Z_2 symmetry

Equivariance to reflections along longitudinal axis

Given the longitudinal axis of the vehicle, we, there is an equivariance to reflections about this axis.

Left-right world reflections for the vehicle flip its sensory readings.





G-equivariant POMDPs

An "objective" world with symmetries

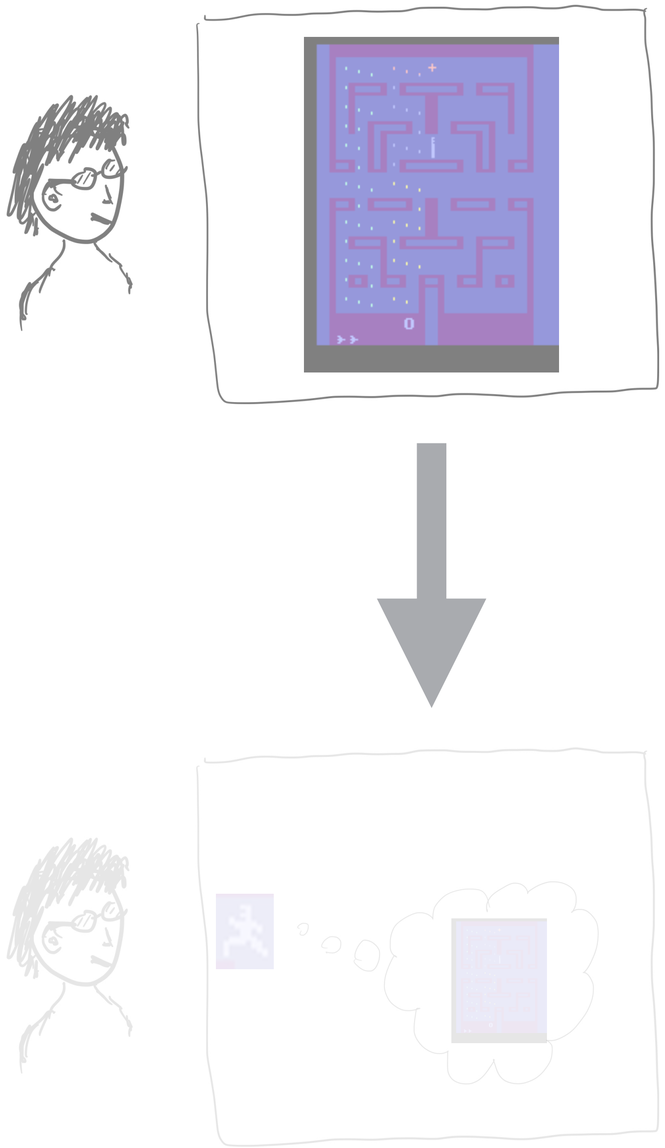
A G-equivariant POMDP $(S, A, O, T, M, G, \alpha_S, \alpha_A, \alpha_O)$ induced by a POMDP given a group G is given by

- POMDP $P = (S, A, O, T, M)$
- A group G
- Group actions $\alpha_S : G \times S \to S, \alpha_A : G \times A \to A, \alpha_O : G \times O \to O$, also written as $g \cdot s, g \cdot a, g \cdot o$

With equivariance conditions on:

  - transitions, $g_* T(- \mid s, a) = T(- \mid g \cdot s, g \cdot a)$
  - observations, $g_* M(- \mid s, a) = M(- \mid g \cdot s, g \cdot a)$

where for $x \in X, p \in \Delta(X)$, the pushforward map $g_* : \Delta(X) \to \Delta(X)$ is defined as $g_* p(x) := p(g^{-1} \cdot x)$.



Bisimulation equivalence induced by G

Compressing world using symmetries

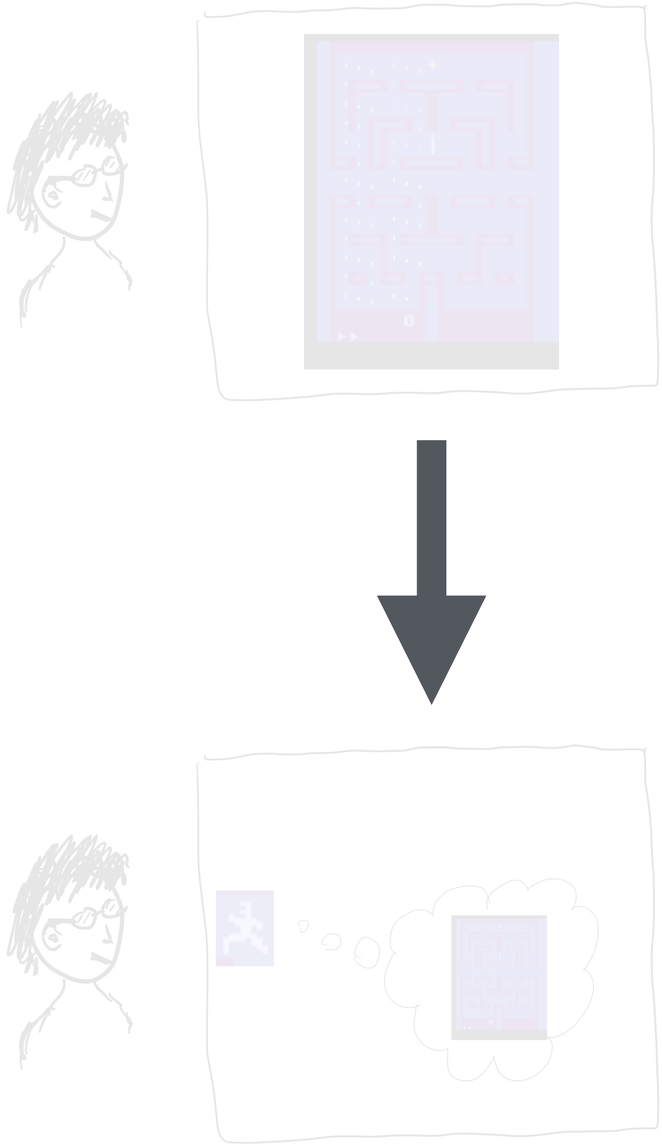
Given a G-equivariant POMDP and the orbit equivalence relation induced by the action of G on S, R_G,

$$s \sim_G \bar{s} \iff \exists g \in G : \bar{s} = g \cdot s$$

R_G is a bisimulation equivalence if the following conditions hold for $[s] \in S/R_G$:

$$\sum_{s' \in [s]} T(s' \mid s, a) = \sum_{s' \in [s]} T(s' \mid \bar{s}, a) \qquad \text{(transitions)}$$

$$M(o \mid s, a) = M(o \mid \bar{s}, a) \qquad \text{(observations)}$$



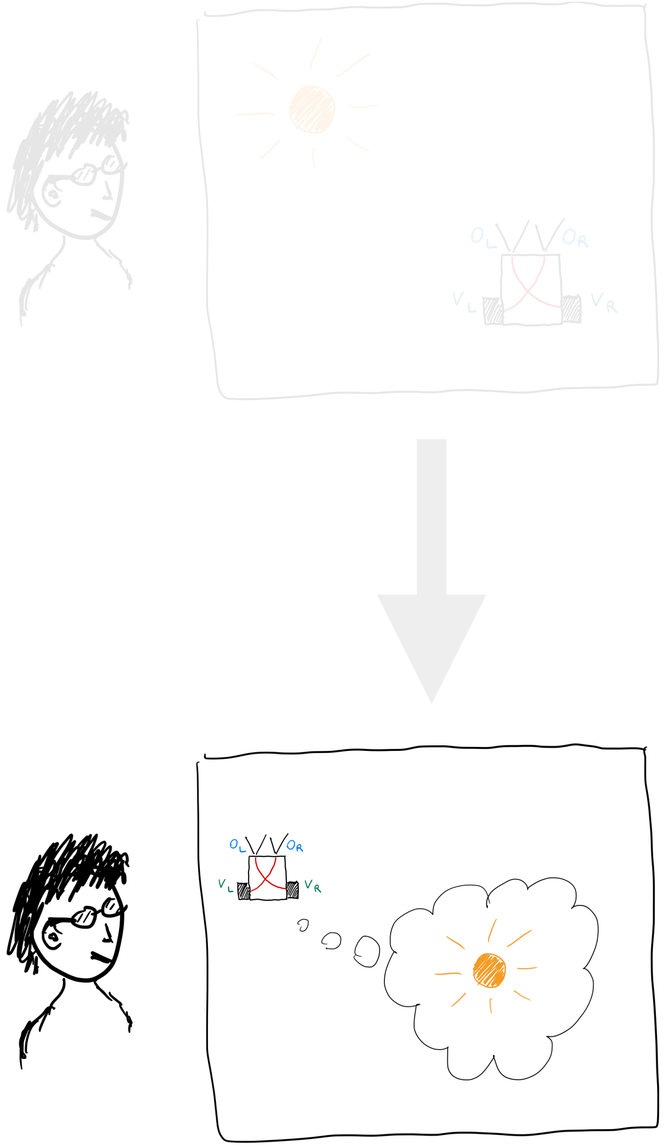
Braitenberg quotient POMDP

Subjective world of a vehicle

Symbol	Description
$X_{\text{Centre}} \subseteq \mathbb{R}^2$	Centre of vehicle's body
$\Theta \subseteq (-\pi, \pi]$	Absolute heading of the vehicle
$\Theta^{\text{rel}} \subseteq (-\pi, \pi]$	Heading relative to the stimulus
$X_{\text{Stim}} \subseteq \mathbb{R}^2$	Position of the stimulus
$R_{\text{Body}} \subseteq \mathbb{R}_{\geq 0}$	Radius of the vehicle body
$V_l \subseteq \mathbb{R}, V_r \subseteq \mathbb{R}$	Left/right linear velocities of wheels
$O_l \subseteq \mathbb{R}, O_r \subseteq \mathbb{R}$	Left/right sensory readings
$R \subseteq \mathbb{R}_{\geq 0}$	Radius of polar coord. (X_{Centre})
$\Phi \subseteq (0, 2\pi]$	Angle of polar coord. (X_{Centre})

Given the group $G = O(2) = SO(2) \rtimes Z_2$

- State space, $S = S/R_G := R \times \Theta^{\text{rel}} \times X_{\text{Stim}} \times R_{\text{Body}}$ (keep attitude Θ^{rel}, forget absolute angle Θ)
- (Same) Action space, $A := V_l \times V_r$
- (Same) Observation space, $O := O_l \times O_r / \{(o_l, o_r) \sim_G (o_r, o_l)\}$
- Transition dynamics, radius $f_R : R \times V_l \times V_r \to R$ and attitude $f_{\Theta^{\text{rel}}} : \Theta^{\text{rel}} \times V_l \times V_r \to \Theta^{\text{rel}}$
- Observation map, deterministic, given by the composition of $f_{\text{Sensor}} : R \times \Theta^{\text{rel}} \times X_{\text{Stim}} \times R_{\text{Body}} \to R_l \times R_r \times X_{\text{Stim}}$, and $f_{\text{Stim}} : R_l \times R_r \times X_{\text{Stim}} \to O_l \times O_r / \{(o_l, o_r) \sim_G (o_r, o_l)\}$.



A vehicle's "qualia"?

Relation among light, vision and $O(2) = SO(2) \rtimes Z_2$

Qualia modalities (group G):

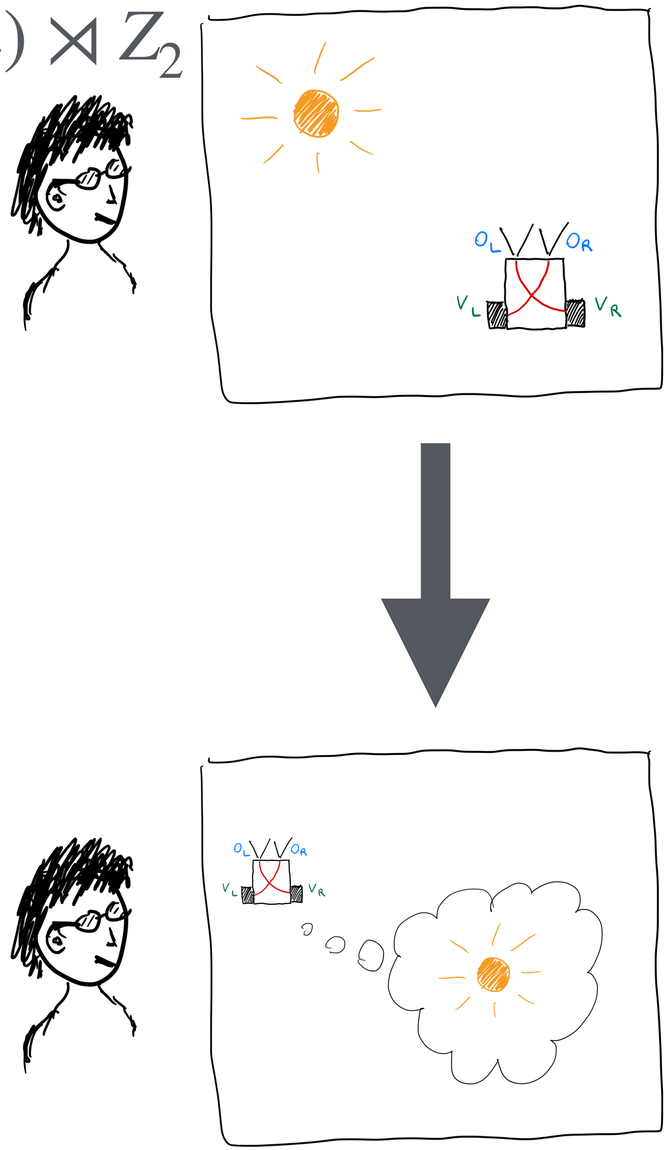
- $SO(2)$ —> rotations around the light
- Z_2 —> reflections about longitudinal direction

Qualia signatures (orbits $[s]$):

- $[s] \in S/R_{SO(2)}$ —> orbits are circles with same sensory readings centred around light
- $[s] \in S/R_{Z_2}$ —> orbits are half-planes given by longitudinal axis of vehicles

Qualia attributes (elements of orbits s):

- $s \in [s] \in S/R_{SO(2)}$ —> position on circles
- $s \in [s] \in S/R_{Z_2}$ —> position in a half-plane



Today the earwig, tomorrow man?

Options

- Toy model only (a simulation of water doesn't make *you* wet)
- Taking seriously consciousness in artificial systems, even small ones
- Mixing measurement/observation with qualia
- …?

My take

These algebraic structures can provide a *proxy* for subjective experience, maybe Umwelt.

They can capture what *cannot* be part of experience, because agents are insensitive to it.

Let's not take toy models too seriously, but also not just discard them.

On the utility of toy models for theories of consciousness

Larissa Albantakis[1,*]

[1] Department of Psychiatry, University of Wisconsin, Madison, WI, USA
* albantakis@wisc.edu

Abstract

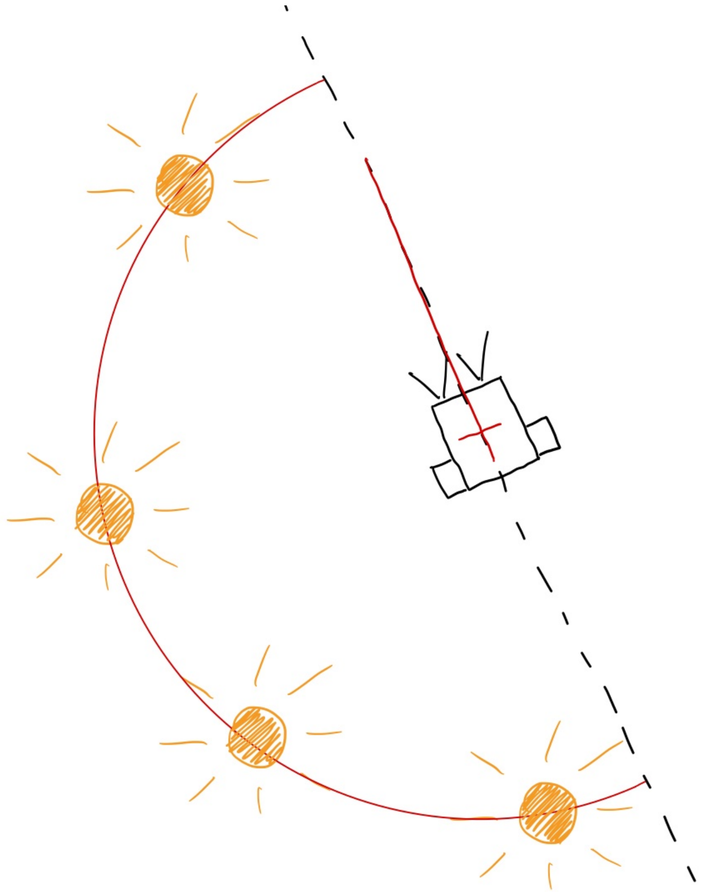
Toy models are highly idealized and deliberately simplified models that retain only the essential features of a system in order to explore specific theoretical questions. Long used in physics and other sciences, they have recently begun to play a more visible role in consciousness research. This chapter examines the potential utility of toy models for developing and evaluating scientific theories of consciousness in terms of their ability to clarify theoretical frameworks, test assumptions, and illuminate philosophical challenges. Drawing primarily on examples from Integrated Information Theory (IIT) and Global Workspace Theory (GWT), I show how these simplified systems could make abstract concepts more tangible, enabling researchers to probe the coherence, consistency, and implications of competing frameworks. In addition to supporting theory development, toy models can also address specific features of experience, as exemplified by the account of spatial extendedness and temporal flow provided by integrated information theory (IIT) and recent theory-independent structural approaches. Moreover, toy models bring philosophical debates into sharper focus, such as the distinction between functional and structural theories of consciousness. By bridging abstract claims and empirical inquiry, toy models provide essential insights into the challenges of building comprehensive theories of consciousness.

Special cases

Angle-insensitivity

Symmetry variations (1)

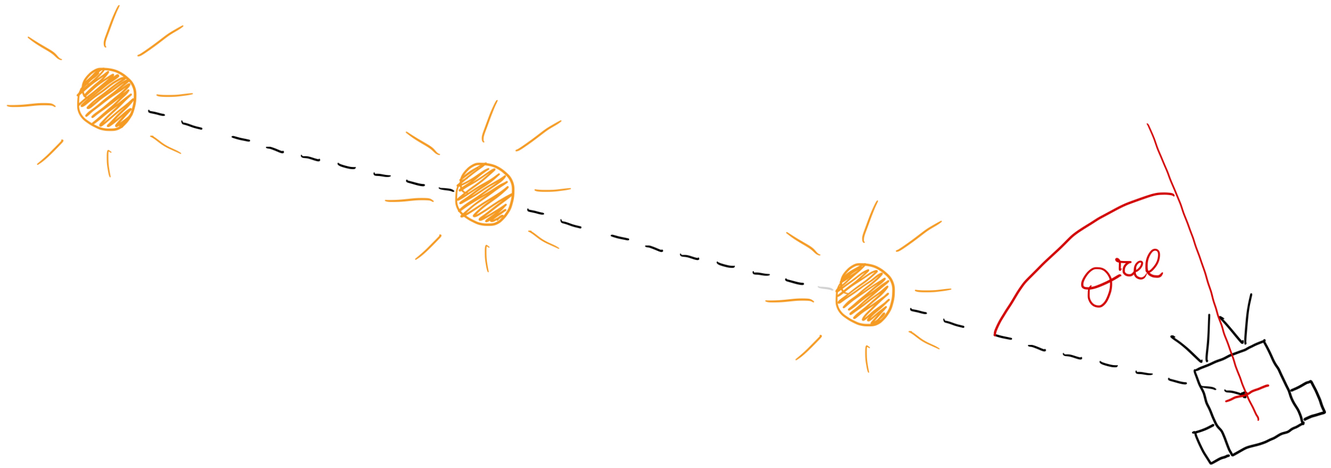
- $G = O(2)$
- Group action on observations is trivial: gives only one value per side
- Maybe a “better" group that doesn’t rely on this trivial group action?



Distance-insensitivity

Symmetry variations (2)

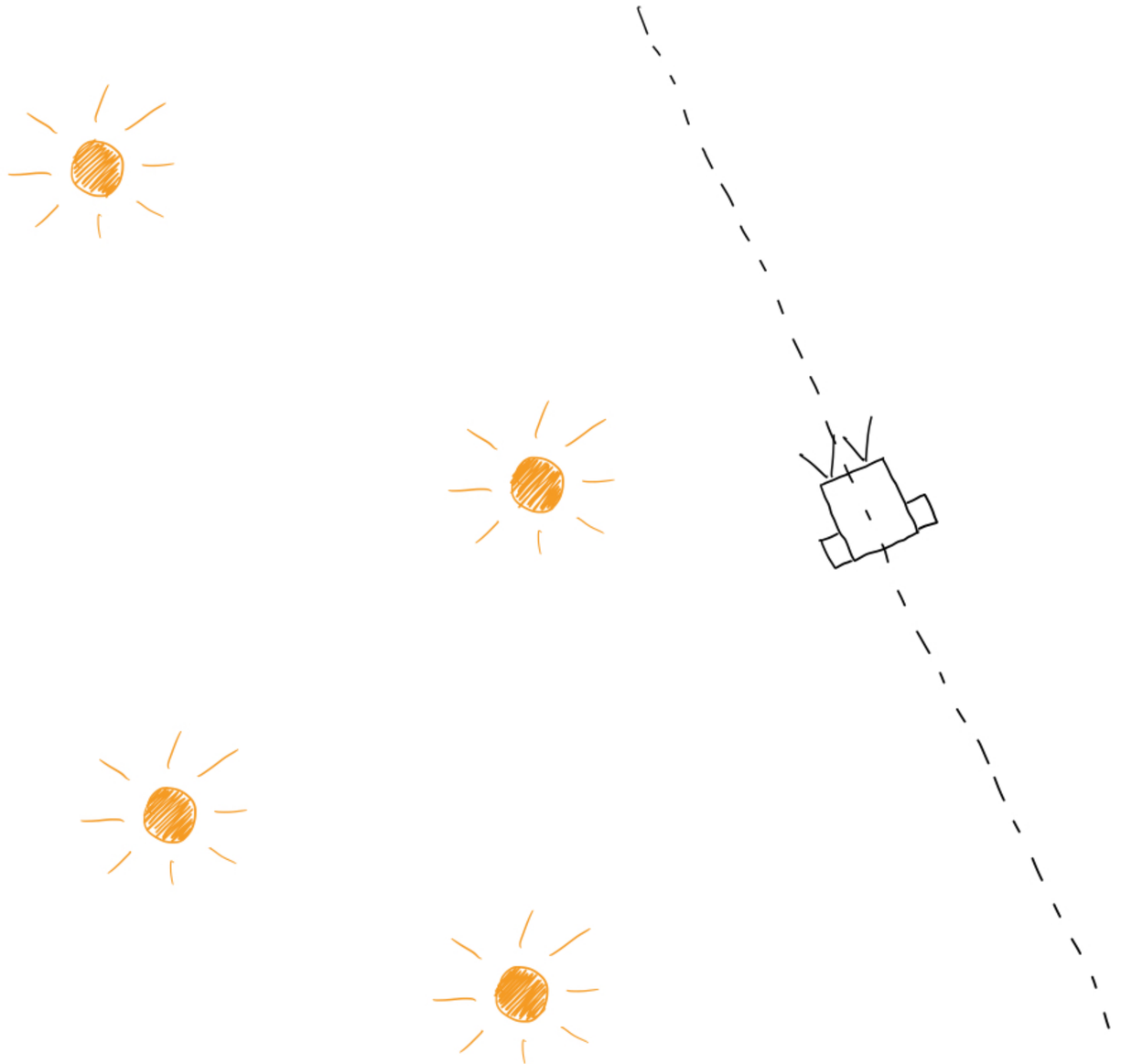
- $G = O(2) \times \mathbb{R}_{>0}$
- Invariance for $\mathbb{R}_{>0}$ and $SO(2)$, equivariance for Z_2



Distance- + angle-insensitivity

Symmetry variations (3)

- $G = ?$

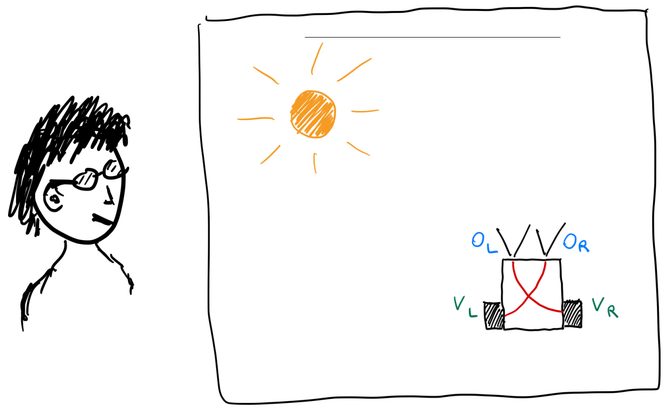


More rants

“Objective”?

Is the objective physical world “objective” here? Objective for an observer?

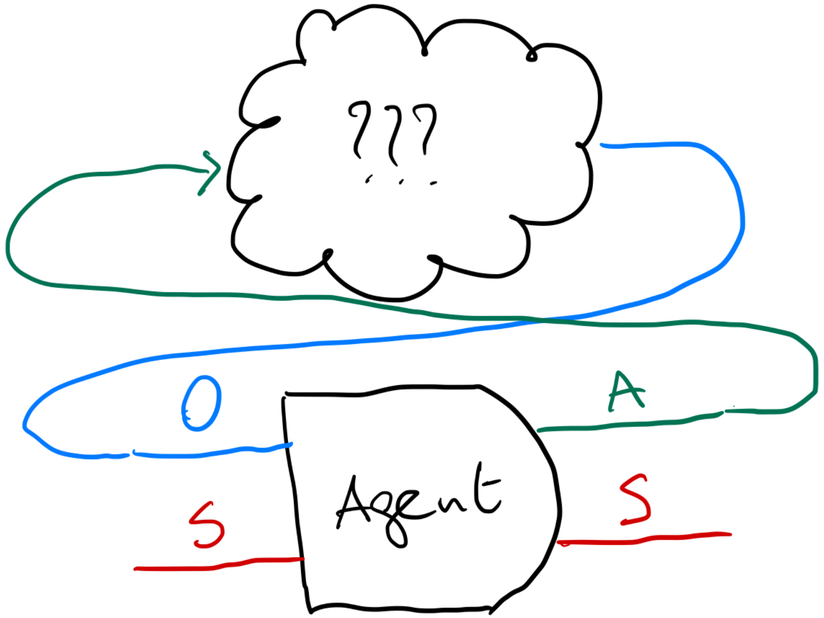
- Are there other alternative world models that we could start from?
- What is the relation between world models?

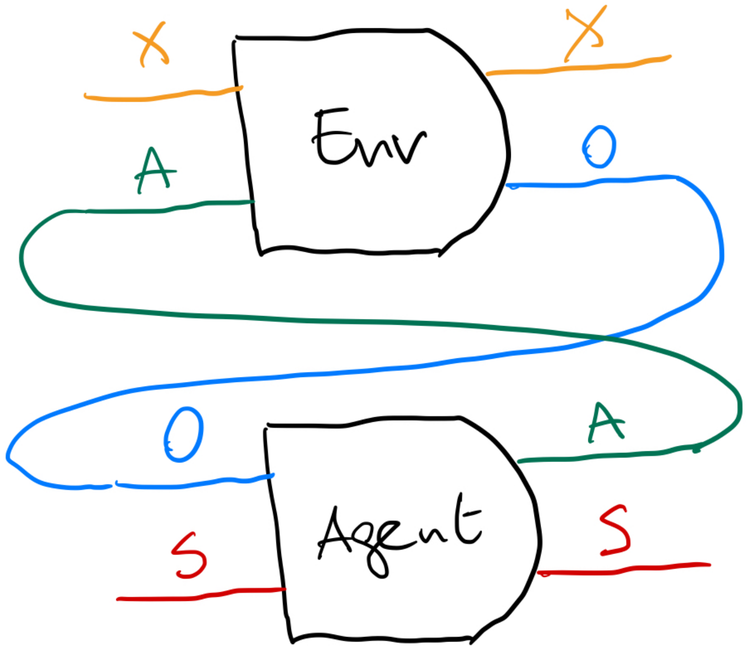


What does a vehicle believe in?

Or rather, what can we interpret it to believe in?

Technically: bisimulations of partially observable processes are not at all trivial, many degrees of freedom (see “angle-insensitivity”)





Qualia vs umwelt?

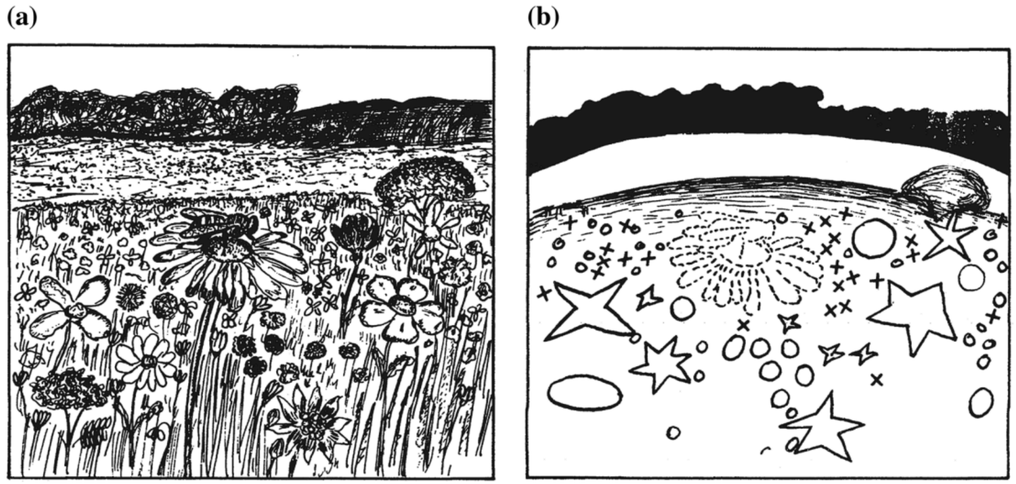


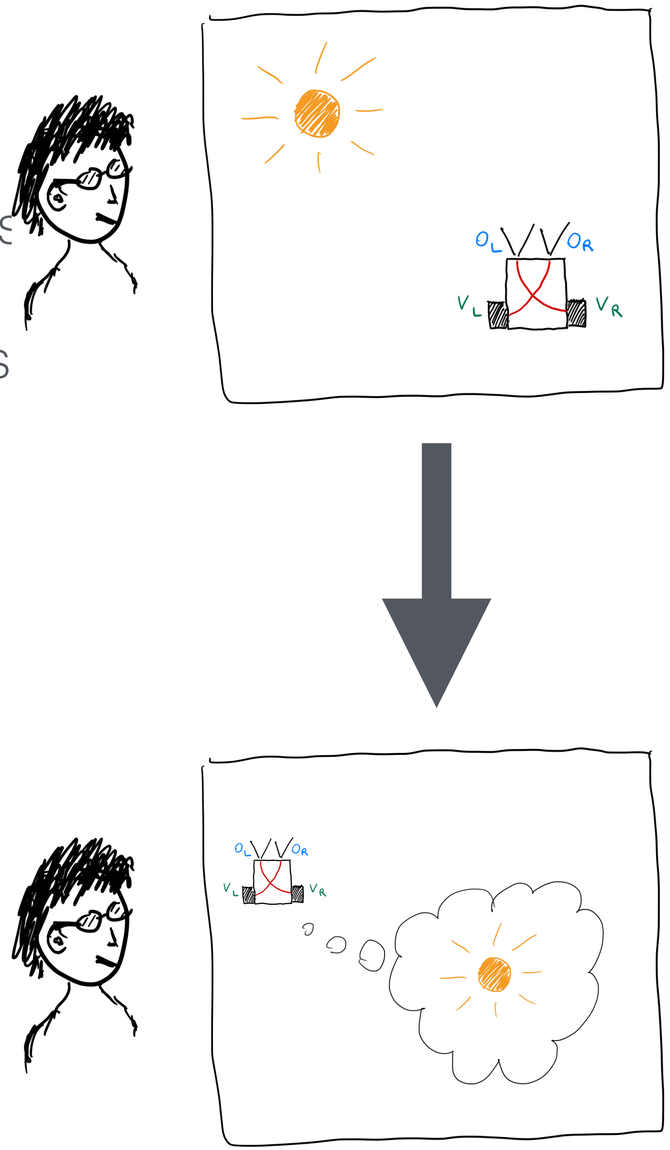
Fig. 1 The *Umwelt* of a bee as illustrated in Von Uexküll (1934). **a** The environment of a bee how we perceive it as an external observer. **b** The same bee perceives only particular aspects of the same world, which constitute its *Umwelt*

Future work?

- We need a ~~category of~~ lot of world models (next)
  - *Minimal* models
  - Relations between models
  - Qualia as capturing some minimality property?
  - …

Summary

- Agent's interface: agents are time-dependent open systems
- Recipe to capture an observer's interpretation of an agent's
- Studying qualia in a toy model using equivalence relations
- ...or maybe "just" an agent's Umwelt



Acknowledgments

For this work: **Fernando Rosas** (if I'm right, otherwise that's on me).

For related works: **Nathaniel Virgo, Martin Biehl, Matteo Capucci**.

Funding: This work was supported by JST Moonshot R&D Grant Number JPMJMS2012.