

Virtual Embodiment: A Scalable Long-Term Strategy for Artificial Intelligence Research

Douwe Kiela¹, Luana Bulat², Anita L. Vero², Stephen Clark^{2,3}

¹Facebook AI Research ²University of Cambridge ³Google DeepMind

dkiela@fb.com

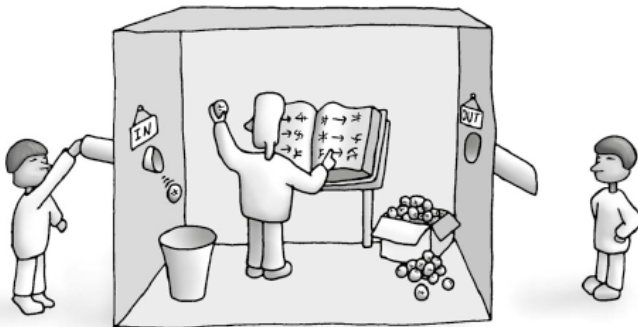
December, 2016

Meaning is AI's holy grail (too)

“Meaning is the ‘holy grail’ not only of linguistics, but also of philosophy, psychology, and neuroscience.”
— Jackendoff, 2002




Symbol grounding problem



*How can you know the meaning of a symbol
if it is defined through other symbols?*

Grounding problem in semantics

democracy

/di'mɒkrəsi/ 

noun

a system of government by the whole population or all the eligible members of a state, typically through elected representatives.
"a system of parliamentary democracy"

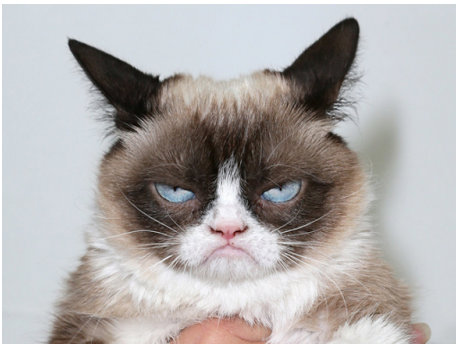


cat¹

/kæt/ 

noun

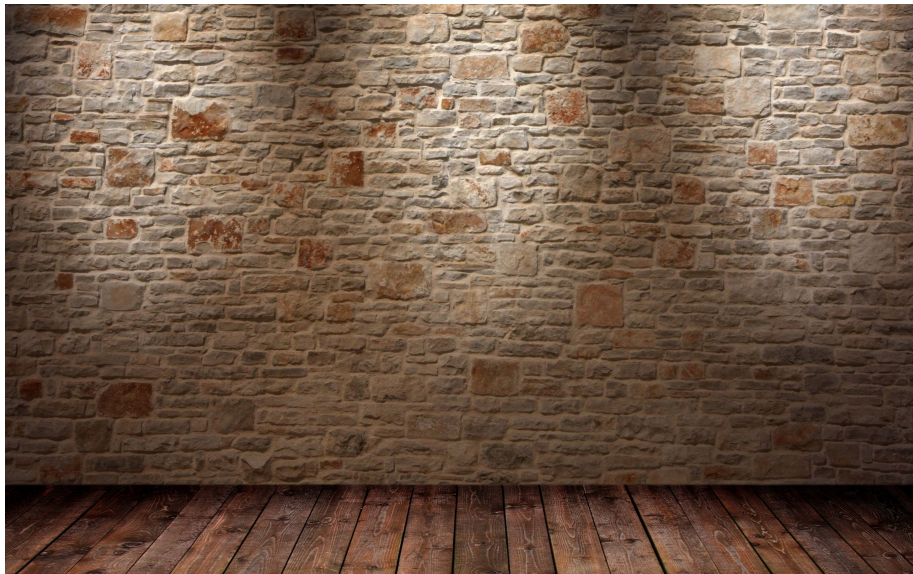
1. a small domesticated carnivorous mammal with soft fur, a short snout, and retractile claws. It is widely kept as a pet or for catching mice, and many breeds have been developed.



Meaning is multi-modal and grounded in sensori-motor experience!

Glenberg & Robertson 2000; Barsalou 2008; Andrews et al. 2009; Baroni et al. 2010; Riordan & Jones 2011; Bruni et al. 2014

The meaning of bumping into walls



Problems with physical embodiment



- Difficult with current technology
- Very expensive
- Not scalable
- Ethically questionable

Virtual embodiment



Advantages of virtual embodiment

- Scalability
 - Complexity of virtual worlds can scale with artificial agent capabilities
- Long-term feasibility
 - Feasible now, but control over virtual world complexity implies focused long-term challenge
- Rapid iteration
 - We can improve iteratively, at great speed
- No human-in-the-loop requirement
 - Humans are useful to learn from, but agents may learn by communicating with each other
- Ethical testability
 - Virtual paperclip maximizers are harmless

Video games with a purpose



Desiderata

- Useful to outline the characteristics of virtual embodiment-compatible video games.
- Inspired by Kardashev scale for complexity of civilizations in physics, we propose a type hierarchy.



Type hierarchy of (virtual) world complexity

- **Type 0:** Agents perform basic first-order interactions with the world, with full or limited access to the objective world state. No intra-agent communication is required.
- **Type 1:** As above, but without any state access. Communication may be used for sharing knowledge about the state of the world.
- **Type 2:** As above, but with higher-order interactions, i.e., with an element of planning, strategy and non-monotonic reasoning. Communication is essential for sharing knowledge about the world.

Type hierarchy of (virtual) world complexity

- **Type 3:** The world should be strictly non-deterministic and multi-modal. This makes communication essential for not only sharing knowledge about the world, but also for sharing plans and strategies.
- **Type 4:** Agents should be multi-objective, that is, an agent's objective or reward function should be a weighted function of various objectives or rewards, that depend both on the state of the world and current plans and strategy.
- **Type 5:** Multi-objective agents interact with and communicate about a non-deterministic world in such a way that it allows for them to plan ahead and form and execute sophisticated strategies.

Conclusion

- Long way to go before we can achieve Type 5 embodiment, which is closest to physical reality.
- Virtual embodiment is a realistic, scalable, long-term strategy for artificial intelligence research

