

Representations: Useful, useless or harmful?

Giacomo Aldegheri and Sushrut Thorat





A representation is a measure

- Shows that a certain part of the brain contains information about a variable
- That information can *potentially* be read out by some downstream process

How is representational analysis useful?



- It can potentially constrain process models of the brain (e.g. visual hierarchy)
- It provides explanations at a relevant scale
 (e.g. in terms of the atoms of interaction with the world
 representations of objects)

Issues with representations



The brain is a highly complex, nonlinear, dynamical system.

It deals with an informationally, but also *structurally*, rich world with a rich repertoire of behaviours.

Information processing is highly contextual => Inferred representations are a function of the experiment performed.

The information a neuroscientist reads out from a piece of cortex might not be read out by the rest of the brain, making the inferred representation generally uninteresting. (*co-varying variables,* e.g. animacy in V1)

Gilbert and Li, Nat. Rev. Neur. 2013; Jonas and Kording, PLOS Comp. Bio. 2017

What are the alternatives?

nature neuroscience

FOCUS | PERSPECTIVE https://doi.org/10.1038/s41593-019-0520-2

A deep learning framework for neuroscience

Blake A. Richards^{1,2,3,4,42*}, Timothy P. Lillicrap^{5,6,42}, Philippe Beaudoin⁷, Yoshua Bengio^{1,4,8}, Rafal Bogacz⁹, Amelia Christensen¹⁰, Claudia Clopath¹¹, Rui Ponte Costa^{12,13}, Archy de Berker⁷, Surya Ganguli^{14,15}, Colleen J. Gillon^{16,17}, Danijar Hafner^{15,18,19}, Adam Kepecs²⁰, Nikolaus Kriegeskorte^{21,22}, Peter Latham²³, Grace W. Lindsay^{22,24}, Kenneth D. Miller^{22,24,25}, Richard Naud^{26,27}, Christopher C. Pack³, Panayiota Poirazi²⁸, Pieter Roelfsema²⁹, João Sacramento³⁰, Andrew Saxe³¹, Benjamin Scellier^{1,8}, Anna C. Schapiro³², Walter Senn¹³, Greg Wayne⁵, Daniel Yamins^{33,34,35}, Friedemann Zenke^{36,37}, Joel Zylberberg^{4,38,39}, Denis Therien^{5,42}

Study cognition in terms of objectives (tasks), architectures, and learning.

Representations are just an outcome of that. However, most representations are *not* interpretable => shortcoming of the current direction of the field? (e.g. modelling primate V4)

"Good" representations as a task?



If you have many (complex) tasks, having good representations can help in *transfer* and *few-shot* learning

Creating low-dimensional representations might be a useful *auxiliary* task (e.g. predicting how a scene would look might help in constructing useful representations for a motor control task)

Eslami et al., Science 2018; https://ruder.io/transfer-learning/

The usefulness of representations



Discussion points

- 1. Do we agree upon the definition of a representation?
- 2. Just how serious are the issues with representations that we mentioned?
- 3. What should we do? Parallel development of goal-oriented process models *and* characterizing neural representations?
 - 1. Or do the two frameworks have two different goals, and cannot interact?
 - 1. What is the example of a complete loop of interaction?
 - 2. What questions might lead to such a loop?

Supplementary info





Gilbert and Li, Nat. Rev. Neur. 2013



Objective functions



Richards et al., Nat. Neuro. 2019

Architectures

Observation 1



Eslami et al., Science 2018