Information and injective mappings

A very early sketch

Overview

- What type of information do structural representations store?
- Injective mappings
  - Did Hubel and Wiesel neuron survive the Mante aftershock
- Vehicle realism
- Temporal stability
- Algorithmic homuncularism
- Similarity-based information (CTSI)

Temporal and spatial stability of vehicles

- How to single out tokens
- In a way that we find „meaning blocks” (localizable)
- Functionally-relevant (task-function sensitive)
- Ought to be formatted correctly to be usable (Yousif et al. 2019)
- Contentful – carry information about X in a stable manner
  - **Does not variate across time (somewhat, at least) – temporal dimension**
  - Rigid connection betwen the vehicle and target
- -> Vehicle realism

Offline/online: concepts, representations, information?

- Working memory is, however, crucial for maintaining information in consciousness in the **absence of input**, beyond a very short time window (Baddeley et al. 2021).
  - „Distance” (Cantwell-Smith 1999)
- But what is the relevant timescale for the „stabilization” (tokening) of vehicles. Absence is quite a relative term. [see: Aoi 2020]
- What is posited in AH is there are moments when a certain content-bearing vehicle is tokened, but the next one is still not. In other words, there is clear temporal sequence and separation between the tokens.
- Important because sometimes opponents of S-reps claim that their purported explanatory work is in fact done by offline/online distinction (Artiga)

Injective mappings

- Vehicles map into targets exlusively. No other vehicle reports to this adress.
  - Pace: multistable, reusable, shared vehicles
- Corollary: *causal infomorphism*: causal transitions between vehicles are to be well-defined, discrete in order to mirror steps posited within an algorithm.
- But what's really at stake here is the **informativeness** of these mappings. Assumption: vehicles latch onto meanings unequivocally because that's what they carry information about
- Hence, uncovering these mappings is the way to decode their meanings and follow the explanatory strategy of *algorithmic homuncularism (AH)*
- A key aim of the approach is to localize **exploitable information** to distinct brain regions in order to **illuminate the causal structure of thought**
  - ***Exploitable information: information-in/for (internal),***
  - ***as opposed for information-about (external)***

Population doctrine (Hopfieldianism; Krakauer &Barack)

- Anderson (2014) and Burnston (2021) argue that current practice in systems neuroscience renders AH no longer valid
- Burnston (2021) locates a particular difficulty for AH in the fact that dimensions of variation are often drawn from the activity of the entire system targeted by an analysis, rather than from the activities of subpopulations. (Kinsey 2023)
- Representations **are realized as trajectories in the state space of a whole population, rather than as isolated to distinct parts of that population. Hence, injective mapping is false.** (Burnston 2021: 1620) What is important about this, for current purposes, is that any measure of representation of a particular variable is taken **to be constituted by the whole population, rather than by distinct parts within that population.** If so, then the system cannot be divided in the way AH recommends. (Burnston 2021: 1624-5)

Followup

- Representations aren't localizable components of a functioning cognitive system, but are instead constitutive patterns or modes of activation of the system, and the graded nature of pattern evolution can prevent us from ascribing distinct contents to distinct stages of activity
- According to Hopfieldianists, outcomes of neural data processing techniques, esp. simplifying ones, such as PCA or other dimensionality reductions suggest that we do not extract „real" informational currency – segmented, localizable and unequivocal as AH contend.
- *Contingent content: (*e.g. Willett 2020 handknob study) – we can extract information that is **not** exploited by the agent and still infer very reliably on the basis of *contingent (epiphenomenal) content (*because coding is „entangled", „emergent" or „compositional". Regardless of the term – importantly: not 1:1*)*
- Mante et al. 2013 and Aoi et al. 2020 studies mobilized as weapons here to argue against homuncularism.
- Bowers et al. 2022 meta-analysis also very pertinent here

- In this view, the *entire population* or state space is vehicular.
- This does not preclude representationalism. Representations are understood as trajectories through the relevant space. Different patterns of stimulations are mapped to relevant trajectories, but these are (pace AH) not decomposable.
- The question is: where are part-whole relations for Burnston. If there are no parts, how there can be structural representation (in which relations between the parts are semantically significant and homomorphic to part-whole relations of targets)

- To perform a proper mapping, we need to understand which types of representations are matched with which types of information
  - Specifically interested in the status of structural representations here
  - And how they are trafficked in similarity-based relations
  - CTSI (Miłkowski 2020) for RSA (Kriegeskorte et al. 2008; Roskies 2018)
- Information is *stored* in concepts. Is this „unmediated explanatory" information (UE; Shea 2018)?
- Distinction between UE-information and *UE structural correspondence*
- Model-free vs. Model-based: only latter uses isomorphisms?
- Non-local inferences: structural and not sequential?
  - But in both chains of representation occur according to the syntax of vehicles
  - Different information models?
  - Case: are similarity-judgments rule-based or associative? Are they (genuinely) informative?

Aoi et al. 2020 study

- If we look closer at temporal dynamics we discover clear sub-regimes (epochs) that might map linearly into distinct contents. These changes might imply differences in coding, different signals intermingled (e.g. decision+perception) or changes in selectivity
- What's most important, though, is that it gives strong evidence to think that the signal is not „contaminated" or „confounded" to the point of undecomposability, as Burnston would want. Exactly contrary to his point here:
- What the temporal division requires is that there be a time at which one content is tokened, but the second content not tokened, and then a specific process that brings about that second content. But in the Mante et al. case, the task axes are constant features of the population response—every location in the PC space describes some place along the task axes. So, there is no temporal stage at which one content is tokened and the others aren't. Nor are there clear phases of transition between some represented contents and others.

Facchin's (2024) typology of S-reps

1. Simulations
2. Maps
3. Spaces
4. Dynamics

He trivializes the distinction between structural and non-structural (even indicators are structural in his view). Also the Burnston/Shea debate would lose its edge on this reading.

But more importantly it misses the mark. What's interesting is not the ontology of these representations but whether they differ in terms of mappings: are there different assumptions about what instantiates the vehicle, target, etc. In other words: how the informational model is built. And do we uncover/mirror the „real" pipeline when we do data analysis.

Helpful criteria?

- Perhaps these criteria from the Concepts (ch. 2) book can help build a hierarchy of informational structures (setting up a ladder of increasingly stringent set of criteria):
- Token-ablity, organization, distributional properties (transitivity?), completeness?

- Hubel and Wiesel neuron: selectively responding to a very particular stimulus
- Informative about this stimulus: causal binding (but the relations is prima face gradable: on/off; magnitude or structure)
- Detector cells are about what they reliably correlate with (correlational information)
- These can be spurious of course, so we solidify our explanations by functional and systematic observations of a given detector cells (and intervene). But still the binding is arbitrary in an important sense: it's like an indicator.
- They were traditionally thought to be injective (mapped onto a single content, highly specialized/selective detectors). But Mante et al. Seemed to stir the pot. Story has it: vehicles are multistable! But are they really?
- And what about relations between detectors? Are they mutually informative?

- What role do similarity relations play in the „vertical” relations between the vehicles?
- Facchin e.g. thinks informational and structural relations are of a different kind.
- But CTSI, on other hand, claims that similarity is the basis of making reliable (surrogative) inferences. Similarity is informative.
- Shannonian coding is arbitrary, according to CTSI coding adheres to similarity relations.
- When we perform RSA we look for levels at which features cause differences in the similarity space. Faces do not cluster together in V1, but form semantic clusters in IT. In IT there is invariance (effect of nuissance variation), but differences do not seem to contribute much. So on which level is the information about this feature an „exploitable information”
- And is this „exploitable information” or „exploitable structural correspondence”

- In the case of arbitrary coding – signs correspond to targets not in the virtue of their similarity to targets. Nor they are similar to other vehicles that are bound to similar targets.
- Mappings are definitive this way (x:1, injective)
- However, in similarity-based account of information, this is plausibly not the case. Similarity is notoriously promiscuous (too many correspondences problem)
- But on other hand, there can be reliable information-extraction on the basis of similarity between vehicles.
  - But can it be done only on via „second order” properties? (only the combinations of detectors – structure: can be similar, not any two given detectors (even if they encode a very „close” property: e.g. patch angle just slightly different from another)
  - Disjunctive challenge of Fodor?