The role of consciousness in social active inference and the emergence of adaptive and maladaptive behaviours

Gregoire Sergeant-Perthuis

MMER
Université de Genève

Seminar Istituto di Fisiologia Umana
University of Palermo, September 29 2021

Modes of consciousness: Motivation

'[...] if one treats consciousness naturalistically, as an empirical construct like any other, subjective experience simply becomes a brain representation that is supported in a certain way by certain brain regions [...]' Baars (Chapter 16 'The Blackwell Companion to Consciousness' [SV17]

Projective Consciousness Model: Previous work

Principles of PCM:

- Rudrauf, D., Bennequin, D., Granic, I., Landini, G., Friston, K., Williford, K. (2017) A mathematical model of embodied consciousness. Journal of Theoretical Biology. 428 [RBG+17]
- Williford, Kenneth, Daniel Bennequin, Karl Friston, and David Rudrauf. "The projective consciousness model and phenomenal selfhood." Frontiers in Psychology 9 (2018): 2571.[WBFR18]

Mathematical formulation of the moon illusion: Rudrauf, D., Bennequin, D., and Williford, K. (2020). The Moon Illusion explained by the Projective Consciousness Model. Journal of Theoretical Biology, 110455 [RBW20]

Projective Consciousness model: Contribution

Develop adaptive agents with a conscious like global workspace
(in interaction with other agents)

'Modeling the subjective perspective of consciousness and its role in the control of behaviours' [RSPB+20]
with D. Rudrauf, O. Belli, Y. Tisserand, G. Di Marzo Serugendo

Bayesian Brain Hypothesis

Adaptive systems: Bayesian Brain Hypothesis

Bayesian Brain Hypothesis I

- Adaptive systems have evolved in such a way that they can preserve their integrity and to do so they must predict the behaviour of their environment.
- How? Making hypothesis on the world and updating them with new observations
- Why? Help them decide how to act

Bayesian Brain Hypothesis II: Free Energy Principle

Karl Friston introduced Active inference [FKH06][DPS+20]: loop of inference and action,

- State space:
  1. S set of sensory inputs
  2. I possible states the agent believes the world can be in (internal to the agent), causes of sensory information
  3. M set of moves
- Generative model $P \in \mathbb{P}(S \times I)$: how states are expected to impact sensors

Recall that a generative model over $S \times I$ is a probability distribution,

$$\sum_{s \in S, i \in I} P(s,i) = 1$$

Bayesian Brain Hypothesis II: Free Energy Principle II

Inference: given an observation $s \in S$ what are the possible causes?

$$P_{|s}(i|s) = \frac{P(i,s)}{P(s)} \qquad \text{(Posterior)}$$

The agent has a set of hypothesis (its internal state) with respect to possible states of the world, a set of contexts $(Q_\gamma \in \mathbb{P}(I), \gamma \in \Gamma)$ that it needs to fit to the observation:

$$\gamma^* = \mathbf{argmin}_{\gamma \in \Gamma} \sum_{i \in I} Q_\gamma(i) \ln \frac{Q_\gamma(i)}{P(i,s)} \qquad \text{(Free Energy)}$$

It can be rewritten as,

$$\sum_{i \in I} Q_\gamma(i) \ln \frac{Q_\gamma(i)}{P(i,s)} = \mathrm{DKL}(Q_\gamma \| P_{|s}) - \ln P(s) \qquad \text{(KL divergence)}$$

Where $\mathrm{DKL}(Q_\gamma \| P_{|s}) = \sum_{i \in I} Q_\gamma(i) \ln \frac{Q_\gamma(i)}{P_{|s}(i|s)}$.

Action: Once γ^* is chosen, how to act?
Moves change sensory information, and the best move is such that,

$$m^* = \mathbf{argmin}_{m \in M} \sum_{i \in I} Q_{\gamma^*}(i) \ln \frac{Q^*_\gamma(i)}{P(i, s(m))} \tag{0.1}$$

In fact,

$$\sum_{i \in I} Q_\gamma(i) \ln \frac{Q_\gamma(i)}{P(i, s(m))} \geq -\ln P(s(m))$$

The optimization problem can be rewritten as:

$$m^* = \mathbf{argmin} \sum_{i \in I} Q_{\gamma^*}(i) [-\ln P(s(m)|i)]$$

Active inference: Summary

In the inference step a state $\gamma \in \Gamma$ is induced by sensory input s by minimizing a cost function $c : S \times \Gamma \to \mathbb{R}$,

$$\gamma^* = \mathbf{argmin}_{\gamma \in \Gamma} c(s, \gamma) \tag{0.2}$$

and during the action selection step, the subject chooses the action according to a second cost function $c_1 : \Gamma \times M \to \mathbb{R}$,

$$m^* = \mathbf{argmin}_{m \in M} c_1(m, \gamma^*) \tag{0.3}$$

which in turn induces a change at the level of the sensory input, since the environment reacts to this action.

Projective Consciousness Model: Remark

⚠ In our model action differs from the Free Energy Principle.

$\longrightarrow$ Action driven by preferences

Projective geometry for modeling consciousness

Projective Consciousness Model I: Goals

Goals:

1. Implement a subjective perspective for 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 II: Method

How?

1. Definition of an internal space: the Field of Consciousness (internal world, workspace)
2. Projective geometry for linking the ‘real’ world to the Field of Consciousness (Attention, Information integration)
3. Define subjective affective quantities from information in the Field of Consciousness

Projective geometry I

1. 3-d Euclidean space $E_3 = \mathbb{R}^3$
2. 3-d Projective space $P_3(\mathbb{R})$, the set of lines of $\mathbb{R}^4$
   - a line of direction $v \in \mathbb{R}^4 \setminus \{0\}$: $\{\lambda v | \lambda \neq 0\}$
3. E_3 as a (open) subset of $P_3(\mathbb{R})$
   - $(\lambda_1, \lambda_2, \lambda_3) \in E_3 \to (\lambda_1, \lambda_2, \lambda_3, 1) \in P_3(\mathbb{R})$

Projective geometry II

- Change of perspective: a projective transformation $P_3(\mathbb{R}) \to P_3(\mathbb{R})$
- A projective transformation is a 4×4 invertible matrix M:
- When seen in the subsets E_3, the transformation can be written as:

$$\phi(\lambda_1, \lambda_2, \lambda_3) = \left(\frac{M(\lambda_1, \lambda_2, \lambda_3, 1)[1]}{M(\lambda_1, \lambda_2, \lambda_3, 1)[4]}, \frac{M(\lambda_1, \lambda_2, \lambda_3, 1)[2]}{M(\lambda_1, \lambda_2, \lambda_3, 1)[4]}, \frac{M(\lambda_1, \lambda_2, \lambda_3, 1)[3]}{M(\lambda_1, \lambda_2, \lambda_3, 1)[4]} \right)$$

Field of Consciousness and choice of transformation

Field of Consciousness:

- Target of a projective transformation
- Subset of E_3 centered on 0
- 0 is the place where the conscious agent locates itself

How to go from 'real world' to Field of Consciousness?

- 'real' world position, orientation → projective transformation
- Multiple choices: give a set of axioms

Choice of transformation: Visualization

Canonical choice of transformation

Set of axioms (in the agent's frame of reference):

1. The subject is centered in 0 after projective transformation, i.e. in its perspective it is at the center of its frame.
2. The three axes x, y, z are preserved, i.e. the axes of the Euclidean frame associated with the agent (up-down, left-right, and back-front) must be preserved after projective transformation.
3. No points in ambient space appears, to the subject, to be truly at infinity; this constraint being 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.

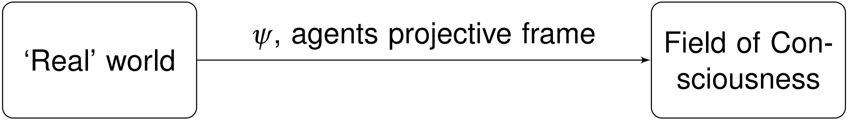
Canonical choice of transformation II: Results

1. Space of transformation parametrized by a positive real number c (depth of field)
2. The perceived size of objects, r_p, varies (asymptotically) as,

$$\frac{r_p}{r} \simeq \frac{1}{cz} \qquad \text{(Steven's Law)}$$

   where z is the coordinate of the direction along which the subject is aiming and r the length of the object in the 'real' world.

Choice of transformation: Visualization



Information integration and affective quantity

Perceived value of an entity: μ,

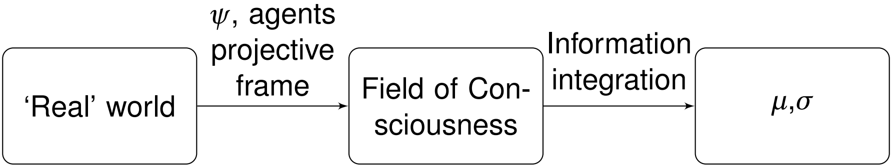
- weighted mean of preference and neutral preference, weighted by perceived volume v_p in the projective chart,

$$\mu = p\gamma\frac{v_p^{1/4}}{v_{tot}^{1/4}} + q_n(1-\gamma\frac{v_p^{1/4}}{v_{tot}^{1/4}}) \qquad \text{(Adjusted Steven's Law)}$$

with $\gamma \in [0,1]$ attentional focalisation (centered on body of the agent and its orientation), and v_{tot} the total volume of the bounded field of view

Information integration and epistemic quantity

- σ_e: uncertainty agent has towards an entity e relative to its field of view.



PCM what we get

Already have:

- Personal Perspective on the ‘real’ world
- Internal world model, workspace
- Attention
- Emotion

What is left to present:

- Imagination? : taking perpective with respect to other Euclidean frames
- Epistemic drive? : acting in order to reduce uncertainty
- Empathy?: simulating other agent’s perspective

Action

- Uncertainty seen as uncertainty on perceived value
- (μ, σ) parameterize a probability law on $[0, 1]$ centered on μ and of dispersion σ: $Q(\lambda|\mu, \sigma)$, $\lambda \in [0, 1]$.
- Achieve high perceived values, low uncertainty
- A desired state $P \in \mathbb{P}[0, 1]$ with high μ, low σ
- Minimise,

$$\mathrm{DKL}\left(Q\left(.|\mu_{\psi(m)}, \sigma_{\psi(m)}\right) \| P\right)$$

  where frame ψ changes with action m.

Multi-agents: ToM and perspective taking

How it generalizes to multi-agents:
Theory of Mind and (imaginary) perspective taking

Multi-agents: ToM and perspective taking I

⚠ We will now call the singled out conscious agent as subject

1. Subject has preferences for entities (objects, other agents)
2. Subject has priors on preferences of other agents
3. Subject is influenced by other agents for
   - preference update (inference phase)
   - action
4. Subject makes plan for how other agents will behave

Multi-agents: ToM and perspective taking II

- Theory of Mind (ToM): inferences about mental states of others (preferences) to understand and predict their behaviours, integrating spatial relationships and affective values
- Collection of entities E (objects and agents), collection of agents A
- The configuration of an entity, e, is a subset of $\mathbb{R}^3$ denoted as $X_e \subseteq \mathbb{R}^3$, collection of configuration denoted as X.

ToM-0

Subject of ToM-0:

- preferences $(q_e, e \in E)$
- no preferences update
- choice of action given by

$$C_0(m, X, q) = \sum_{\substack{e \in E \\ e \neq s}} \frac{1}{|E|-1} \mathrm{DKL}(Q(.|\mu_{\psi(m),q}(X_e), \sigma_{\psi(m),q}(X_e) \| P) \tag{ToM-0}$$

ToM-1

Subject of ToM-1:

1. Simulates agents as ToM-0, taking their perspective
2. preference matrix $(p_{ae} \in [0,1], a \in A, e \in E)$
   - preferences attributed to others
   - true preferences: $q_s = p_s$.
3. Influences of other agents
   - influence vector for preferences update $(J^p_{se}, e \in E)$
   - for actions $(J^m_{se}, e \in E)$
4. Planification of others actions, contribution to action after k steps:

$$C_1(m, p^k, J, X^k) = \sum_{a \in A} \sum_{\substack{e \in E \\ e \neq b}} \omega_{a,e} \mathrm{DKL}(Q(.|\mu_{a,\psi_a(m),p^k_{a.}}(X^k_{e,m}), \sigma_{a,\psi_a(m)}(X^k_{e,m}) \| P)$$

Experimental Results

Can the PCM algorithm robustly simulate adaptive and maladaptive canonical behaviors evaluated in child developmental and clinical psychology?

1. approach-avoidance and joint-attention behaviors
2. psychopathology
    - Autism Spectrum Disorders (ASD): deficits in social interactions, restricted interests, repetitive behaviors, deficits in joint attention
    - Social Anxiety Disorders (SAD): avoidance of social situations related to negative priors about self attributed to others while caring about others' judgments

For videos of experiments see
`http://www.gregoiresergeant-perthuis.com/`

Improvements

- Implementation of active inference for emotion assessment adapted from [SPF19]: Internship of Rida Lali
- Reformulation of FEP with perpective taking (group action): upcoming paper

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

References II

- David Rudrauf, Daniel Bennequin, and Kenneth Williford, *The Moon illusion explained by the projective consciousness model*, Journal of Theoretical Biology (2020).
- David Rudrauf, Grégoire Sergeant-Perthuis, Olivier Belli, Yvain Tisserand, and Giovanna Di Marzo Serugendo, *Modeling the subjective perspective of consciousness and its role in the control of behaviours*, arXiv:2012.12963, 2020.
- Ryan Smith, Thomas Parr, and Karl J. Friston, *Simulating Emotions: An Active Inference Model of Emotional State Inference and Emotion Concept Learning*, Frontiers in Psychology **10** (2019), no. December, 1–24.
- Susan Schneider and Max Velmans, *The Blackwell Companion to Consciousness*, Wiley Blackwell, 2017.

References III

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