EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Problems with the LoTH

Or: Fodor gonna get ya

- *The Language of Thought: computational cognitive science approaches to category learning*
- Who: Fausto Carcassi
- When: Sommer semester 2022

From two weeks ago

- A good point of discussion from two weeks ago was the functionalist concept of ‘belief’.
- The problem can be put as follows:
  - The functionalist account is meant as an *analysis* of belief*
  - E.g., ‘believing* p’ behaves in a certain way wrt other beliefs
  - However, it seems like for any set of rules we can give, real ascriptions of beliefs that we do in everyday life fail to follow those rules.
  - E.g., ‘believing* p and q’ implies ‘believing* p’
- One possible response:
  - Functionalism is meant as a scientific analysis. Compare to the analysis of ‘water’ as H2O. Substances that were thought to be water with the pre-chemical definition might have turned out to not be water. But H2O is not a *definition* of water, it’s an analysis (it’s empirical & could turn out to be false!)
  - So whatever it is we are doing with the functionalist analysis of belief*, it’s the same as whatever we are doing with the chemical analysis of water.
- This is made more plausible by the fact that mental state types are plausibly natural kinds.

From last week

- One point of discussion last week was the following:
- In which sense are the symbol manipulation rules distinct from their meaning?
- From a computational perspective conjunction can be *defined* with the following three rules:

$$\frac{A\ true \qquad B\ true}{A \wedge B\ true}\ \wedge\mathrm{I} \qquad \frac{A \wedge B\ true}{A\ true}\ \wedge\mathrm{E_L} \qquad \frac{A \wedge B\ true}{B\ true}\ \wedge\mathrm{E_R}$$

- The point is that these rules make no reference to semantic concepts such as ‘truth’ or ‘reference’. They are just rules for *manipulating symbols*.
- If all it took to understand conjunction was the ability to manipulate these rules, a water system implementing the rules would understand conjunction.

Review of the situation so far

- In the previous two weeks, we have seen a *picture* of the mind from the point of view of cognitive science.
- Namely, the mind as a system that runs computations over representations that are structured like a language.
- Moreover, we have seen a variety of arguments in support of this picture.
  - Can you recall what they are?
  - Fodor's main type of argument is the 'Only Game in Town' argument
- The picture we have seen has many components, and each of them can be attacked.
- This week let's look at some of the possible ways that the picture can be attacked.

The formal-syntactic view of computation

- Fodor's *purely formal* view of computation in the LoT has been criticized
  - Do you see what could be the problem with it?
- Basically, the problem is that intuitively the *content* of our thoughts (which is a semantic property) is part of what explains the actions.
- Fodor's answer is too complicated to discuss it here.
- What do you make of this?

Computationalism has more trouble with another problem about mental causation. Generally, when an event c causes an event e, we can ask what it is *about* c in virtue of which it causes e.[1] Some of c's properties are causally relevant. For example, when the bowling ball causes the pins to fall, some properties of the ball play a role, and others do not: the weight of the ball is causally relevant, but not the color or the price of the ball. Intuition endorses the claim that the content of Zelda's thought is causally relevant. Nothing could be more obvious: her plant-watering thought causes plant-watering precisely because it is a plant-watering thought. Generally, intuition agrees that

(REL) Content is causally relevant: thoughts cause behaviors (as well as subsequent thoughts) in virtue of their content.

But can computationalists accept (REL)?

In my view, the answer is No. Computationalism has the counterintuitive implication that our thoughts do not have their effects because of their contents: that thought content is "epiphenomenal."

Kazez (1993), Computationalism and the causal role of cor

Searle’s Chinese room

- NOTE: Searle is a really awful guy (google it), nonetheless I’d be doing you a disservice if I didn’t mention this.
- Searle (1980), *Minds, Brains, and Programs,* The Behavioral and Brain Sciences.
- Here’s the thought experiment:
  - Imagine yourself in a closed room with two holes, call them I and O.
  - You get pieces of papers with symbols though I
  - Following a big book of instructions, you write stuff on more paper
  - And send it through O
- In principle, the room could this way instantiate any intelligent behaviour, e.g. understanding of Chinese (assuming Chinese symbols), even if you don’t understand Chinese.
- The idea is that a purely computational view (symbols manipulation) cannot make sense of things like *understanding*.

Searle’s Chinese room

- The ‘robot reply’ is Fodor’s answer to the Chinese room thought experiment.
- The reply is something like: part of what constitutes our meaning is our interaction with the world. Therefore, if a Chinese-room-like system was implemented in a robot that interacts with the world, it would acquire intentionality.
- Fodor changed his mind a couple times on the detail of this!
- Searle isn’t very convinced: ‘Of all the zillions of criticisms of the Chinese Room argument, Fodor’s is perhaps the most desperate.’ (in Rosenthal 1991, p. 525)
- What do you make of this counterargument?

Chinese Room & Cole (1984)

- Imagine each of your neuron is itself conscious and aware of its chemical processes (transmitting chemicals, etc).
- Your conscious neurons might find it implausible that out of their combines activity emerge things like experience, pain, etc.
- But experience tells us that such things do in fact emerge.
- What do you make of this?

‘Everything is a computer’

- Section 2.2 in *The Mind as the Software of the Brain*
- Primary source: Putnam, Hilary, 1988. *Representation and Reality* (pp.95-96 and 121-125).
- This is a long attack on functionalism by Putnam, one of the people who invented it in the first place.
- The argument is too complicated to analyze it here (see book’s Appendix)
- The conclusion is: any physical system can be *interpreted* as instantiating any computation.
- For instance, with the right interpretation we could see a wall as instantiating Doom or Crysis (Can it run Crysis? *Yes*)
- And since there is no fact of the matter as to what interpretation is ‘correct’, computation cannot identify mental states.

Regress objections I

- As we saw last week, Fodor's account of learning the meaning of words is by 'hypothesis formation and testing'.
- This raises the question: How do children learn the LoT?
- Fodor's answer is obviously: they don't. It's *innate*.
- In other words, hypothesis formation and testing does not apply to *concepts*:
- 'Now, according to HF, the process by which one learns C must include the inductive evaluation of some such hypothesis as "The C things are the ones that are green or triangular". But the inductive evaluation of that hypothesis itself requires (inter alia) bringing the property green or triangular before the mind as such […] Quite generally, you can't represent anything as such and such unless you already have the concept such and such. All that being so, it follows, on pain of circularity, that "concept learning" as HF understands it can't be a way of acquiring concept C […] Conclusion: If concept learning is as HF understands it, there can be no such thing. This conclusion is entirely general; it doesn't matter whether the target concept is primitive (like green) or complex (like green or triangular).' (2008, p.39)

Regress objections I

- But consider the fact that we represent concepts like ‘iPhone’. According to Fodor, these have to be innate too, since they’re expressed in the LoT.
  - This is the ‘carburetor’ objection
- By ‘innate’ here Fodor simply means ‘not learned’
- Still, clearly concepts can be acquired.
- We’re in a bit of a pickle! Various objections and counterobjections possible.
- In the second half of the course we’ll see a way in which we can get out of this pickle.
- What do you make of the regress objection?

Regress objection II

- How do we *understand* a sentence/word in the LoT?
- With natural language words, understanding means representing in an appropriate LoT expression.
- But how do we represent the meaning of LoT expressions?
- Answer: We don’t. Rather, we think *in* the LoT.
- By comparison: A computer interprets the high-level input it receives into some machine language, but it doesn’t need an interpreter for the *machine language* itself.

Dennett (1975)’s first objection

- Section 3.1 in *The Mind as the Software of the Brain*
- We have infinitely many beliefs!
- But how can we be in a believes* relation to infinitely many LoT expressions?
- For instance, ‘2 is smaller than 3’, ‘2 is smaller than 4’, etc.
- This would seem to require infinite memory.
- Block’s answer: draw a distinction between *ordinary* and *scientific* notions of belief (recall discussion earlier today!)
- In ordinary sense, we have infinitely many beliefs.
- In scientific sense, we only have finitely many beliefs.
- Namely, those beliefs that are *causally active*: they cause behaviour or other mental states, or they are caused by perception or other mental states.

Dennett (1975)’s second objection

- Dennett notes that in some cases we ascribe beliefs that are not explicitly represented.
- For instance, someone playing guitar (who doesn’t know music theory) might in a way believe that at some point in a song they have to switch to an A-
  - Despite lacking any such explicit representation!
  - In a way this is an emerging fact about the way they play.
- The argument is that our pre-theoretical notion of belief does not align with the notion of belief that the LoT proposes.
- However, again, we point out that the LoT is making sense of some scientific, *cognitive* notion of belief rather than a pre-theoretic one.

Dennett (1975)’s third objection

- Finally, a third objection is contained in the following thought experiment:
- A surgeon changes your brain by inserting the representation that you have a pet turtle in a belief* functional role.
- When you wake up you are asked whether you have pets and you say you indeed have a pet turtle.
- When you are asked how old it is and what its name is, you can’t answer, etc.
- So it seems like you don’t really believe it, in contradiction with the LoT account.
- LoTH answer: the functional role for belief* includes integration with other beliefs, other mental states, and behavior. And isolated belief isn’t a belief.

Connectionism

- The last point we should discuss is the relation between LoTH and connectionism.
- Connectionism is the project of explaining or studying aspects of cognition with artificial neural networks (ANNs).
- The crucial point here is that ANNs do not perform symbolic computation, and therefore they seem prima facie incompatible with the LoT.
- The big discussion of the connections is in Fodor and Pylyshyn (1988)
- However, as always in philosophy the situation is more complicated.

Connectionism

- We can distinguish two types of connectionism:
- *Eliminative* connectionism
  - An *alternative* to symbolic computational account
  - Exists at Marr's *computational* level
  - Therefore, we should require it to explain the same phenomena as LoT
- *Implementationist* connectionism
  - An account of how cognition is *implemented* in the brain
  - Exists at Marr's *implementation* level
  - Therefore, it is not incompatible with the LoTH

Connectionism (Aydede 2015)

1. Representational mental states and processes exist. An explanatorily adequate account of cognition should acknowledge these states and processes.
2. The representational states and processes that figure in high-level cognition have certain fundamental properties: thought is *productive* and *systematic*; inferential thinking is *systematic*. The states and processes have these properties as a matter of *nomic necessity*: it is a psychological law that they have the properties.
3. A theory of mental computation is explanatorily adequate only if it explains the nomic necessity of systematicity and productivity.
4. The only way to explain the nomic necessity of systematicity and productivity is to postulate that high-level cognition instantiates computation over mental symbols with a compositional semantics. Specifically, we must accept RTT+COMP.
5. Either a connectionist theory endorses RTT+COMP or it does not.
6. If it does, then it is a version of implementationist connectionism.
7. If it does not, then it is a version of eliminative connectionism. As per (iv), it does not explain productivity and systematicity. As per (iii), it is not explanatorily adequate.
8. **Conclusion**: Eliminative connectionist theories are not explanatorily adequate.

Summary

- This week we have considered various objections to the LoT:
  - Syntactic view of computation
  - Searle’s Chinese room
  - ‘Everything is a computer’ objection
  - ‘Carburetor’ objection
  - ‘Understanding the LoT’ objection
  - Dennett’s objections
  - Connectionism
- We didn’t go into much detail on any of them, but hopefully now you have an idea of some ‘weak points’ of the LoTH.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

What’s next

- On Wednesday, we’ll finish the intro to python.
- Next week we’ll leave philosophy behind!
- And get onto the technical stuff.
- We’ll see the concept of formal grammar, formal language, set, function, etc.
- And also introduce the beautiful system of *lambda calculus*.