Embodiment & Reference

- Embodiment — existence in the real world, groundedness, and contact with physical things — is thought to be a precondition for intelligence [2]

- However, evidence from both humans and LLMs support that the lack of embodiment does not always hinder reference

  - Helen Keller, who was both blind and deaf, had a color scheme that she inferred from smells and touches; she projected one dimension on another [2]

  - Experiments show LLMs trained only on text can recover key geometry of color space, navigation, shapes [2, 4, 7]: e.g. text-only LLMs can draw und understand what they are drawing [3]

  - While LLMs cannot actually see or perform actions, they can do so via a surrogate [3]

  - Given a textual feedback from the environment, LLMs can adjust their generations accordingly [3]

- => LLMs end up learning a „great deal of embodied knowledge" [2]

- Imagine an world with another physics we cannot access; given textual description detailed enough, we can understand how things work there and can imaginary „live" there [2]

Internal States

- All said above makes us conclude there should be something more behind meaning than just reference

- Opinion: the interrelation of concepts is the key to meaning, and reference is just one optional aspect in world conceptualization [4]

  - When the associated terms shift in meaning, the resulting meaning adjusts [4]

  - Thus, a concept builds on conception of how the underlying pieces relate [4]

  - => the search for meaning should focus on understanding the way that the systems' internal representational states relate to each other [4]

- LLMs can manipulate complex concepts (but only in a linear manner) [3, 4]

- fMRI evidence supports that the human mental models for representational geometry are similar to LLMs [4] (?)

  - Cf. embedding and their geometrical meaning [7]

- LLM's internal state has some notions of conceptual role, so LLM's utterances have the semantic intent corresponding to these roles. [4]

- => Opinion: LLMs likely already share the foundation of how our own concepts get their meaning [4]