

# Philosophy of AI (WBMV05003)

Janneke van Lith



Universiteit Utrecht

2019-2020 period 3

February 12, 2020

◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

## Two philosophical questions about representation



- ▶ **content**: What are representations about? What determines the content?
- ▶ **function**: How does a physical state actually fulfil the role of representing in a physical or computational process?

# Representational function



The challenge (William Ramsey): provide a description that tells us what it is for something to function as a representation in a physical system.

Do tree rings represent the age of the tree? Do gastric juices represent food? Answering yes to many of such questions leads to **pan-semanticism**: attributing representations to a system then loses its explanatory value.

Is indication sufficient for representation? And similarity?

## Representations: a definition

According to John Haugeland (1991), a system has internal representations if the following conditions are met:

- ▶ It must coordinate its behaviors with environmental features that are not always reliably present to the system;
- ▶ It copes with such cases by having something else (in place of a signal directly received from the environment) stand in and guide behavior in its stead;
- ▶ That “something else” is part of a more general representational scheme that allows the standing in to occur systematically and allows for a variety of related representational states.

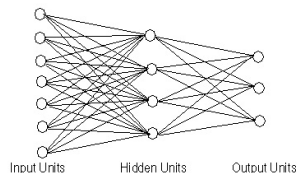
# Cognition without representation?

- ▶ Connectionism
- ▶ Dynamical Systems Theory (DST)
- ▶ DEEDS (dynamical, embodied, extended, distributed, situated cognition)

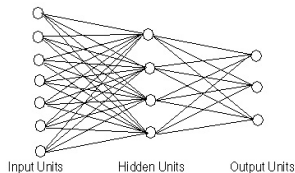
# Representation in connectionism

What kind of representation can be found in connectionist networks?

- ▶ activation patterns (see Ch. 7)
- ▶ connection strengths



- ▶ activation patterns (see Ch. 7)
- ▶ connection strengths



Connectionist representations are **distributed**: many items contribute to the representation of a single item. Moreover, the same items may contribute to very many different representations.



## Harnad on representation in connectionism

Stevan Harnad (1990), *The symbol grounding problem*

- ▶ argues for a hybrid model of the mind: symbolic plus connectionist
- ▶ sensory connections to the real world plus the connectionist part may deliver the grounding
- ▶ the symbolic part delivers the compositionality



## Ramsey on representation in connectionism (1)

William Ramsey, *Representation reconsidered*, CUP, 2007.

### activation patterns:

- ▶ activation patterns are responses to input; indicators
- ▶ 'receptor notion' of representation
- ▶ not truly representational; otherwise we end up with 'pansemanticism'; mere indicators

## Ramsey on representation in connectionism (2)

### connection strengths:

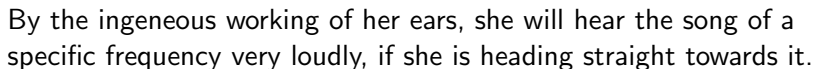
- ▶ 'tacit notion' of representation
- ▶ reversed causal order: structures are alleged to be representational due to their **effects**
- ▶ not truly representational; mere dispositional properties

*'There is no more reason for thinking the weights play a representational role than there is for thinking that the micro-structure of a fragile vase or the molecular constitution of a rubber band play representational roles.'*  
(Ramsey, *Representation reconsidered*, p. 178)

## DST example: the cricket



A female cricket is able to move towards a singing male cricket of her own kind, distinguishing his song even if other potential partners are simultaneously serenading her. How does she do that?



# Dynamical systems theory

Cognitive systems are dynamical systems (in an exact, mathematical sense). A cognitive process is the time evolution of a dynamical system.

Pro DST:

- ▶ Natural cognition happens in real time.
- ▶ Brain, body and environment can be explained in dynamical terms, so DST gives a proper account of the embeddedness of cognition.
- ▶ DST gives an account of self-organization, and thereby of the emergence and stability of cognition within the world.

# Situated cognition

DEEDS approach (Dynamical, Embodied, Extended, Distributed and Situated). Cognition is not limited to what is in the head, but also comprises interaction between a neural system, a body (hence: embodied) and an environment (hence: situated).

Internal representations are not always necessary for cognitive processes.

## Cognition without representation?

- ▶ DEEDS attributes a much smaller role to representations in their explanations of cognitive phenomena than orthodox views in cognitive science.
- ▶ But: there are *representation hungry* tasks, for which representations seem to be necessary (e.g. offline cognitive tasks).
- ▶ Interesting question: can DEEDS account for such higher-level cognitive capacities?



## Situated cognition

Entrance test, Presentations, Class discussions

## Tips on essay writing

# Entrance test



- ▶ Monday, February 17, 11.00-12.30/13.00, Educatorium Beta
- ▶ Handbook Ch. 1-6, lectures up to this point
- ▶ Example questions on Blackboard
- ▶ Have your Solis-id ready!

# Presentations

- ▶ by 2 students
- ▶ 15 minutes including discussion
- ▶ present and comment on a paper related to that week's theme
- ▶ suggestions on Blackboard; check with seminar teacher
- ▶ contains:
  - ▶ outline of presentation itself
  - ▶ main standpoints and arguments from the paper
  - ▶ your own critical comments on the paper

## Class discussions

- ▶ post a thesis and supporting argumentation on Blackboard 24 hours before seminar
- ▶ 100-300 words
- ▶ thesis related to the required reading
- ▶ class discussions mainly based on Blackboard posts, also on presentations

# Procedure

- ▶ Throughout the course: think of an interesting topic / question
- ▶ Tuesday, March 24: research question and outline ready
- ▶ March 25-29: online peer feedback
- ▶ March 30/31: feedback session
- ▶ Sunday, April 5: first version ready
- ▶ April 5-8: online peer feedback
- ▶ Friday, April 10: final version ready

# Formal requirements

- ▶ 3000-4000 words (excl. bibliography)
- ▶ in English
- ▶ philosophical paper, that argues for a certain thesis related to one of the five themes of the course
- ▶ includes a critical discussion of relevant literature
- ▶ makes use of at least 3 good quality papers / book chapters
- ▶ proper reference to all sources used
- ▶ the assessment form is on Blackboard

## Finding sources

- ▶ use the suggestions from Blackboard
- ▶ use the suggestions from your fellow students (presentations; peer feedback)
- ▶ search on [philpapers.org](http://philpapers.org)
- ▶ search (cleverly) on [scholar.google.com](http://scholar.google.com)
- ▶ ask the lecturers

## Using sources

- ▶ don't just tell your own opinion, but build your argument on existing discussions and ideas
- ▶ always refer to the sources you have used
- ▶ refer in some consistent way (APA, Chicago, ...)
- ▶ keep quotations to a minimum
- ▶ using other people's text without proper reference is considered plagiarism



## What a philosophical paper is not (1)

- ▶ a literary confession,  
in which you express your feelings, opinions, intuitions:

*"I believe that people have a free will and that our actions are not predetermined. I think it's simply crazy to deny this. In fact, most people agree with me on this. I also believe ..."*

- ▶ a list of other people's opinions and positions,  
also not of the "big philosophers":

*"Descartes claimed that there are two kinds of substances, that animals have no emotions, that he proved the existence of God ..."*

## What a philosophical paper is not (2)

- ▶ an opportunity to display your impressive knowledge of the literature:

*“Davidson said that we can't justify our beliefs from other beliefs. Sellars claims something similar. Also certain passages in Hegel seem to point in this direction, and already Aristotle knew that ...”*

- ▶ a collection of quotes and passages from other people's work:

*“Descartes discussed in the first meditation the reasons for why we can doubt the existence of all things. Later he says: “Everything that I hold to be true, I've encountered via my senses. But now I found out that my senses sometimes deceive me, and it's a matter of prudence never to trust someone completely, who even deceives you only once”. He continues saying ...”*

## What is a good philosophy paper? (1)

- ▶ a reasoned and structured defense of some claim (so not *just* a formulation of your private opinion)
- ▶ keep it small: don't argue that computers can (or cannot) think, but choose a more modest claim
- ▶ broaden the scope both in the introduction and the conclusion, and zoom in to defending your specific narrow claim in the middle part
- ▶ be original, i.e. show independent thinking and respond critically to existing literature

## What is a good philosophy paper? (2)

Examples of the general aim of a philosophical paper:

- ▶ defending a claim or an argument
- ▶ criticizing a claim or an argument
- ▶ evaluating the strength and weaknesses of opposing positions
- ▶ discussing the consequences of a thesis, i.e. making the commitments of certain views explicit
- ▶ giving a new formulation or interpretation of an argument or thesis to avoid common criticisms
- ▶ developing or discussing counterexamples to a commonly accepted thesis
- ▶ developing or discussing examples to explain or reinforce commonly accepted claims

## Do's and don'ts

- ▶ Rule of thumb: never just say 'I think / believe that p' without explaining why

*If there is no space for explaining a presupposition, be explicit about that.*

- ▶ Make sure you yourself understand what you write
- ▶ Don't make excessive use of jargon, or of colloquialisms
- ▶ Avoid category mistakes

*Example: 'One of the presuppositions of functionalism is the concept of causality'*

- ▶ Anticipate counterarguments
- ▶ Avoid platitudes

## Structuring the paper

- ▶ **Title**: choose a catchy and representative title
- ▶ **Introduction**: formulate the question you are going to address, specify context and motivation, and give an overview of the rest of the paper
- ▶ **Body**: give a critical exposition of relevant literature, and your arguments leading to your answer to the research question
- ▶ **Conclusion**: recap of the arguments given; explicit answer to research question
- ▶ **List of sources**: consistent, precise, and complete

## Writing an outline

- ▶ write a draft version of your introduction: research question, context, motivation
- ▶ specify the subsections
- ▶ give some indication of what will be written in each subsection, preferably by stating a question, your answer, and some indication of the arguments underlying your answer
- ▶ include a (preliminary) list of sources

## Further tips

- ▶ How to read and write philosophical texts, by Michael Pohl and Johannes Korbmacher
- ▶ Guidelines on Writing a Philosophical Essay, by Menno Lievers
- ▶ Guidelines on Writing a Philosophy Paper, by Jim Pryor
- ▶ A sample philosophy paper, by Angela Mendelovici

(links are on Blackboard)