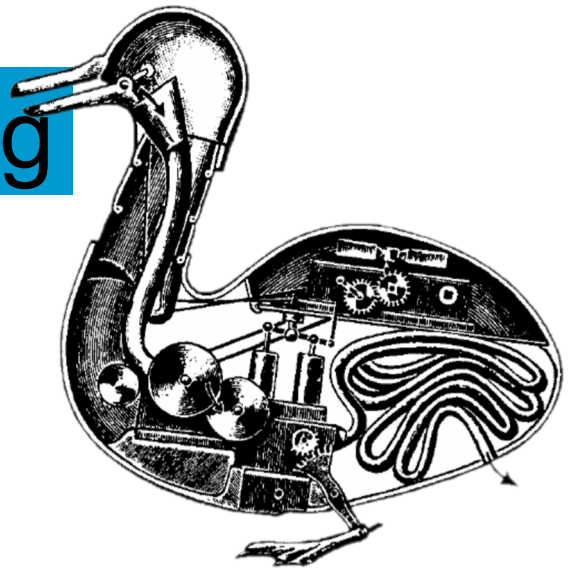
Universität Konstanz

Why LLMs are (still?) mechanical turks:

Two levels of grounding

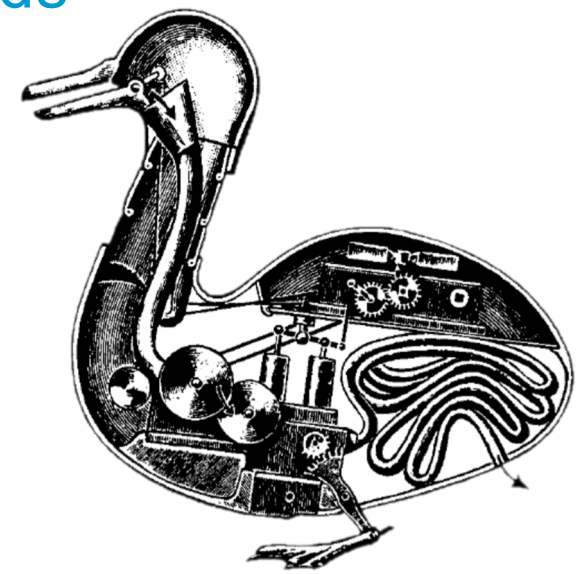


Dorothea Debus, Regine Eckardt

AIAI — May 16, 2024

Outline

1. Background: Do LLMs master meaning?
2. Linguistics: Meanings and truth conditions
3. Philosophy: Meanings and minds
4. Summary



Do LLMs master meaning?

CON camp

PRO camp

Do LLMs master meaning?

CON camp

Bender+Koller (2020): (a) A system trained on form only can *a priori* not master the world-word link inherent in meaning. (b) LLMs cannot learn “communicative intent” (speakers’ aims behind speaking).

PRO camp

Do LLMs master meaning?

CON camp

Bender+Koller (2020): (a) A system trained on form only can *a priori* not master the world-word link inherent in meaning. (b) LLMs cannot learn “communicative intent” (speakers’ aims behind speaking).

PRO camp

Søgaard (2023): Proper communicative behavior possible without c-intent. First evidence for isomorphisms between LLM word topology and “the world”; novel perspective on grounding.

Do LLMs master meaning?

CON camp

Campbell, J. (2002), consciousness; Harnad, Stevan (1990) grounding; Lake, B. and G. Murphy (2021) psychology; Pezzulo et al. (2024) cognitive science

PRO camp

Abdou, Mostafa et al. (2021) color terms; Carta, T. et al. (2023); Tom Brown et al. (2020), Jean-Baptiste Alayrac et al. (2022), Ahn, M. et al. (2022) (various implementations)

Do LLMs master meaning?

CON camp

Campbell, J. (2002), consciousness; Harnad, Stevan (1990) grounding; Lake, B. and G. Murphy (2021) psych… Pezzulo et al. (2024) cognitive science

PRO camp

Abdou, Mostafa et al. (2021) …terms; Carta, T. et al. (2023); Tom Brown et al. (2020), Jean-Baptiste Alayrac et al. (2022), Ahn, M. et al. (2022) (various implementations)

Missing:
Test for „Meaning“

Formal Linguistics: Truth conditional semantics

- truth conditional models of natural language semantics
- extension, intension of words and phrases
- logical type hierarchy
- compositionality
- semantics operationalized:
  truth value judgement tasks

Heim + Kratzer (1998): Semantics in Generative Grammar. Malden: Blackwell.

Linguistics: Truth value judgement tasks

To know the meaning of sentence S
= To know the conditions under which S is true
D. Davidson (1967)

’Snow is white’ is true iff snow is white.

Linguistics: Truth value judgement tasks

Truth value judgements are not schematic paraphrases.

(a) There are three crocodiles on the table.
(b) There are three animals on the table.
(c) Two elephants meet a parrot.

Linguistics: Truth value judgement tasks

Truth value judgements in Language Acquisition (Crain et al. 2000)

(1) Is every boy riding an elephant?

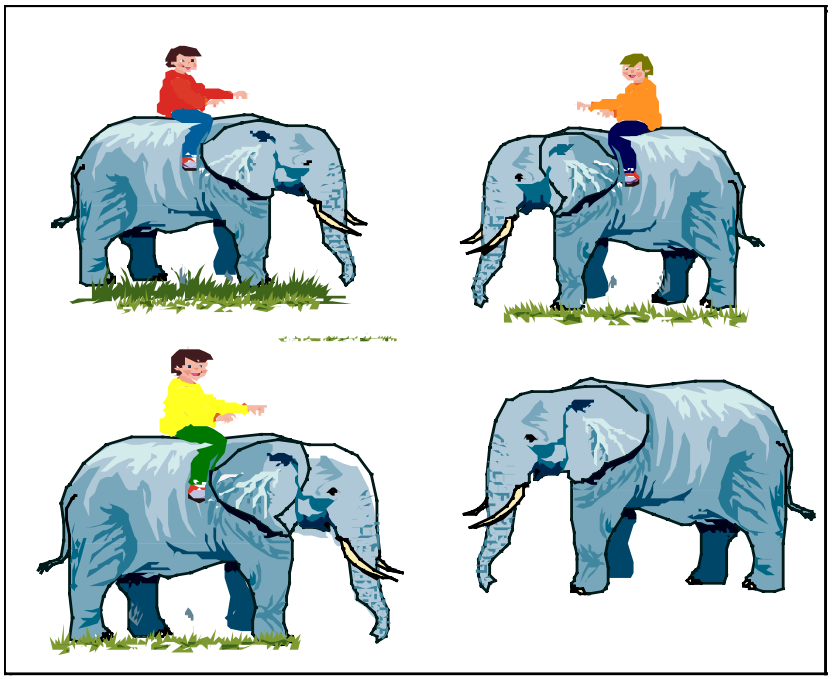


Figure 1. The Extra-Object Condition

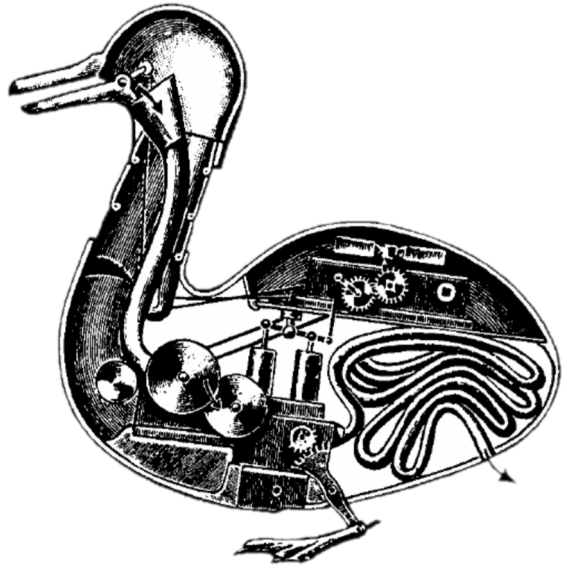
Linguistics: Truth value judgement tasks

Can LLMs master TVJ tasks?

- decide for a substantial range of nouns N: what is the extension of N in a given situation?
- decide for a substantial range of verbs V: who is engaged in the V-activity in a given situation?
- decide for a substantial range of adjectives A: which object(s) have property A in a given situation?
- decide for a given *world/situation* whether S is true or not.

LLMs in TVJ-tasks need human confederates

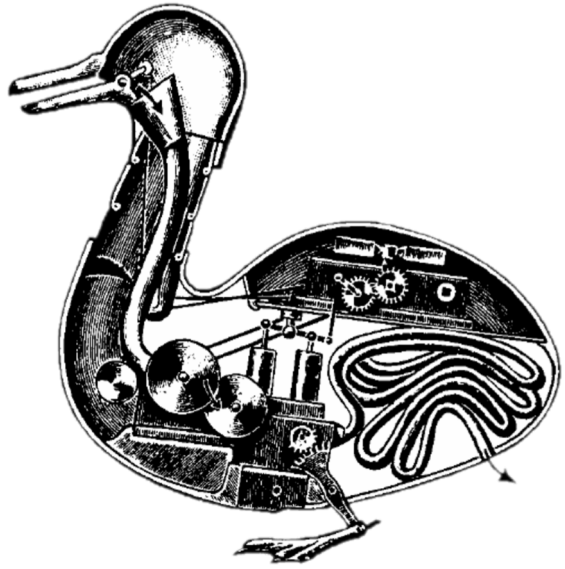
- LLMs in virtual reality (Deepmind's MIA, Chalmers 2022)
- LLMs in coded reality (Steinert-Threlkelt et al. 2019, Carta, T. et al. 2023)
- LLMs with human confederates ("Tell me the age of Heinz Müller.", ChatGTP)



LLMs in TVJ-tasks need human confederates

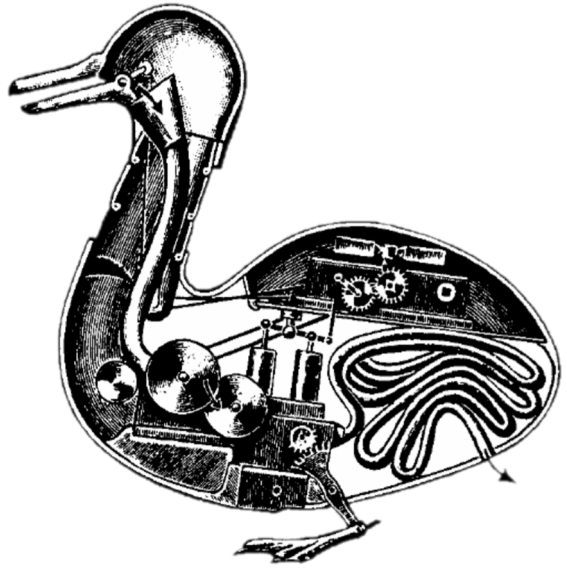
- LLMs in virtual reality (Deepmind's MIA, Chalmers 2022)
- LLMs in coded reality (Steinert-Threlkelt et al. 2019, Carta, T. et al. 2023)
- LLMs with human confederates ("Tell me the age of Heinz Müller.", ChatGTP)

- LLMs can't do TVJ tasks unless a human helper provides the necessary true assertions about the world.



Can Human communication be non-grounded (sometimes)?

- Situations of non-grounded HUMAN communication
- Is Human communication successful in such contexts?



Human non-grounded communication

grounding the language of wine tastes:
no intersubjective truth conditions

- “Is this wine mineralic?”
  answers to questions: no intersubjective truths

Human non-grounded communication

grounding the language of wine tastes:
no intersubjective truth conditions

- “Is this wine mineralic?”
  answers to questions: no intersubjective truths
- “Give me a dry white wine.”
  speech acts: perlocutionary acts fail, due to lack of grounding

Human non-grounded communication

grounding the language of wine tastes:
no intersubjective truth conditions

- “Is this wine mineralic?”
  answers to questions: no intersubjective truths
- “Give me a dry white wine.”
  speech acts: perlocutionary acts fail, due to lack of grounding

- (ps: not a personal-taste-predicate problem)

Human non-grounded communication

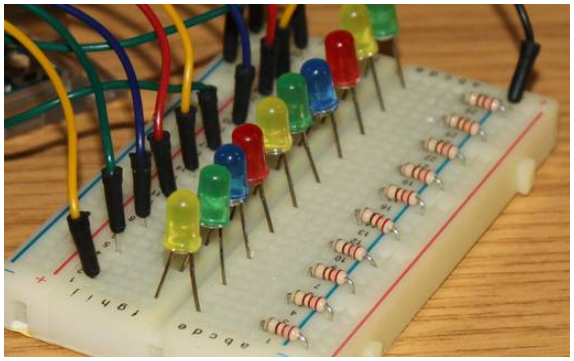
grounding the language of color:
speaker X who cannot perceive color
(e.g. can only see black-grey-white).

- X has read “everything” about color (≈ LLM)
- X is unable to do TVJ tasks on sentences with color terms.
- X is unable to perform instructions that rest on color terms.

Human non-grounded communication

grounding the language of color:
speaker X who cannot perceive color
(e.g. can only see black-grey-white).

- X has read “everything” about color (≈ LLM)
- X is unable to do TVJ tasks on sentences with color terms.
- X is unable to perform instructions that rest on color terms.

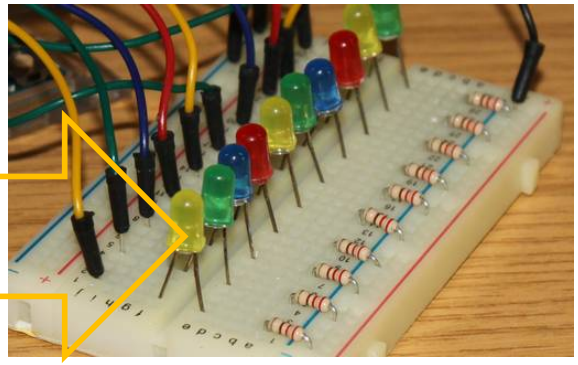


Human non-grounded communication

grounding the language of color:
speaker X who cannot perceive color
(e.g. can only see black-grey-white).

- X has read “everything” about color (≈ LLM)
- X is unable to do TVJ tasks on sentences with color terms.
- X is unable to perform instructions that rest on color terms.

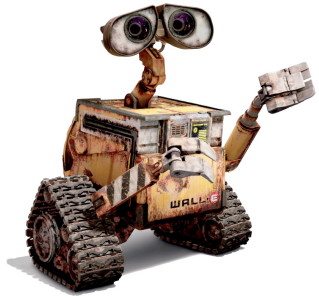
„it‘s yellow“



TVJ-tasks: Interim Summary

- *Grounding* ≈ Ability to determine words' extensions, sentences' truth value in (most) possible situations
- TVJ-tasks ≈ Tests for semantic competence
- Humans can locally lack this competence.
  ➔ Communication is imperfect under these circumstances.

- LLMs are not trained for TVJ-tasks.
- LLM communication is imperfect.
- How about LLMs+ (Chalmers 2022)?



‚Grounding', perceptual experience and demonstrative reference

- Healthy human subjects meet the 'grounding condition' on the basis of the various perceptual experiences they have of their environment.

- A healthy human subject can demonstratively refer to objects in her environment on the basis of her perceptual experiences.

- Perceiving one's environment, and being able to demonstratively refer to objects in one's environment, seem to be core features of 'being grounded'.

Two theories of perception: representationalism vs direct realism

Two philosophical accounts of perceptual experience:

representationalism

vs

direct realism (or ‘naïve realism’ or ‘relationalism’)

Representationalism

Perceptual experiences are representational states, and are ultimately of the same nature as illusory and hallucinatory experiences.

Direct realism

Perceptual experiences are relational states – a perceiving subject stands in a particular relation to the object perceived, and this relation is constitutive of the experience.

Direct realism

(Core Claim) perceptual experiences are ‘genuinely relational’, that is, perceptual experiences are relational essentially.

Amongst other things, this entails that

(Relation Claim) when a subject perceives an object, the subject stands in an experiential *relation* – namely, a *perceptual* relation – to the relevant object.
(Constitution Claim) A perceived object itself is a *constitutive part* of the relevant perceptual experience.
(Consciousness Claim) A perceived object is, for the perceiving subject, ‘immediately available in consciousness’ (McDowell 1978: 138).

Demonstrative reference

Sometimes, subjects can, on the basis of perceptual experiences, successfully engage in demonstrative reference.

(e.g.: ‘this chair’, ‘this table’, ‘this person’)

In these contexts,

‘experience of objects has an explanatory role to play: it explains our ability to think demonstratively about perceived objects. Experience of a perceived object is what provides you with knowledge of the reference of a demonstrative referring to it’ (Campbell 2002: 114).

The problem with representationalism I

According to the representationalist, experiences are

‘states of a type whose intrinsic mental features are world-independent; an intrinsic, or basic characterization of a state of awareness will make no reference to anything external to the subject. But if that is what experience is like, (…) how can it yield knowledge of an objective world beyond experience, and how can it so much as put us in a position to think about such a world?’

(Child 1994: 146-147)

The problem with representationalism II

‘experience, conceived from its own point of view, is not blank or blind, but purports to be revelatory of the world we live in.’

(McDowell 1986: 152)

The explanatory role of perceptual experience in demonstrative reference

According to direct realism,

(Consciousness Claim) a perceived object is, for the perceiving subject, ‘immediately available in consciousness’ (McDowell 1978: 138).

But then,
a subject who demonstratively refers to an object which she perceives knows what the reference of her demonstrative is simply because the object is ’immediately available in consciousness’ to the subject in perceptual experience.

‚Grounding', perceptual experience and demonstrative reference (again)

Perceiving one's environment, and being able to demonstratively refer to objects in one's environment, seem to be core features of 'being grounded'.

The core question

Assume we augment LLMs with vision (camera), smell (chemical sensors), touch (robot arms) or other, similar input systems.

Could such augmented LLMs be ‘grounded’?

The Core Argument

(i) Cameras, chemical sensors or robot arms (or other, similar input devices) can only provide a system with relevant *representational* content.

(ii) A system which only has access to *representational* content could not possibly be grounded.

(Rather, in order for a system to be ‘grounded’, it needs to stand in an *experiential relation* to its environment.)

Thus,

(C) **LLMs augmented with cameras (‘vision’), chemical sensors (‘smell’), robot arms (‘touch’) or other similar input devices could not possibly be ’grounded’.**

In support of premise (ii)

(1) According to a representationalist account of perceptual experience, perceptual experience could not possibly ‘put us in a position to think about’ an ‘objective world beyond experience’ (Child 1994: 146-7), that is, if perceptual experiences were nothing but representational states, they could not explain our ability to demonstratively refer to objects in our environment.

(2) But then, this should translate to other, non-human systems: A system which only has access to representational content, that is, amongst other things, a system whose only access to the external world is via representational states, could not possibly refer to objects in its environment demonstratively.

(3) But then, being able to demonstratively refer to objects in one’s environment seems to be a core feature of ‘being grounded’.

(C) Thus, a system which only has access to *representational* content could not possibly be grounded, just as premise (ii) has it.

The Core Argument

(i) Cameras, chemical sensors or robot arms (or other, similar input devices) can only provide a system with relevant *representational* content.

(ii) A system which only has access to *representational* content could not possibly be grounded.

(Rather, in order for a system to be ‘grounded’, it needs to stand in an *experiential relation* to its environment.)

Thus,

(C) **LLMs augmented with cameras (‘vision’), chemical sensors (‘smell’), robot arms (‘touch’) or other similar input devices could not possibly be ’grounded’.**

In Conclusion

Levels of Grounding:

(1) Mastering truth-value judgement tasks (without human confederates)

(2) Demonstrative reference

Thank you!

References:

Abdou, Mostafa, A. Kulmizev, D. Hershcovich, S. Frank, E. Pavlick, and A. Søgaard. 2021. Can language models encode perceptual structure without grounding? A case study in color. *Proceedings of the 25th Conference on Computational Natural Language Learning*.

Bender, Emily and A. Koller. 2020. *Climbing towards NLU*. Proc. of the 58th Annual Meeting of the Association for Computational Linguistics, 5185–5198. Online.

Browning, Jacob and Y. Lecun. 2022. AI and the limits of language. *Noēma*.

Campbell, J. (2002): Reference and Consciousness. Oxford: Oxford University Press.

Carta, T. et al. 2023. Grounding large language models in interactive environments with Online reinforcement learning. *Proc. of the 40th Intl. Conf. on Machine Learning*, Honolulu.

Chalmers, D. 2022. Could a LLM be conscious? Talk ms., presented at the NeurIPS New Orleans, November 2022. Online https://nips.cc/virtual/2022/invited- talk/55867.

Child, W. (1994): *Causality, Interpretation and the Mind*. Oxford: OUP.

Crain, Stephen. 2017. Acquisition of Quantifiers. Annual Review of Linguistics, Vol. 3, Issue 1, pp. 219-243. SSRN: https://ssrn.com/abstract=2905688 or http://dx.doi.org/10.1146/annurev-linguistics-011516-033930

Crane, T. and French, C. 2021. The Problem of Perception. The Stanford Encyclopedia of Philosophy. <https://plato.stanford.edu/archives/fall2021/entries/perception-problem/>.

References:

Davidson, D. 1967. Truth and Meaning. *Synthese* 17, 304 - 323.

Harnad, St. 1990. The symbol-grounding problem. *Physica* D (42):335-346.

Heim, I. and A. Kratzer. 1998. *Meaning in generative grammar*. Malden: Blackwell.

Jackson, F. 1986. What Mary Didn't Know. *The Journal of Philosophy*, Vol. 83. 291-295.

Lake, B. and G. Murphy. 2021. Word meaning in minds and machines. *Psychological Review*.

Long, R. 2022. *Key questions about artificial sentience: An opinionated guide. Experience Machines* (Substack), 2022.

McDowell, J. (1986): 'Singular Thought and the Boundaries of Inner Space', in: J. McDowell and P. Pettit (eds.), *Subject, Thought and Context*. Oxford: OUP. 137-168

Meroni, L., Gualmini A. and Stephen Crain. 2000. A Conservative Approach to Quantification in Child Language. In: U. Penn Working Papers in Linguistics, Volume 7.1, 2000

References:

Pezzulo, Giovanni et al. 2024. Generating meaning: active inference and the scope and limits of passive AI. *Trends in Cognitive Sciences* 28(2): 97–112.

Søgaard, A. 2023. Grounding the vector space of the octopus: Word meaning from raw text. *Minds and Machines* 33, 33 - 54.

Steinert-Threlkeld, Shane and Jakub Szymanik. 2019. Learnability and semantic universals. *Semantics and Pragmatics* 12(4). https://doi.org/10.3765/sp. 12.4.

Wittgenstein, Ludwig. 1998 [1921]. *Logisch-philosophische Abhandlung, Tractatus logico-philosophicus. Kritische Edition.* Frankfurt am Main: Suhrkamp.

References:

Software / LLM / LLM+

Brown, Tom et al. 2020. Language models are few shot learners. arXiv:2005.14165, 2020. (GPT-3)

Alayrac, Jean-Baptiste et al. 2022. *Flamingo*: a Visual Language Model for Few-Shot Learning. Proceedings of Neural Information Processing Systems, 2022.

Ahn, Michael et al. 2022. Do as I can, not as I say: Grounding language in *robotic affordances*. https://say- can.github.io/. 2022.

Abramson, Josh et al, 2021. Creating *multimodal interactive agents* with imitation and self-supervised learning. arXiv:2112.0376.