




Marcin Miłkowski, IFiS PAN

ROLE OF REPRESENTATION IN COMPUTATIONAL EXPLANATIONS





Presentation Plan

- Representations in Explanations: A primer
 - Classical accounts and how they fail
 - Four case studies without a funeral
- 



Representation

- The aim of the talk: give an account of the role of representation in computational explanations in cognitive science.
- I will not:
 - Defend a full-blown philosophical theory of representation
 - Take programmatic manifestos at face value



Representation in Explanation

- Representation can be both used as *explanans* and as *explanandum* (Ramsey)
- If it is used as *explanans*, it has a specific theoretical role. It should shed light on some other phenomenon. You can take it as fundamental.
- As *explanandum*, it is the focus of explanation (usually with something else than representation). The notion is not treated as fundamental.

Representation

- Start with a simple Dennettian point.
- Two levels of representation:
 - Personal level
 - Subpersonal level
- I will use “agent” instead of “person”. Nothing from Dennett is lost this way.

Agent-level Representation


- On the agent-level, we ascribe representations to the whole agent, based on rationality considerations, etc.
- This is the level at which we talk of beliefs, desires, etc.
- It is the focus of Dennettian intentional stance, Newell's knowledge level, folk psychology, classical economical theories...


Agent-level Representation

- Note: Agent-level representation is usually taken as the (part of the) explanandum phenomenon.
- It is used as *explanans* in folk psychology and in the humanities. It is not a part of the explanatory framework of computational cognitive science.




Sub-agent representation

- Sub-agent representation is usually considered in relation to agent-level representation.
- 



Agent vs. Sub-agent representation

- Three possible options:
 - Anti-realism: There is no agent- or sub-agent level representation, so nothing to relate to (instrumentalism, eliminativism)
 - Functional localization claim: the agent-level representations are localizable in functional parts of the agent
 - Lack of functional localization (holism, emergence, implicit representation)
- 

Anti-realism

- Anti-realism is not necessarily a bad option for behaviorism and its modernized versions
- But it means that there is no special role for representations to play, so I leave it aside
- Note: if there is real explanatory gain from representation, some anti-realist claims will be undermined (and I argue to this effect)

Functional localization (or lack thereof)


- There would be straightforward correlates of your beliefs (or propositional attitudes, or emotions) in the functional parts of the agents.
 - Note: it does not have to be at the neurological level.
- Mentalese hypothesis seems to involve some claim like this.

Representation as *Explanans*

- In using representation as *explanans*, you take it as contributing to the *explanandum* phenomenon
- In the explanatory framework that I am using (mechanistic explanation), it means causal contribution
- But if it is *just* causal contribution, it ain't representation (Ramsey on Dretske-style detectors)



Representation as *ExpLanandum*

- If you want to explain something as a representation, you can treat representation as an epiphenomenon (explaining away).
 - But if it is a part of another phenomenon, again you need to make it causally relevant as representation.
- 

Classical Account

- The classical account is motivated by the hope that computation will explain representation.
- Representations are taken to be symbols that correspond to their referents, and computation understood as manipulation of symbols.
- Computation is at the same time physical, so we naturalize representation this way.
(Pylyshyn, Newell) Hooray!

Classical Account Won't Work

- Computation will not explain representation. At least not reference (symbol grounding problem means that you cannot take it for granted).
- "Symbol" is an ambiguous word. The whole argument rests on equivocation.
- Computation cannot be used to naturalize representation. There is no content there.

Symbol, schmymbol

- A symbol in computation theory is just a token over the alphabet, so this is just a syntactic entity.
- It has no content.
- Such symbols can be *ascribed* content by users but if you explain *mental* representation, you posit undischarged homunculi that ascribe content to your Mentalese tokens.

Symbol, schmybol

- At the same time, it can have certain syntactic features that suggest compositionality or systematicity.
- So you can try naturalizing ascription without the internal observer.
- Natural meaning relations are the usual candidate.

Symbol, schmymbol

- But note, symbols that acquire natural meanings are not the same symbols as in computation theory.
- They are schmymbols.
- Symbols remain meaningless if you have no theoretically-grounded justification of their identification with (some) schmymbols.

Symbol, schmymbol

- The moment you looked at schmymbols meant you don't believe that computation explains representation.
- So all natural-meaning-based theories actually presupposed that Pylyshyn and Newell cannot be right.

Non-classical Accounts


- Non-classical accounts all concede that computation will not explain representation.
- Most rely on two relations:
 - Covariation
 - Resemblance
- ... sometimes with some teleology tacked on.
- But they won't work either (cf. Bickhard on encodiginism and Ramsey's challenge).

Non-classical Accounts

- Covariation and resemblance accounts are:
 - Circular: you do not know what to look at as relevant in the world to see if it covaries or is similar
 - Too broad: all regularities and laws, including laws of arithmetic, would be representations
 - Too narrow: empty representations do not have any referents, so there is nothing to relate them to.
 - Usually cannot account for misrepresentation.



So what do we need?

- We need an account that:
 - Has a *functional* role for representation (and content!) *in* the cognitive system
 - Makes misrepresentation possible
 - Does not reduce representation to an encoding relation
 - Makes misrepresentation detectable by the agent
- 

Mechanistic Explanation

- Mechanistic Explanation Thesis:
Cognitive Science explains cognition by positing *mechanisms* that display cognitive capacities

Stuart Glennan



Bill Bechtel



Carl Craver




Representational Mechanisms

- My project is to sketch a meta-theory of representational mechanisms
- General specification of the representational mechanism:
 - Its capacity to represent
 - Explained in terms of its organization
- My problem right now is to build the proper specification that matches my desiderata



Four cases

- I will show that some classical and modern computational explanations crucially involve representation in various roles.
 - I assume the mechanistic framework of explanation.
- 

Cryptarithmic

- Suppose you're given this task: Find natural numbers that are hidden behind letters here:

SEND
+ MORE
= MONEY

Cryptarithmic


- Or another. Start from assigning $D = 5$:

DONALD
+ GERALD
= ROBERT

- You probably notice that $T = 0$, given that we have two D s above T




Cryptarithmic

- Newell & Simon (1972) explained how people solve such problems using an impressive computer simulation
 - On one hand, they analyzed how the task could be solved and programmed a computer model appropriately
 - On the other hand, they gather verbal protocols (and eye movement data) of people solving the task
- 



Cryptarithmetics

- The general methodology is as follows:
 - Analyze the task, including possible representations of the problem being solved and the problem *search space*
 - Gather behavioral data
 - Build production system rules (rules of rewriting symbols) that could be used to search for the solution in the space
 - Test it on data
- 

Cryptarithmetics: Summary

- This is an example explanation in the research tradition of *symbolic* information-processing
- Cognition is explained via rule-based symbol manipulation by showing a model that matches human *performance*
- Representations are not *explananda*. They are used to explain the task and remain unexplained themselves.
- The task is representation-hungry.

Past-tense learning

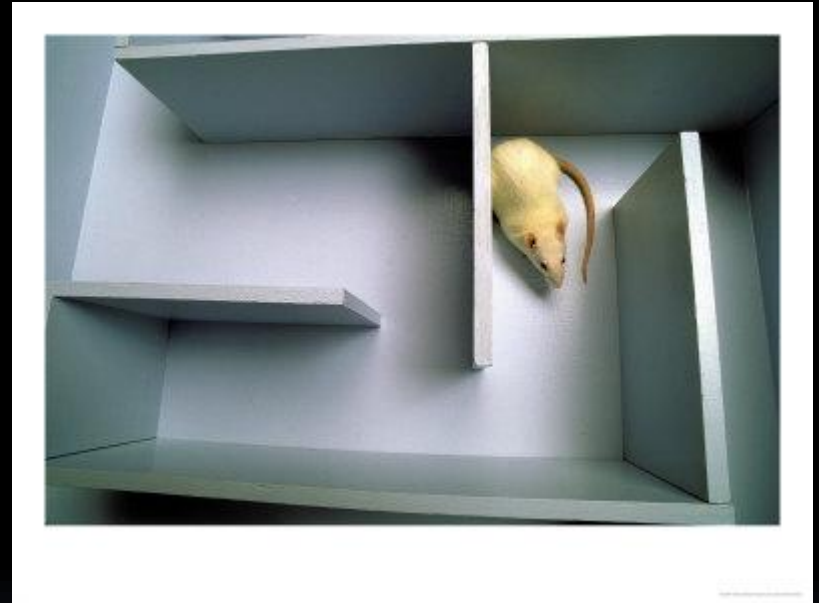
- Rumelhart & McClelland (1986) famously modeled the acquisition of the past-tense forms of English verbs
- The model reflected the known behavioral data on stages of development in children but without any explicit rule-based processing
- The match with *performance* was high

Past-tense learning

- Representations were used as part of *explanantia*: encoded as Wickelfeatures (structurally rich phonological patterns)
- Isolated verbs were used in learning
- There is no evidence that such patterns are neurologically plausible
- The explanation is incomplete: it does not say what role *exactly* these representations have in normal speech production.

Rat Navigation

- Conklin & Eliasmith (2005)
- Rats are known to return directly to the starting location after exploring the environment in a somewhat random fashion.






Rat Navigation

- The only cue available to the rat is his own movements, and the widely accepted hypothesis is that the rat represents the environment in a mental (or neural) map.
- Using Neural Engineering Framework, Conklin & Eliasmith built a model that uses neural *representation* (2D bell-shaped function as a bump of neural activity that is the rat's estimated location)



Rat Navigation

- In NEF, the neural representation is included in a dynamical mechanism and is considered to be a part of a control mechanism (modeling using control theory).
 - At the same time, there is a sketch of a high-level mechanism responsible for navigation to the starting location based only on self-motion velocity commands from the vestibular system.
- 

Neural Engineering Framework

- *Representation*. Neural representations are defined by a combination of non-linear encoding and optimal linear decoding.
- *Transformation*. Transformations of neural representations are functions of the variables that are represented by a population.
- *Dynamics*. Neural dynamics are characterized by considering neural representations as control theoretic state variables.



Rat Navigation

- Representation is *explained* here in terms of the control theory, neural coding and decoding (whereas decoders are only theoretical entities). Encodingism!
- But it has a role in real animal behavior as a sensomotor representation of its location: it guides behavior.
- If NEF considered the role of animal-detectable error, however, the encodingism would be gone.

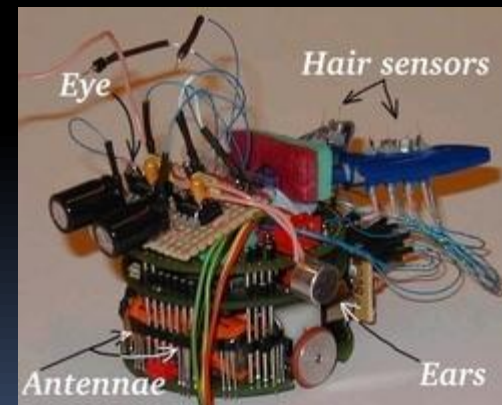
Cricket phonotaxis

- Crickets display phonotaxis: ability to walk towards the location of the male by the female based on the male's calling song



Crickets and biorobotics

- One of the ways to explain how crickets are able to do it, is to build a synthetic model (biorobotics in explaining cognitive animal behavior)



Cricket phonotaxis

- Barbara Webb and her collaborators build synthetic models of crickets to explain their capacities
- Existing behavioral data is rarely enough to build a full robot: new data is needed and you know it only when building a real robot





Cricket phonotaxis


- Behavioral data: cricket ears are quite special as they filter and delay some sounds and that makes the task of locating the source easier
- The carrier frequency of the male cricket's song is around 4-5 kHz, and it is used by females to detect conspecifics
- Females select the sound that has the right temporal pattern

Cricket phonotaxis: Hypothesis

- The significant cue for the filtering could be the onset of the sound.
- The hypothesis was tested using a simple neural circuit: the left auditory neuron excited the left motor neuron and inhibited the right motor neuron, and vice versa. This is the only representation here, used as *explanans* again.
- As a control mechanism, the circuit was able to reproduce a number of behavioral experiments in crickets.



Summary

- Four cases:
 - Three with representation as *explanans*
 - One with representation as *explanandum*
- 




Summary

- In classical cognitivism, the content was ascribed by the researcher.
- In connectionism, there was no reference and probably no content (only transformation).
- In NEF, the content was the location (but could be changed).
- In biorobotics, the content was the strength of signal.




Summary

- Could you reinterpret the results?
 - You don't need to buy Physical Symbol System hypothesis to analyze Newell & Simon.
 - Rumelhart & McClelland had no role for real representation, only for some phonological processing.
 - NEF detections could be replaced with interactive representation or something close to it.
 - For crickets, you can even be anti-representational.
- 



Summary

- To quote Eliasmith:

“The explanatory power of representations comes from the fact that they can be manipulated internally without manipulating the actual, external, represented object.”
 - So, representations help solve problems. And problem solving is explained in cognitive science.
- 



Summary

- It's a book-length project (*Explaining the Computational Mind*), so there are much more details...
- 