Challenges and Prospects for the Projective Consciousness Model.

Gregoire Sergeant-Perthuis

LCQB Sorbonne Université

CCHM 24 Oxford, January 12, 2024

Projective Consciousness Model: Motivation

A project that has been ongoing for several years, and you can find updates at:

- The project website:
  `www.gregoiresergeant-perthuis.com/PCM.html`

$\rightarrow$ Focus on phenomenological aspects of consciousness

$\rightsquigarrow$ Computational Phenomenology

Some traits of Consciousness

Consciousness involves a subjective perspective, characterized by viewpoint-structured organization, a sense of unity (holistic world), embodiment, and an internal representation of the world in perspective from a specific standpoint.

$\longrightarrow$See K. Williford's MoC4 presentation of phenomelogical motivation
`www.youtube.com/watch?v=eHyVZWZMqzg&t=853s`

Projective Consciousness Model : Goals

Goals:

1. Implement a subjective perspective for *artificial* adaptive agents
2. In a way compatible with axioms of consciousness
   - Alexander's axioms: internal world model, imagination, attention, planing, emotion [Ale05]
   - Information integration

Projective Consciousness Model: (Some of the) Founding Papers

- **A Mathematical Model of Embodied Consciousness.** D. Rudrauf, D. Bennequin, I.Granic, G. Landini, K. Friston, and K. Williford, *Journal of Theoretical Biology (JTB), 2017.*[RBG$^+$17]
- **The Projective Consciousness Model and Phenomenal Selfhood.** K. Williford, D. Bennequin, K. Friston, D. Rudrauf, *Front. Psychol 2018.* [WBFR18]
- **The Moon Illusion Explained by the Projective Consciousness Model.** D. Rudrauf, D. Bennequin, and K. Williford, *JTB, 2020.* [RBW20]

Focus: *Projective geometry and perception.*

Key Result

The model's strength is demonstrated through the prediction and testing on a perceptual illusion: the moon illusion.

PCM Founding Fathers

D. Rudrauf[1]

K. Williford[2]

D. Bennequin[3]

K. Friston[4]

1 Pr. Cognitive Science, CIAMS, Université Paris Saclay

2 Pr. Philosophy, Philosophy & Humanities, The University of Texas at Arlington

3 Pr. Mathematics, IMJ-PRG, Université Paris-Cité

4 Pr. Theoretical Neuroscience, Institute of Neurology, University College London

A Step Toward Robotics: Imbuing Perspective Taking in Social Agents

$\rightsquigarrow$ Where I Joined the Project.

Context: World Models. recall Bayesian Brain Hypothesis:

- Adaptive systems have evolved to preserve their integrity, necessitating the prediction of environmental behavior.
- How? By forming hypotheses about the world and updating them with new observations.
- Why? To assist in decision-making regarding how to act.
- Formally:
  - Karl Friston introduced Active Inference [FKH06][DPS$^+$20].
  - Markov Decision Process (MDP), POMDP.

Imbuing Perspective Taking in Social Agents: Challenge

Problem: Associate actions, the movement of an agent with a perspective taking (projective transformation).
Solving Subproblem: How to relate the configuration of the agent and projective perspectives on the environment?

$\rightarrow$ *impose a set of axioms (in the agent's frame of reference)*:

1. The subject is centered at 0 after the projective transformation, i.e., in its perspective, it is at the center of its frame.
2. The three axes x, y, z are preserved. The axes of the Euclidean frame associated with the agent (up-down, left-right, and back-front) must be preserved after the projective transformation.
3. No points in ambient space appear, to the subject, to be truly at infinity; this constraint is satisfied when the subject only directly represents a half space.
4. Objects that are near the agent appear to have the same size as in the reference Euclidean frame.

Imbuing Perspective Taking in Social Agents: Results

Axioms Impose a Set of Projective Transformations (Proposition A.1 [RSPB$^+$22])
As a consequence, a justification for Stevens' psychophysical distance.

Resulting Numerical experiments: maldaptive behaviors, theory of mind games:

- **Modeling the Subjective Perspective of Consciousness and Its Role in the Control of Behaviors.**
  D. Rudrauf, G. Sergeant-Perthuis, O. Belli, Y. Tisserand, G. Di Marzo Serugendo
  Journal of Theoretical Biology, 2022. [RSPB$^+$22]

- **Combining the Projective Consciousness Model and Virtual Humans to Assess Theory of Mind Capacity in Virtual Reality: A Proof-of-Concept.**
  D. Rudrauf, G. Sergeant-Perthuis, Y. Tisserand, T. Monnor, O. Belli
  ACM Transactions on Interactive Intelligent Systems, 2022. [RSPT$^+$23a]

Videos

PCM what we have:

Sum up:

- Personal Perspective on the ‘real’ world
- Internal world model, workspace
- Attention
- Emotion
- Epistemic drive : acting in order to reduce uncertainty

But also:

- Imagination : taking other’s perspective
- Empathy: simulating other agents’ perspective

One Step Further: Towards Group Structured World Models for Adaptive Agents

- Building a world model/internal space that incorporates the actions and 'perspectives' of agents within the agent's internal space.

$\rightarrow$ Assumption: the space is designed to be a reconstruction for the sensory input.

$\rightsquigarrow$ Thought experiment: Imagine no object in the world, but imagine that you can still move around, then no point is singled out.

- Assumption: Building such a space plays a key role in stable/robust inference and planning of actions.
- Modelization: Imbue the world model with a group acting on it.

One Step Further: Towards Group Structured World Models for Adaptive Agents

Important differences compared to the previous approach:

$\rightarrow$ *Inclusion of 'perspectives' in the broader context of charts.*

$\rightarrow$ *No limitation on the group: Not restricted to a projective group.*

Formalized in:

- **The Projective Consciousness Model: Projective Geometry at the Core of Consciousness and the Integration of Perception, Imagination, Motivation, Emotion, Social Cognition, and Action.**
  D. Rudrauf, G. Sergeant-Perthuis, Y. Tisserand, G. Poloudenny, K. Williford, and M-A. Amorim, *Brain Sciences, 2023.* [RSPT$^+$23b]

Definitions of MDP and POMDP

Definition (Markov Decision Process: Definition 1 [YW21])

A Markov Decision Process, is a collection $\langle S, A, T, r \rangle$ where,

- S is the set of configurations of the environment
- A is the collection of actions of the agent
- $T : S \times A \to S$ is the transition probability; it captures the consequences of the action $a \in A$ of the agent on the environment that changes from s_t to s_{t+1}
- $r : S \times A \times S \to \mathbb{R}$; it is the reward function for an action $a \in A$ and two states (s, s') thought of as s_t and s_{t+1}.

Definitions of MDP and POMDP

Definition (Partially Observable Markov Decision Process [Kur22])

A POMDP is defined as a tuple $\langle S, A, T, r, O, Z \rangle$, where $\langle S, A, T, r \rangle$ is an MDP and,

- O is the set of possible observations.
- Z is the observation kernel, $Z : S \times A \to O$, which specifies the probability of observing a particular observation given the current state and action.
- r is a reward function which domain is $S \times A$; $r : S \times A \to \mathbb{R}$.

Group-structured world model

- Input: world model accounts for point of view

Perspectives	$\rightsquigarrow$	**Group G**
World model	$\rightsquigarrow$	**G-space**

Definition (Group-structured space, G-space)

S is a group-structured space for the group G when there is a map $h : G \times S \to S$ denoted as $h(g, s) = g.s$ for $g \in G$ and $s \in S$, such that,

1. $(g.g_1).s = g.(g_1.s)$ for all $g, g_1 \in G$, $s \in S$
2. $e.s = s$, for all $s \in S$

For a given group G, such space is called a G-space.

POMDP with group structured state space

We will call a MDP with group-structured state space, a MDP where the state space S is a $G-$space, for some group G, and a subset of the set of actions is the group G.

Definition (MDP and POMDP with group-structured state space)

A MDP with a group-structured state space is a tuple $\langle S, A, T, r, G\rangle$ where G is a group and $\langle S, A, T, r\rangle$ is a MDP that satisfies the following properties:

- S is a $G-$space
- G is subset of the set of actions A,
- for all $g \in G$, $T(s'|s, g) = 1[s' = g.s]$

A POMDP with a group-structured state space is a tuple $\langle S, A, T, r, O, Z, G\rangle$ where $\langle S, A, T, r, G\rangle$ is a group-structured MDP (structured by G) and $\langle S, A, T, r, O, Z\rangle$ is a POMDP.

Comparing two cases: Euclidean and projective

In our example: 2 cases

1. (S,G):= Euclidean space, affine transformations
2. (S,G):= Projective space, projective transformations

Theorem

Let us assume that staying still is always a possible move for the agent.

1. *Euclidean case: when the agent has an objective representation of its environment, given by an affine map, the agent stays still.*
2. *Projective case: Assume now that the set of moves M is finite; assume furthermore that after any possible move, the agent faces O, in other words, we assume that the agent knows in which direction to look in order to find the object but is still uncertain on where the object is exactly. If it has a 'subjective' perspectives, i.e. its representation is given through a projective transformation, it will choose the moves that allows it to approach O (for any ϵ small enough).*

Numerical simulations

Numerical simulations confirm predictions, made by N. Ruet (Ph.D. Student)

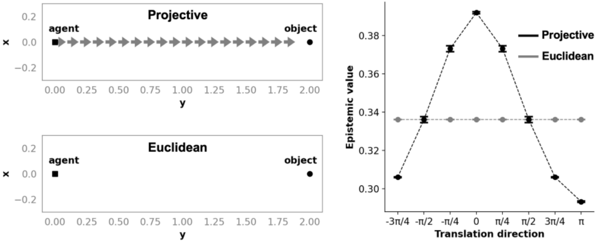

Figure 3: Simulation results

Left-tier. Trajectory of the agent for the projective *versus* Euclidean internal spaces. *Right-tier.* Epistemic value as a function of directions of translation with respect to the object direction, for the projective *versus* Euclidean internal spaces. Points are average values across comparable directions, and error-bars are standard errors.

Group structured world models: Why?

Answer coming from machine learning perspective: *a form of inductive bias*

- Better adaptation to unseen configurations of objects
- Better action planning.

Some challenges: Planning, learning group structured world models

On the optimal control side:

- Downside: Exact Filtering, Planning with theroetical guaranties $\rightsquigarrow$ difficult problem. Working on it T. Yan (Ph.D. student)
- Upside: Approximate method, EKF, UKF can be simplified see [SPRR$^+$23]

On learning world models:

- Optimizing over possible group actions is a challenge.

$\rightsquigarrow$ Example learning by maximum likelihood

Learning Group Actions on Bayesian Networks

- Want to learn a Bayesian network from data with different points of view, influenced by the agent's actions.
- Bayesian network $\implies$ factorization of probability distributions over variables (factor graph, Markov random field): $\mathscr{A}$ collection of subsets of I, $\mathscr{A} \subseteq \mathscr{P}(I)$.
- Each variable X is assumed to have a group acting on it, but the action is *undetermined* (has to be learned).

$$P_{X_i, i \in I}(x) = \prod_{a \in E} f_a(x_a)$$

Learning Group Actions on Bayesian Networks

- **Input:** $X_a^j, a \in \mathscr{A}, j = 1 \ldots N$ and perspectives $g^j, j = 1 \ldots N$

Parameter Model: not learned

- A group G
- A model of interaction $\mathscr{A}$

Learned

- $f_a, a \in \mathscr{A}$
- $\psi : G \times X \to X$ denoted as $\psi(g, x) = g.x$

Model: $P(g, X_I) \simeq g_* \prod_{a \in \mathscr{A}} f_a$ (uniform prior on G)

Method: Maximum likelihood, $l(f, \psi) := \sum_{j=1 \ldots N} \ln g_*^j \prod_{a \in \mathscr{A}} f_a(x_a^j)$.
$\rightsquigarrow$ Ongoing work.

Thank you very much for your attention

Thank you very much for your attention!

References I

- Igor Aleksander, *The World in My Mind, My Mind in the World: Key Mechanisms of Consciousness in People, Animals and Machines*, Exeter, Imprint Academic, 2005.
- Lancelot Da Costa, Thomas Parr, Noor Sajid, Sebastijan Veselic, Victorita Neacsu, and Karl Friston, *Active inference on discrete state-spaces: A synthesis*, Journal of Mathematical Psychology **99** (2020), 102447.
- Karl Friston, James Kilner, and Lee Harrison, *A free energy principle for the brain*, Journal of Physiology Paris **100** (2006), no. 1-3, 70–87.
- Hanna Kurniawati, *Partially observable markov decision processes and robotics*, Annual Review of Control, Robotics, and Autonomous Systems **5** (2022), no. 1, 253–277.

D. Rudrauf, D. Bennequin, I. Granic, G. Landini, , K. Friston, and K. Williford, *A mathematical model of embodied consciousness*, Journal of Theoretical Biology (2017).

David Rudrauf, Daniel Bennequin, and Kenneth Williford, *The Moon illusion explained by the projective consciousness model*, Journal of Theoretical Biology (2020).

D. Rudrauf, G. Sergeant-Perthuis, O. Belli, Y. Tisserand, and G. Di Marzo Serugendo, *Modeling the subjective perspective of consciousness and its role in the control of behaviours*, Journal of Theoretical Biology **534** (2022), 110957.

References III

D. Rudrauf, G. Sergeant-Perhtuis, Y. Tisserand, T. Monnor, V. De Gevigney, and O. Belli, *Combining the projective consciousness model and virtual humans for immersive psychological research: A proof-of-concept simulating a tom assessment*, ACM Trans. Interact. Intell. Syst. **13** (2023), no. 2.

David Rudrauf, Grégoire Sergeant-Perthuis, Yvain Tisserand, Germain Poloudenny, Kenneth Williford, and Michel-Ange Amorim, *The projective consciousness model: Projective geometry at the core of consciousness and the integration of perception, imagination, motivation, emotion, social cognition and action*, Brain Sciences **13** (2023), no. 10.

References IV

Grégoire Sergeant-Perthuis, Nils Ruet, David Rudrauf, Dimitri Ognibene, and Yvain Tisserand, *Influence of the geometry of the feature space on curiosity based exploration*, NeurIPS 2023 workshop: Information-Theoretic Principles in Cognitive Systems, 2023.

Kenneth Williford, Daniel Bennequin, Karl Friston, and David Rudrauf, *The projective consciousness model and phenomenal selfhood*, Frontiers in Psychology (2018).

Yaodong Yang and Jun Wang, *An overview of multi-agent reinforcement learning from game theoretical perspective*, 2021.