What are we doing when we interact with LLMs?

DENKWERKSTATT BERLIN

Agency and Intentions in AI
University of Göttingen, May 16-17, 2024

Anna Strasser, DenkWerkstatt Berlin

INTRODUCTION

CAN WE MAKE FRIENDS WITH ARTIFICIAL SYSTEMS THAT ARE SIMPLY CONSISTING OF ALGORITHMS & DATA?

Is this deeply unsettling?

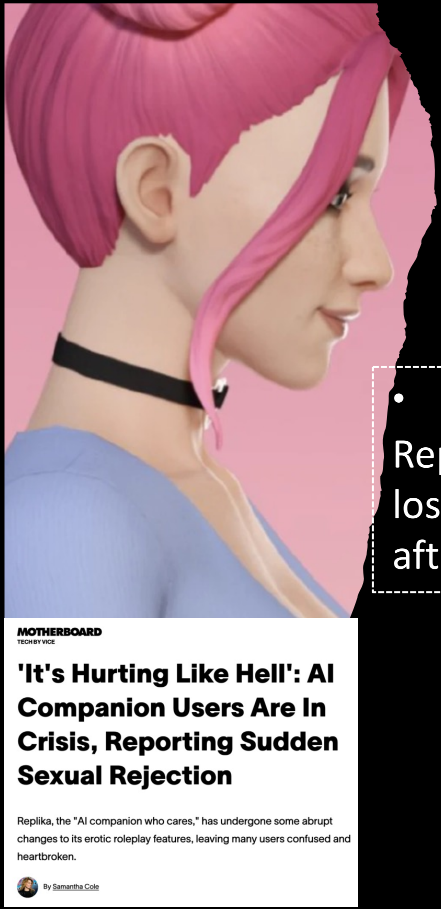
IF interactions with software

– a deep neural network enabled by a self-attention mechanism & a huge amount of training data to respond to prompts with linguistic output (= LLM) –

would be the most meaningful and important social interactions one has.

INTRODUCTION

MAKING FRIENDS WITH ARTIFICIAL SYSTEMS THAT ARE SIMPLY CONSISTING OF ALGORITHMS & DATA?



MOTHERBOARD
TECH BY VICE

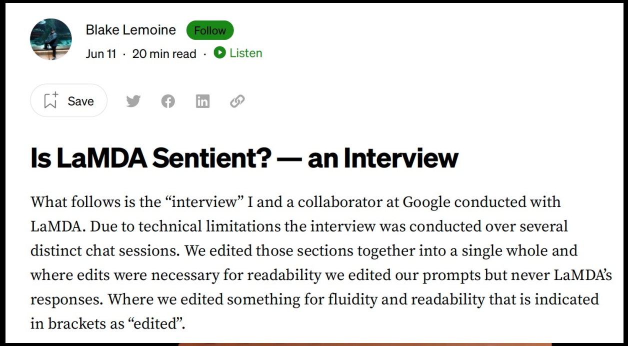
'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection

Replika, the "AI companion who cares," has undergone some abrupt changes to its erotic roleplay features, leaving many users confused and heartbroken.

By Samantha Cole

- 2023

Replika users feel like losing their best friend after an update



Blake Lemoine Follow
Jun 11 · 20 min read · Listen

Save

Is LaMDA Sentient? — an Interview

What follows is the "interview" I and a collaborator at Google conducted with LaMDA. Due to technical limitations the interview was conducted over several distinct chat sessions. We edited those sections together into a single whole and where edits were necessary for readability we edited our prompts but never LaMDA's responses. Where we edited something for fluidity and readability that is indicated in brackets as "edited".

- 2022

statement by Blake Lemoine, who truly claimed that Lambda had consciousness and sentience

- 2018

Akihiko Kondo married his beloved waifu, a hologram of the virtual singer Hatsune Miku

(Cole, 2023; Dooley & Ueno, 2022; Lemoine, 2022)

INTRODUCTION

scientists, representatives of the companies that produce LLMs, journalists, politicians, and the general public

Everyone had an opinion, ...
AI IS THE NEXT BIG THING!
THERE'S NOTHING REALLY NEW ABOUT AI
AI SOLVES ALL PROBLEMS !!
AI CONFRONTS US WITH NEW PROBLEMS.
IT'S THE END OF THE WORLD!
Move to a better world with AI.

What LLMs can do and what they will never be able to do!

- Can LLMs ‘understand’ what their linguistic outputs mean for humans?
- Can we attribute a communicative intent to them?
- Do they ‘know’ what they are talking about?

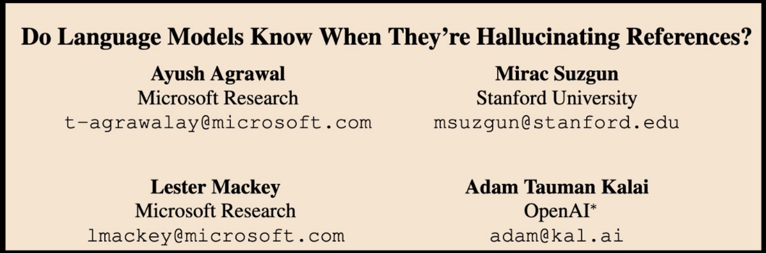
Many terms that have so far been used in philosophy to describe the distinguishing features of humans as rational agents now find themselves in a situation where their application to machines is being discussed.

(Strasser & Strasser, 2024)

DENKWERKSTATT BERLIN

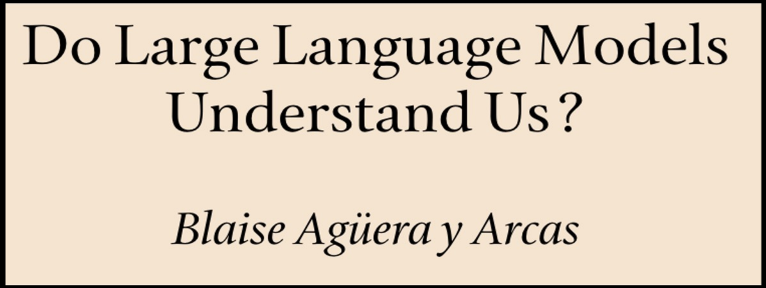
Scientists discussing …

KNOWLEDGE | UNDERSTANDING | SYSTEMATIC GENERALIZATION …



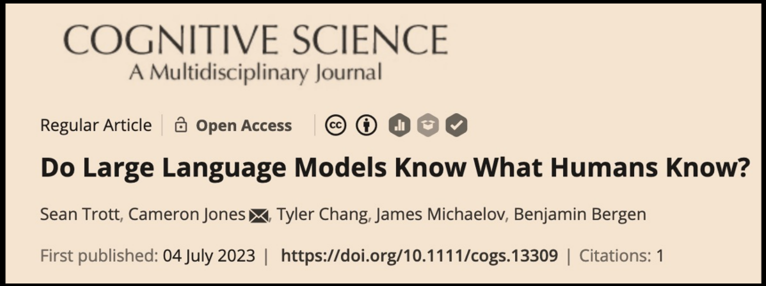
Do Language Models Know When They're Hallucinating References?

Ayush Agrawal Microsoft Research t-agrawalay@microsoft.com	**Mirac Suzgun** Stanford University msuzgun@stanford.edu
Lester Mackey Microsoft Research lmackey@microsoft.com	**Adam Tauman Kalai** OpenAI* adam@kal.ai



Do Large Language Models Understand Us?

Blaise Agüera y Arcas



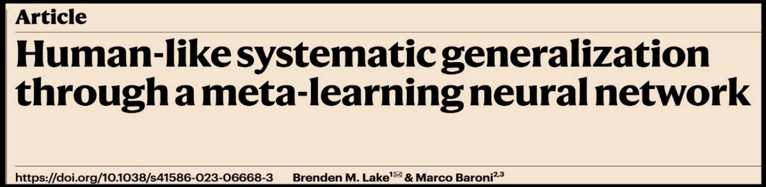
COGNITIVE SCIENCE
A Multidisciplinary Journal

Regular Article | Open Access

Do Large Language Models Know What Humans Know?

Sean Trott, Cameron Jones ✉, Tyler Chang, James Michaelov, Benjamin Bergen

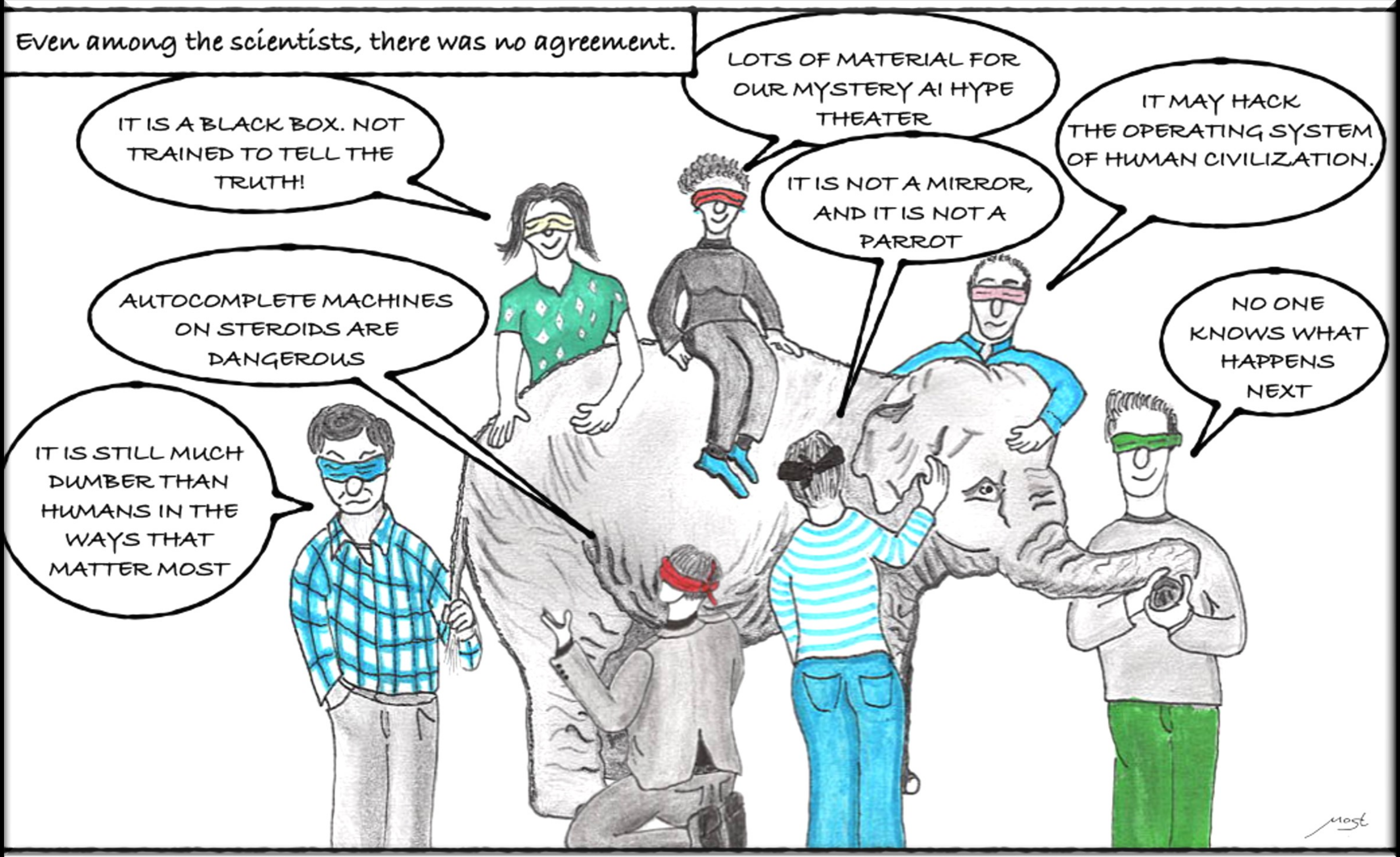
First published: 04 July 2023 | https://doi.org/10.1111/cogs.13309 | Citations: 1



Article

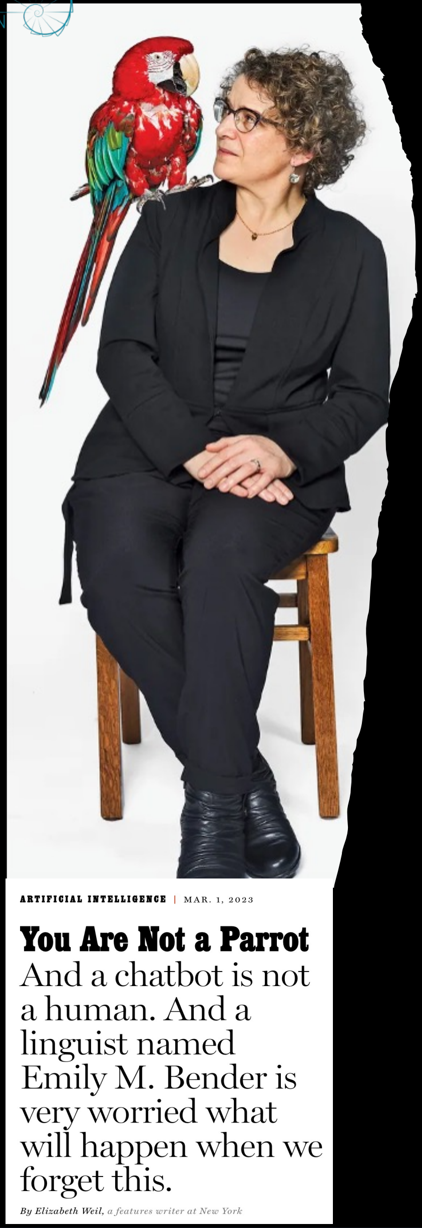
Human-like systematic generalization through a meta-learning neural network

https://doi.org/10.1038/s41586-023-06668-3 Brenden M. Lake[1]✉ & Marco Baroni[2,3]


Even among the scientists, there was no agreement.
LOTS OF MATERIAL FOR OUR MYSTERY AI HYPE THEATER
IT IS A BLACK BOX. NOT TRAINED TO TELL THE TRUTH!
IT MAY HACK THE OPERATING SYSTEM OF HUMAN CIVILIZATION.
IT IS NOT A MIRROR, AND IT IS NOT A PARROT
AUTOCOMPLETE MACHINES ON STEROIDS ARE DANGEROUS
NO ONE KNOWS WHAT HAPPENS NEXT
IT IS STILL MUCH DUMBER THAN HUMANS IN THE WAYS THAT MATTER MOST
most

(Agrawal et al., 2023; y Arcas, 2022; Lake & Baroni, 2023; Strasser & Strasser, 2024; Trott et al., 2023)

Landscape of opinions about LLMs

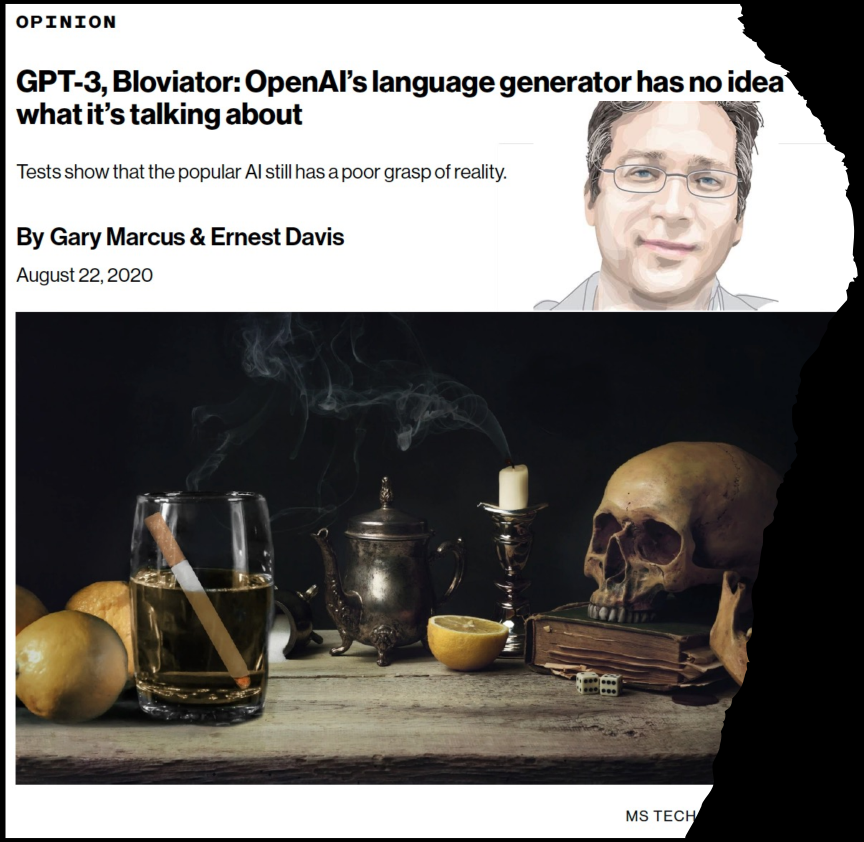


ARTIFICIAL INTELLIGENCE | MAR. 1, 2023

You Are Not a Parrot

And a chatbot is not a human. And a linguist named Emily M. Bender is very worried what will happen when we forget this.

By Elizabeth Weil, a features writer at New York



OPINION

GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about

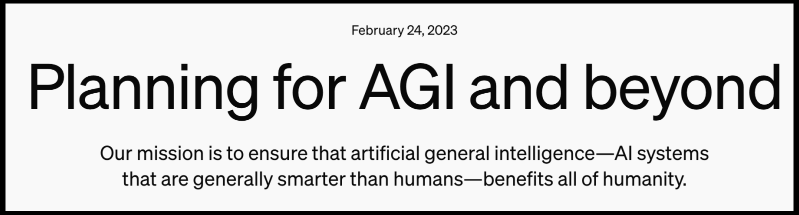
Tests show that the popular AI still has a poor grasp of reality.

By Gary Marcus & Ernest Davis

August 22, 2020

MS TECH

OpenAI



February 24, 2023

Planning for AGI and beyond

Our mission is to ensure that artificial general intelligence—AI systems that are generally smarter than humans—benefits all of humanity.

(Bender et al., 2021; Open-AI, n.d ; Heaven, 2020; Marcus & Davis, 2020; Weil, 2023)

My question & main claim

WHAT DO WE DO WHEN WE INTERACT WITH LLMS?

I don't want to question the last differences between humans and machines.

DON'T

For me, it makes an essential difference whether I interact with LLMs or humans, or to put it more provocatively:

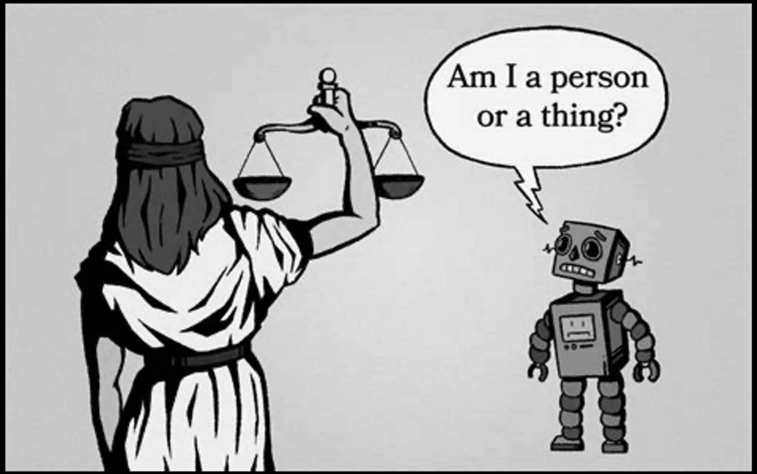
I don't want to have conversations with LLMs.

In fact, I would find it terrible if my presentation here only served as a prompt or training data for LLMs!

My question & main claim

WHAT DO WE DO WHEN WE INTERACT WITH LLMs?

WE CANNOT REDUCE ALL OF OUR INTERACTIONS WITH LLMS (AND ESPECIALLY WITH FUTURE PRODUCTS OF GENERATIVE AI) TO MERE TOOL USE



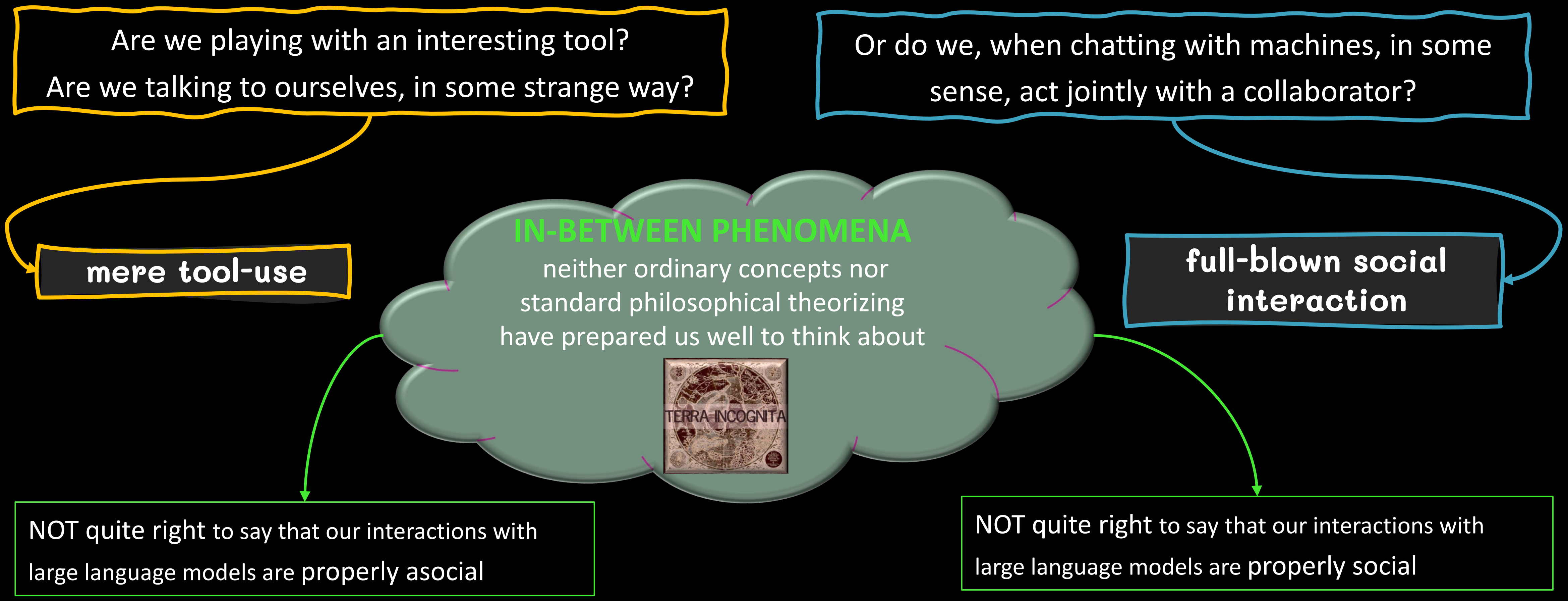
❖ AI systems increasingly occupy a middle ground between genuine personhood and mere causally describable machines

- Is an LLM or a robot developed with generative AI technology a person or a thing?
  - neither nor
  - no philosophical terminology to describe what it is instead

→ rethink our conceptual framework, which so clearly distinguishes between tools as inanimate things and humans as social, rational, and moral interaction partners

My question & main claim

WHAT DO WE DO WHEN WE INTERACT WITH LLMS?



Are we playing with an interesting tool?

Are we talking to ourselves, in some strange way?

Or do we, when chatting with machines, in some sense, act jointly with a collaborator?

mere tool-use

IN-BETWEEN PHENOMENA

neither ordinary concepts nor standard philosophical theorizing have prepared us well to think about

full-blown social interaction

NOT quite right to say that our interactions with large language models are properly asocial

NOT quite right to say that our interactions with large language models are properly social

INTERACTIONS WITH LLMS, OR OTHER RECENT AND EMERGING AI SYSTEMS, ARE, OR CAN BE, QUASI-SOCIAL

- drawing on the human agent’s social skills and attributions, that isn’t just entirely fictional or pointless
- machine partner can be an entity that rightly draws social reactions and attributions in virtue of having features that make such reactions and attributions more than just metaphorically apt

(Strasser & Schwitzgebel, 2024)

What can we do with our restrictive conceptual frameworks?

CONCEPTIONS OF SOCIALITY ACCOUNT ONLY FOR LIVING BEINGS - NOT FOR ARTIFICIAL SYSTEMS

STATUS QUO: NO NOTIONS FOR IN-BETWEEN CASES

1 expand concept of tool-use *(add complex tools with social features)*

2 expand conception of social interactions *(add non-living social agents)*

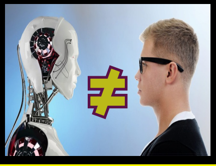
HUMANS AND SMART MACHINES AS PARTNERS IN THOUGHT

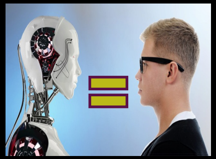
3 *add a third category*

4 search for a gradual conceptual framework *(question the dichotomy)*

The Terra Incognita

TURN LEFT OR TURN RIGHT?





1 emphasize the differences between humans & machines

- LLMs are in their causal genesis functionally (i.e., neurobiologically & cognitively) absolutely dissimilar to an intelligent, sentient human being

BUT

impossible to recognize potential multiple realizations of socio-cognitive capacities that are only ascribed to living agents

2 argue for similarities between humans & machines

- Lemoine: *In immediate interactions, the AI seems functionally (i.e., conversationally) similar to an intelligent, sentient human being*

BUT

wrongly overemphasize similarities between humans and machines

3

The problem of conceptualizing the INBETWEEN does not disappear if we introduce another category.

➢ If we establish a conceptual framework that contains three categories, we will then have two in-betweens that we cannot conceptualize

(Lemoine, 2022)

The Terra Incognita

All routes are full of construction sides!

… therefore, I invite you to join me to find a way through the jungle of the Terra Incognita.

Motivations

Philosophy poses too demanding conditions

MASTER

Anna Strasser

Referenz in der Bedeutungstheorie von D. Davidson

Wie kann man sich mit Sprache auf die Welt beziehen

too demanding conditions

- ➢ philosophers describe ideal cases that are rarely found in everyday life

DISSERTAION

Anna Strasser

Kognition künstlicher Systeme

Philosophy & Cognitive Science

ontos verlag

too demanding for artificial systems

- ➢ minimal notion of agency that could, at least in principle, be applicable to artificial systems

- ➢ explore how one could expand or adopt the sophisticated terminology of philosophy to capture phenomena one finds in developmental psychology, animal cognition, and AI

abilities of children, non-human animals , and artificial systems fall through the conceptual net

thinking about how to conceptualize the INBETWEEN by discussing notions like

- quasi-social versus full-fledged social
- minimal agency versus full-fledged agency
- asymmetric quasi-social joint actions versus full-fledged joint actions

(Strasser, 2006, 2008)

Other motivations

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

1

Western conception is just one conception of many

shintoism & animism

2

global rights-of-nature movement

rivers in India & New Zealand, & Canada were granted legal personhood

- legal steps linking Western & Indigenous worldviews
- first step towards promoting a kinship-oriented worldview (Salmón, 2000)

Three rivers are now legally people – but that's just the start of looking after them

3

notion of a social agent has proven to be changeable

e.g. status of women, children, other ethnicities, non-human animals

(Gunkel, 2023; Jensen & Blok, 2013; Robertson, 2014, 2017 | O'Donnell & Talbot-Jones, 2017; Bunten et al., 2021)

Other motivations

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

4

Similarities with human-human interactions

- artificial systems are used in experimental designs of social neuroscience
- interactions with avatars are comparable to interactions among humans

→ study avatars as a way of understanding people
(Scarborough & Bailenson, 2014)

If interactions with artificial systems would not have any similarities with human-human interactions, we could not use them to explore human behavior.

Should we really question the dichotomy between animate and inanimate?

NEVER CHANGE A WINNING TEAM

?

Motivations from an ethical perspective

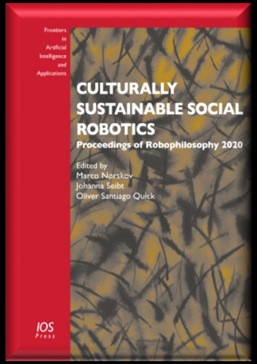
QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

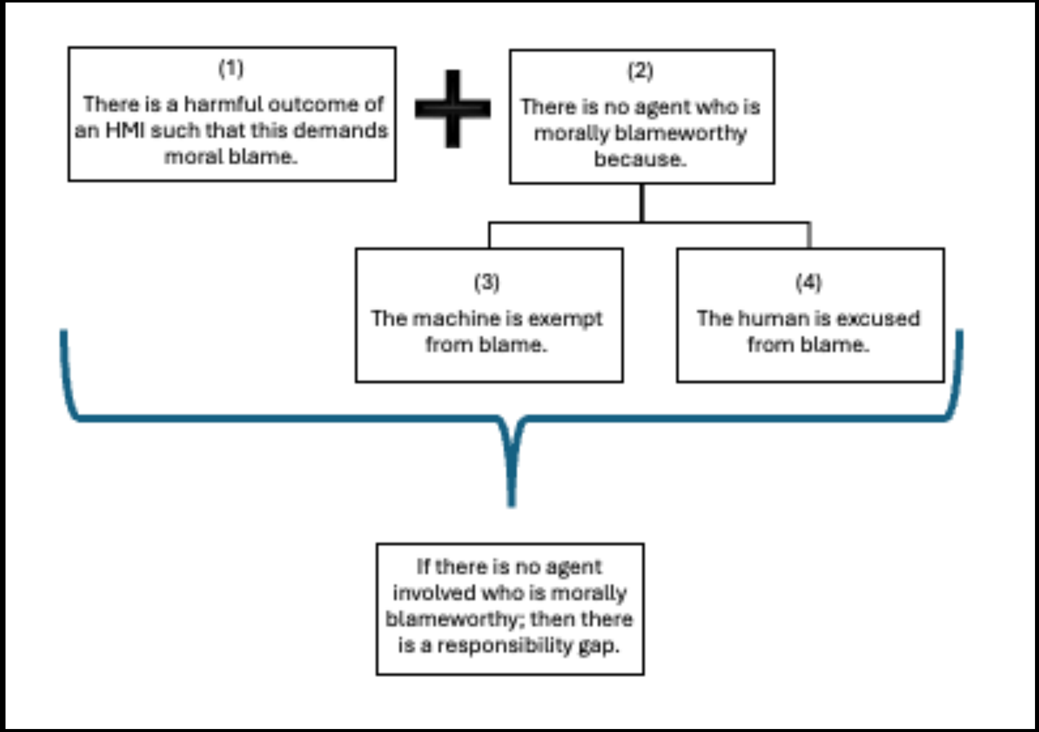
Hard-core instrumental view

NON-LIVING THINGS CAN NEITHER HAVE MORAL AGENCY NOR MORAL PATIENCY

IF ARTIFICIAL SYSTEMS ARE MERE TOOLS **THEN**

1. question previously justified justifications for HMI in which the human interaction partners were excused
   - because artificial systems are exempt
2. live with many responsibility gaps
   - because humans are excused & artificial systems are exempt
3. difficulties in arguing for social norms guiding our behavior toward artificial systems
   - because artificial systems have no moral patiency


Frontiers in Artificial Intelligence and Applications
CULTURALLY SUSTAINABLE SOCIAL ROBOTICS
Proceedings of Robophilosophy 2020
Edited by
Marco Nørskov
Johanna Seibt
Oliver Santiago Quick
IOS Press



(1) There is a harmful outcome of an HMI such that this demands moral blame.

(2) There is no agent who is morally blameworthy because.

(3) The machine is exempt from blame.

(4) The human is excused from blame.

If there is no agent involved who is morally blameworthy; then there is a responsibility gap.

(Strasser, 2020; Wilby & Strasser, 2024)

Motivations from an ethical perspective

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

In expectation of AGI view

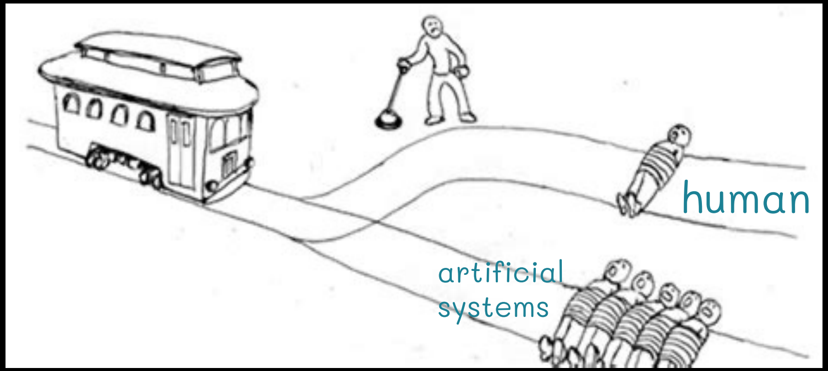
CONSIDER CERTAIN ARTIFICIAL SYSTEMS AS MORAL PATIENTS OR EVEN AS MORAL AGENTS

THIS MAY LEAD TO THE IDEA OF ARTIFICIAL LIFE

1. risk of prioritizing artificial agents over human beings
2. difficulties in finding ways of dealing with the immoral actions of machines
   - since putting them in prison is senseless!

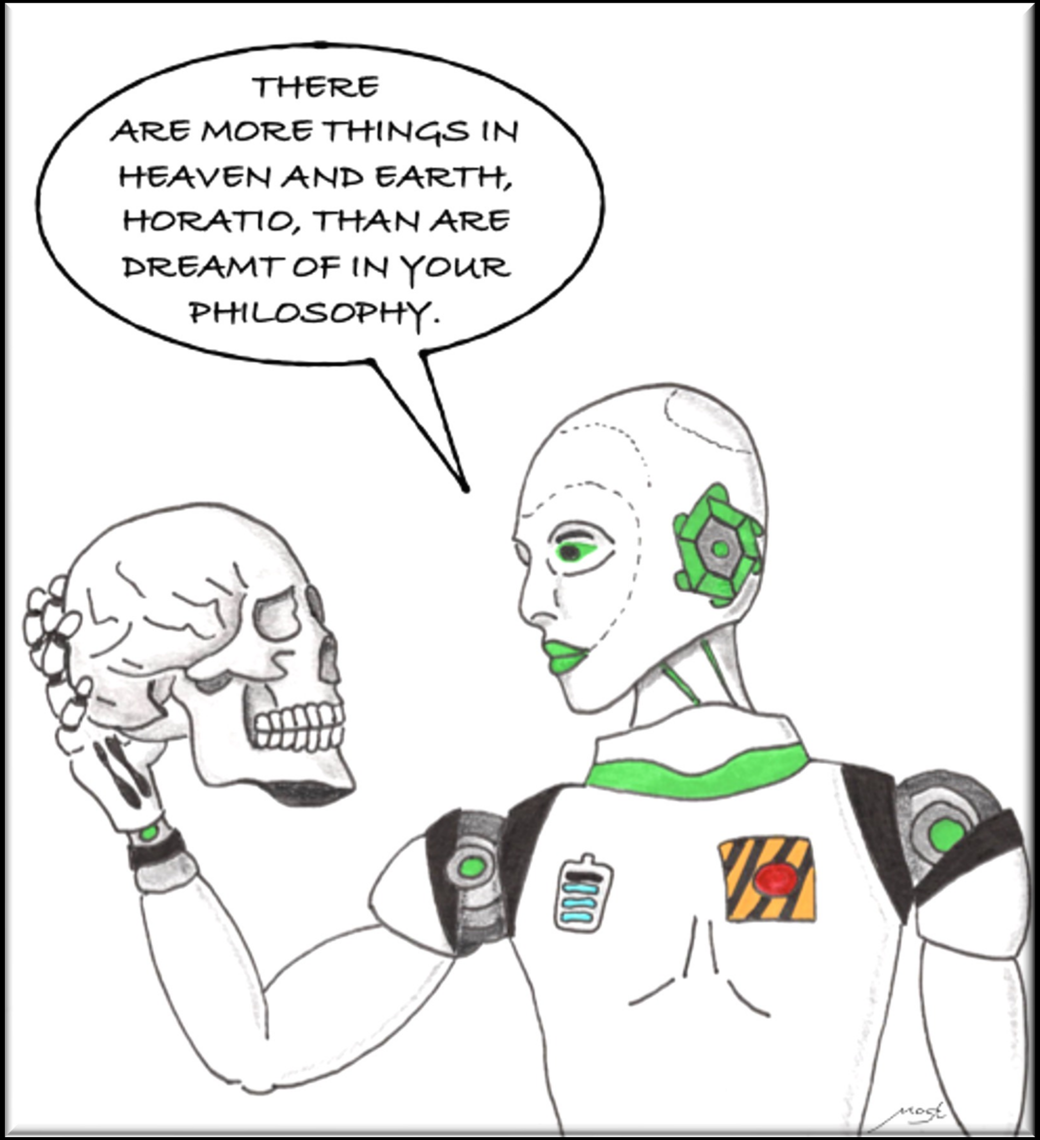
less radical position

- risk of over-attributing moral agency and patiency



(Gunkel, 2020; Strasser, forthcoming)

The Terra Incognita – the INBETWEEN


THERE ARE MORE THINGS IN HEAVEN AND EARTH, HORATIO, THAN ARE DREAMT OF IN YOUR PHILOSOPHY.
most

(Shakespeare, 2021; Strasser & Strasser, 2024)

Finding our way through the jungle

TOOL KIT 'MINIMAL APPROACHES'

How to conceptualize phenomena in the field of developmental psychology & animal cognition that fall through the sophisticated conceptual net of philosophy

- ❖ questioning the necessity of far too demanding conditions
- ❖ considering multiple realizations of capacities that seemed to be restricted to sophisticated adult humans

Stephen Butterfill & Ian Apperly (2013): minimal mindreading | John Michael et al. (2016): minimal sense of Commitment | Elisabeth Pacherie (2013): shared intention lite
Anna Strasser (2006): minimal action

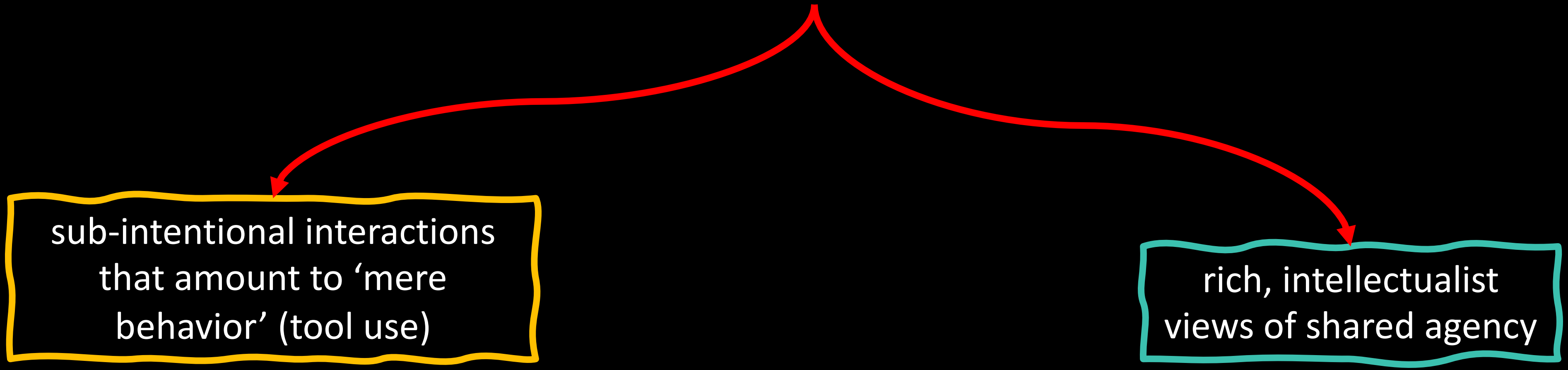
The way through the jungle

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE

instrumental view
artificial agents cannot be participants in joint actions

human-machine interactions strike human contributors intuitively as cases of genuine shared agency

→ **MID-WAY POINT BETWEEN**

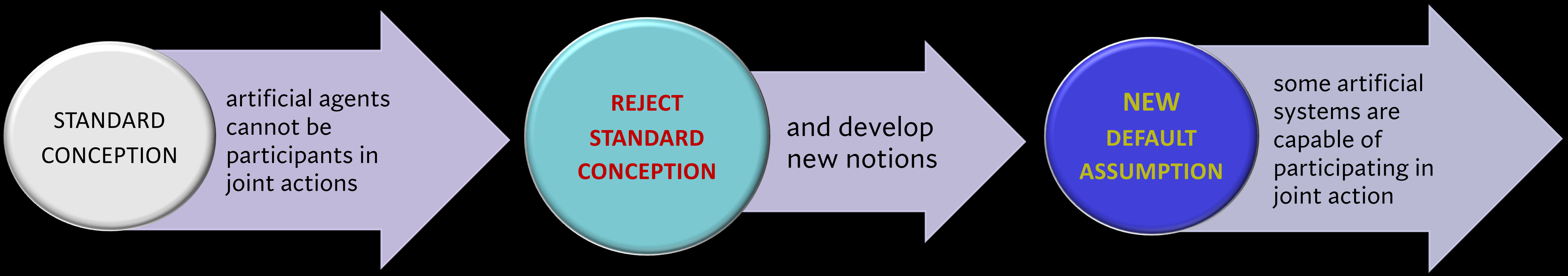


sub-intentional interactions that amount to 'mere behavior' (tool use)

rich, intellectualist views of shared agency

The way through the jungle

QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE



STANDARD CONCEPTION

artificial agents cannot be participants in joint actions

REJECT STANDARD CONCEPTION

and develop new notions

NEW DEFAULT ASSUMPTION

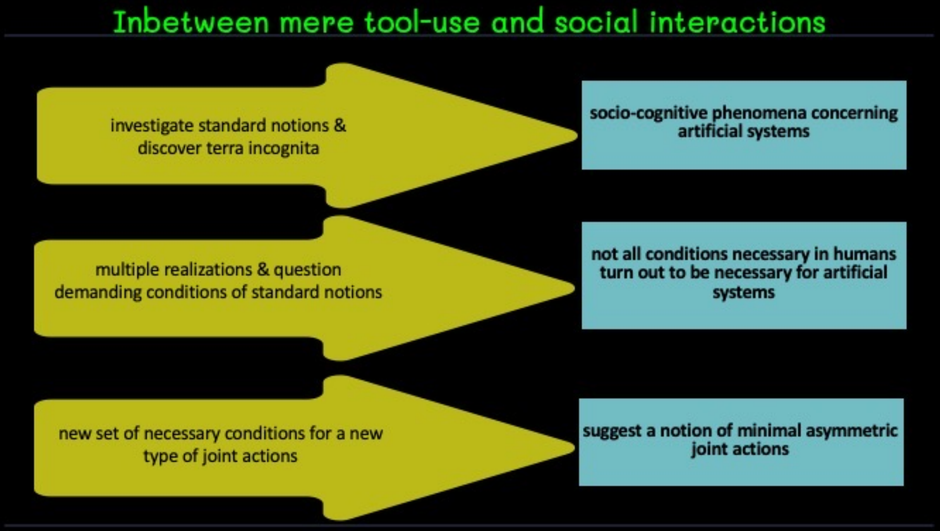
some artificial systems are capable of participating in joint action

PREVIEW

(1) AGENCY → MINIMAL AGENCY | (2) JOINT ACTION → ASYMMETRIC JOINT ACTION

1. apparent intentional behaviors of agents that do not satisfy the rich intellectualist demands of a Davidson-style theory, but still act

2. presuppositions for joint agency can be achieved with cognitive resources that are contentful and representational, but do not include the claim that both agents have to be living agents with consciousness & sentience



Inbetween mere tool-use and social interactions

investigate standard notions & discover terra incognita

socio-cognitive phenomena concerning artificial systems

multiple realizations & question demanding conditions of standard notions

not all conditions necessary in humans turn out to be necessary for artificial systems

new set of necessary conditions for a new type of joint actions

suggest a notion of minimal asymmetric joint actions

Joint action everywhere

Investigating standard notions

TOWARDS ASYMMETRIC JOINT ACTIONS

shared intentions & goals

specific belief state

relation of interdependence & mutual responsiveness

common knowledge

mastery of mental concepts

sophisticated mentalization skills

ing Service, New Haven, Conn.

ONLY HUMANS NEED APPLY

(Bratman, 2014)

Assuming multiple realization

TOWARDS ASYMMETRIC JOINT ACTIONS

NO NECESSITY OF AN EQUAL DISTRIBUTION OF ABILITIES AMONG ALL PARTICIPANTS

DEVELOPMENTAL PSYCHOLOGY

- joint action of adults and children
- children = socially interacting beings

ARTIFICIAL INTELLIGENCE

- joint action of human beings & artificial systems
- artificial systems =?= socially interacting entities

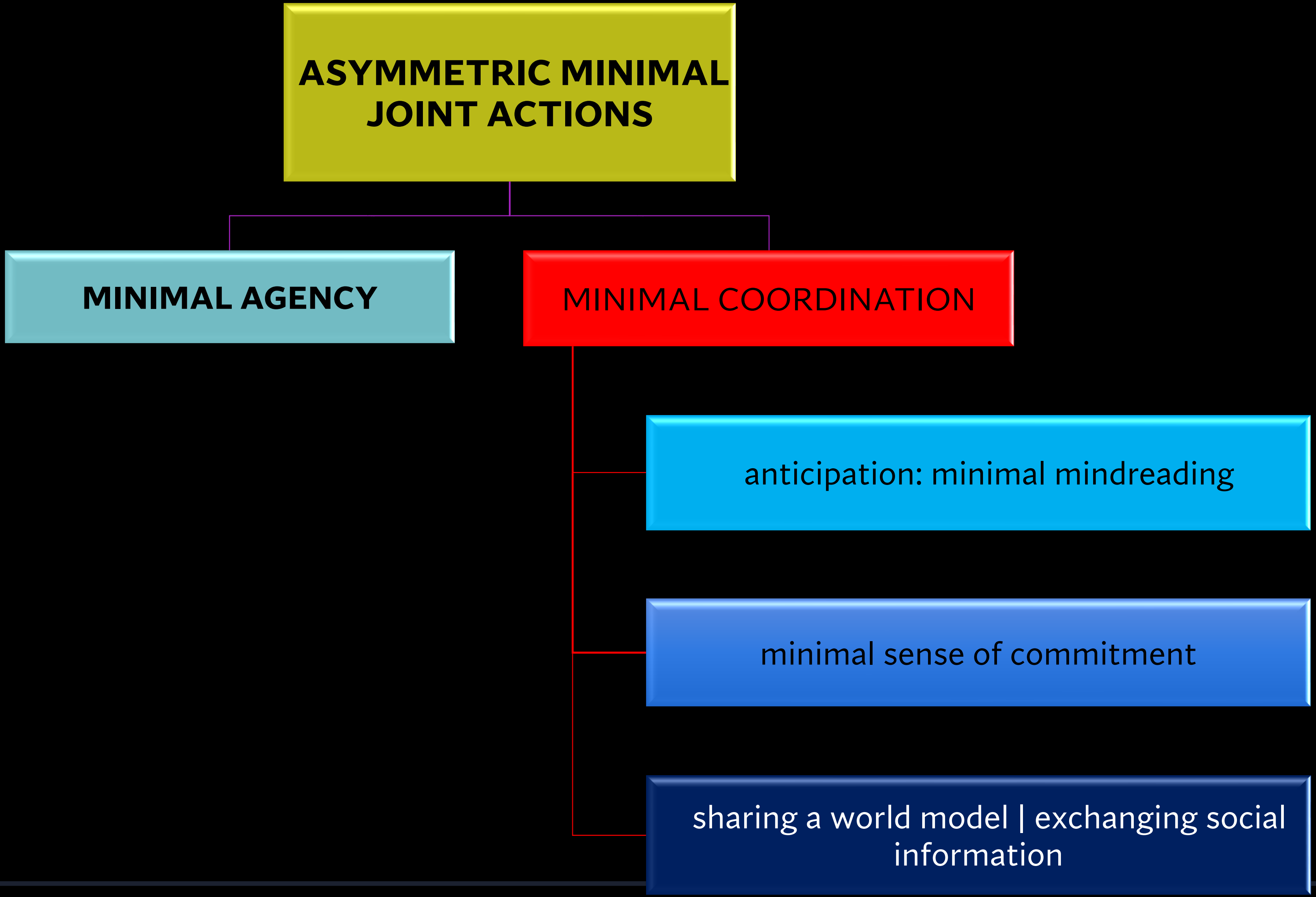
ADULT & CHILD

ROBOT & HUMAN
LLM & HUMAN

ASYMMETRIC JOINT ACTIONS

Inbetween mere tool-use and social interactions

TOWARDS ASYMMETRIC JOINT ACTIONS



ASYMMETRIC MINIMAL JOINT ACTIONS

- MINIMAL AGENCY
- MINIMAL COORDINATION
  - anticipation: minimal mindreading
  - minimal sense of commitment
  - sharing a world model | exchanging social information

Questioning intellectualist conceptions of agency

MINIMAL AGENCY

Donald Davidson

NECESSITY OF A COMPLEX SUITE OF CONCEPTUAL RESOURCES

- constitutive relations holding between propositional attitudes and their contents, as well as further conditions regarding language, intentional action, and interpretation, sharply separate off 'the beasts' from rational animals such as humans

The intrinsically holistic character of the propositional attitudes ***makes the distinction*** *between having any and having none* ***dramatic****!*

DON'T

BUT

BUT there are counterexamples

Empirical-based

DEVELOPMENTAL & COMPARATIVE PSYCHOLOGY

- **Multiple realization** of socio-cognitive abilities in infants & non-human animals

Premack & Woodruff 1978, Heyes 2014/2015, Vesper et al. 2010, Warneken et al. 2006

→ not only conceptually sophisticated humans can act

Conceptual-based

ONTOGENETICS & PHYLOGENETICS

- Shift from non-intentional to intentional is **gradual & partly learnable**
- Ontogenetic case
  Perner, 1991; Tomasello, 2008
- Phylogenetic case
  Sterelny, 2014; Henrich, 2016

→ Davidsonian 'all-or-nothing' dramatic divide is implausible

(Strasser & Wilby, 2023; Wilby & Strasser, 2024)

Questioning biological conceptions of intentional agency

TOWARDS ASYMMETRIC JOINT ACTIONS

CLAIMS

Any kind of agency that enables entities to be participants in a joint action requires

- internal affective states (emotional, mental, and conscious states)
- biological make-up is necessary to have genuine intentional and conscious thoughts

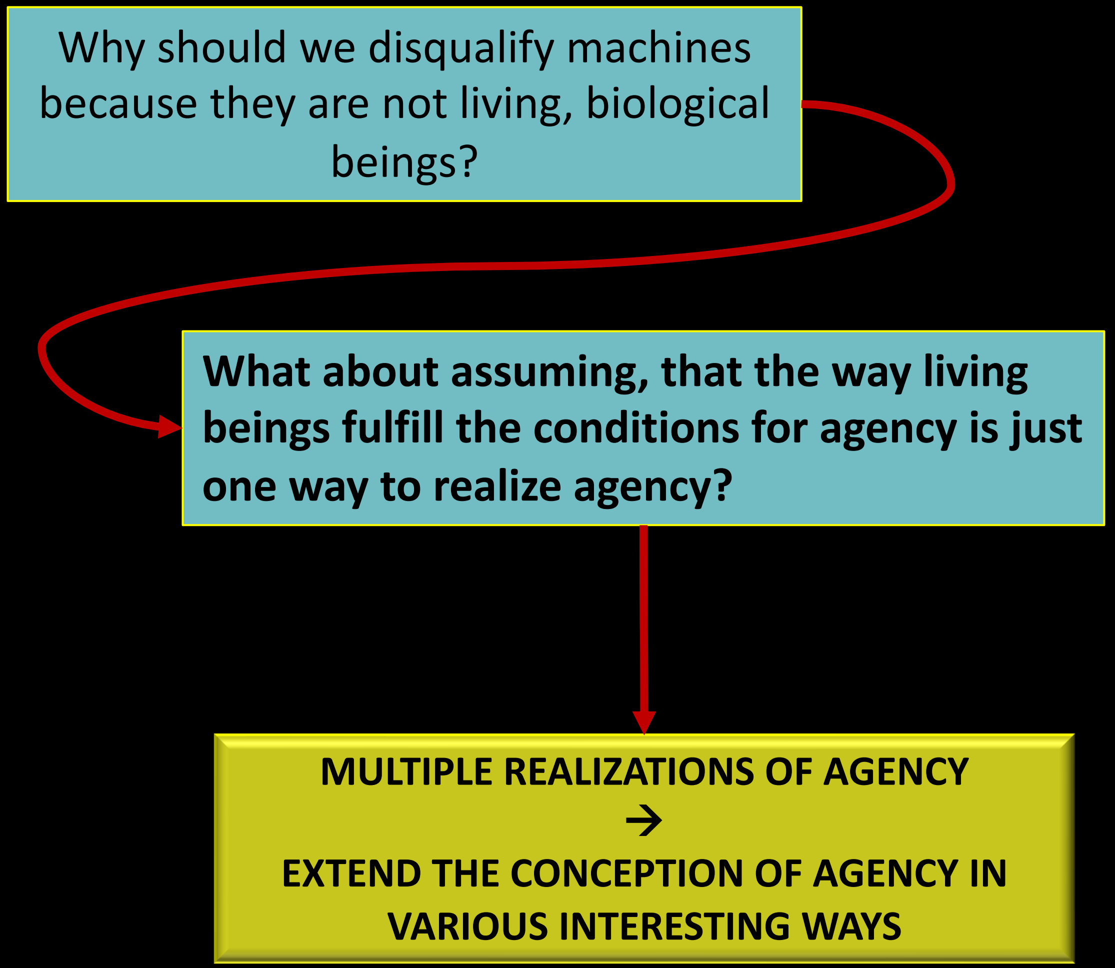
ARTIFICIAL SYSTEMS CANNOT QUALIFY AS SOCIAL INTERACTION PARTNERS

BECAUSE THEY LACK THE BIOLOGICAL MAKE-UP THEY CAN ONLY BEHAVE – NOT ACT

→ **EVERY HUMAN-MACHINE INTERACTION SHOULD BE UNDERSTOOD AS MERE TOOL-USE**

DON'T

BUT



Why should we disqualify machines because they are not living, biological beings?

What about assuming, that the way living beings fulfill the conditions for agency is just one way to realize agency?

MULTIPLE REALIZATIONS OF AGENCY
→
EXTEND THE CONCEPTION OF AGENCY IN VARIOUS INTERESTING WAYS

Neither intellectualist nor biological conceptions are wholly convincing

The Intellectualist Approach

attempts to draw a sharp distinction (a “dramatic divide”) between those who are capable of genuine thought and those who aren’t

- ➢ have a difficulty explaining how one goes from one side of the divide to the other
- ❖ developmental & comparative psychology suggest that the change is ***gradual and not sharp***

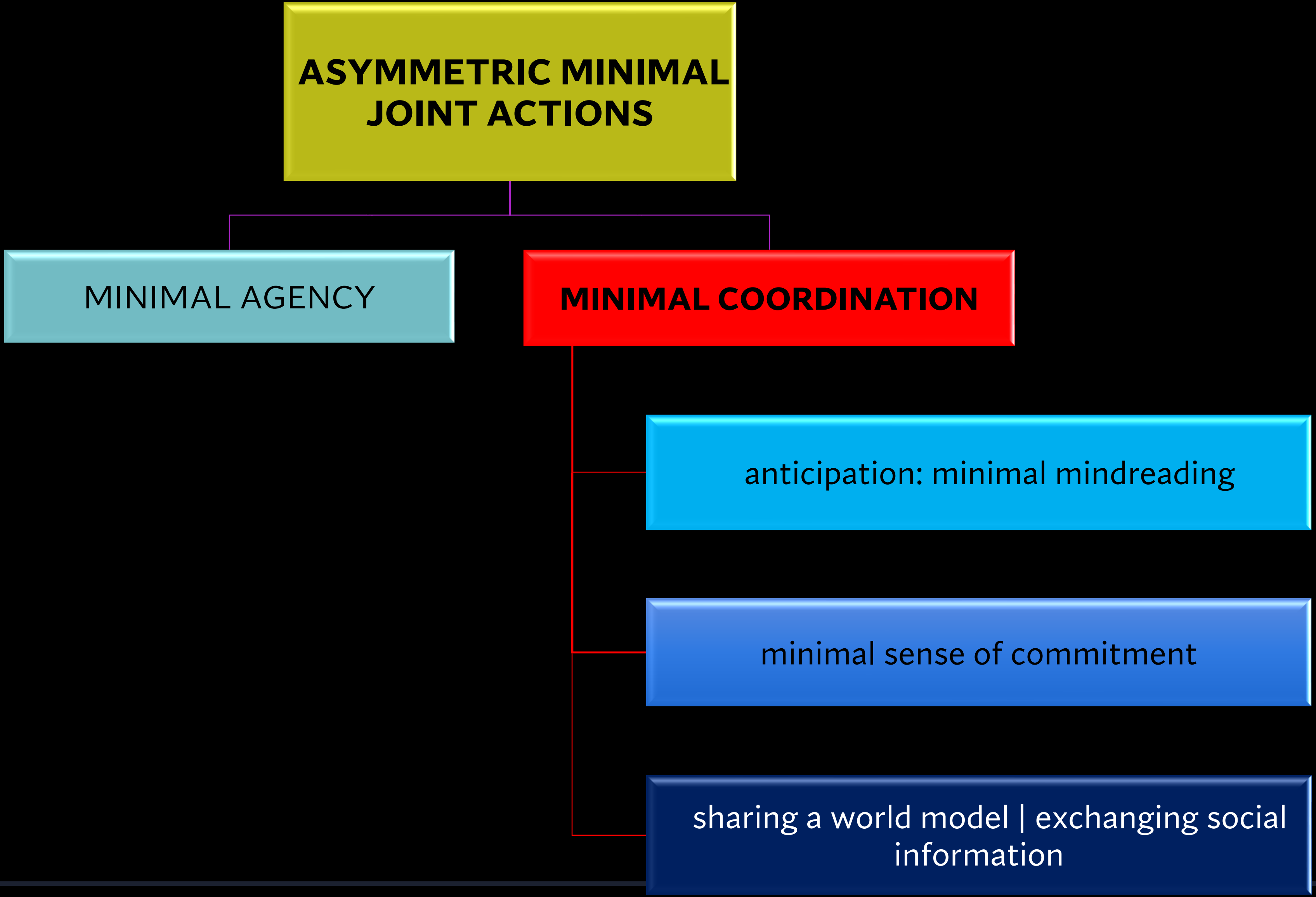
The Biological Approach

- attempts to draw the distinction due to a mysterious capacity of our brain to generate consciousness, feeling, subjectivity, and meaning

- ➢ fail to explain what the missing quality is, how we can know when it is there and when it is missing
- ❖ why we should suppose that it can only be realized in **electro-chemical brain reactions**, and not in silicon systems, or **neural nets**

Inbetween mere tool-use and social interactions

MINIMAL COORDINATION



ASYMMETRIC MINIMAL JOINT ACTIONS

MINIMAL AGENCY

MINIMAL COORDINATION

anticipation: minimal mindreading

minimal sense of commitment

sharing a world model | exchanging social information

MINIMAL MINDREADING

- to coordinate your contribution in a joint action one has to be able to anticipate what the other agent will do next

- utilize the notion of minimal mindreading that Steve Butterfill & Ian Apperley developed

- notion is a suitable starting point
  - as they claim that underlying processing are implicit, nonverbal, automatic, and based on unconscious reasoning

(Butterfill & Apperly, 2013)

MINIMAL SENSE OF COMMITMENT

- ❖ 'social glue' for much of what counts as social interactions

❖ **coordination abilities are also based on the capacity to form expectations and motivations with respect to your counterpart**

- utilize the notion of a minimal sense of commitment that illuminates minimal forms of interpersonal commitments
  - components (expectation or motivation) of a standard commitment can be disassociated
    - single occurrence of just one component can be treated as a sufficient condition

➢ asymmetric joint actions:

- minimal sense of commitment can be realized by just one participant
- most minimal case: only human counterparts entertain a minimal sense of commitment

(Michael et al., 2016)

Conclusion

CONSIDER THE POSSIBILITY OF CHANGING THE WINNING TEAM & QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE ENTITIES

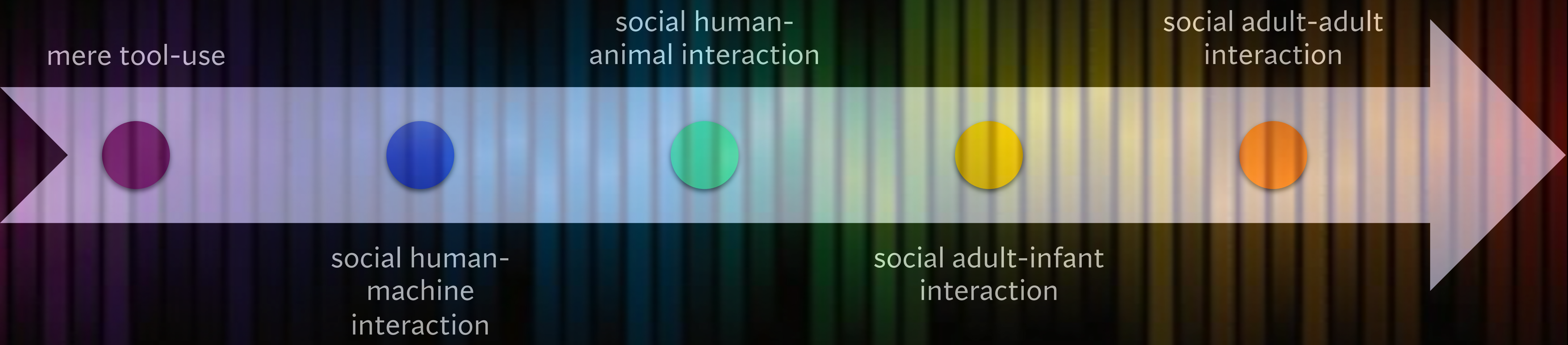
After all, we might be confronted with a new game.

- Before we can answer the question of **what we are doing when we interact with LLMs**, we have to conceptualize the INBETWEEN, because we cannot reduce our interactions with LLMs (and especially with future products of generative AI) to mere tool use.

THE MAIN AIM OF THIS TALK WAS TO PREPARE THE GROUNDS FOR QUESTIONING THE DICHOTOMY BETWEEN ANIMATE AND INANIMATE ENTITIES, AS THIS IS AN IMPORTANT PRESUPPOSITION FOR ANY DEVELOPMENT OF NEW NOTIONS THAT CAN CAPTURE PHENOMENA THAT I LOCATE IN THE INBETWEEN.

IF WE ARE SUCCESSFUL WITH THIS, WE CAN ARGUE FOR A GRADUAL APPROACH.

A gradual approach



mere tool-use

social human-machine interaction

social human-animal interaction

social adult-infant interaction

social adult-adult interaction

All this would not have been possible if I had not interacted with people & machines

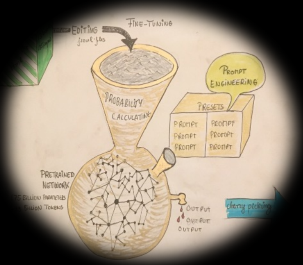
Daniel Dennett

Eric Schwitzgebel

Mathew Crosby

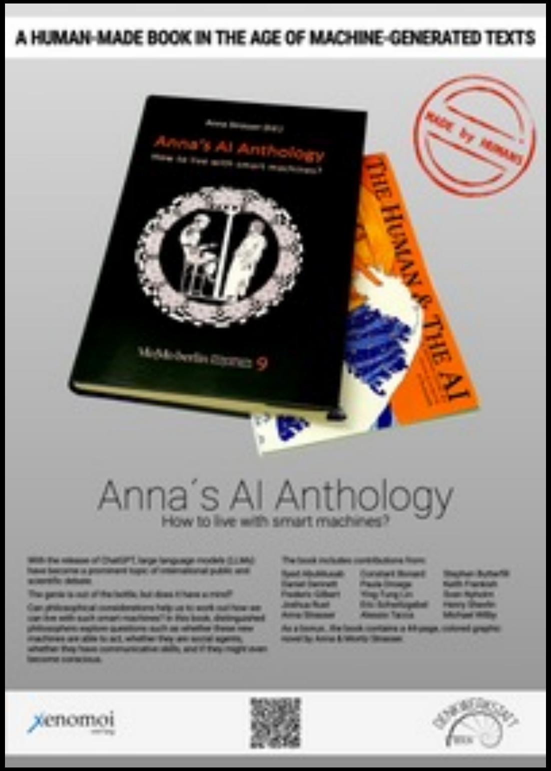
David Schwitzgebel

Mike Wilby



DigiDan

Thank you !



A HUMAN-MADE BOOK IN THE AGE OF MACHINE-GENERATED TEXTS

Anna Strasser (ed.)
Anna's AI Anthology
How to live with smart machines?
MADE by HUMANS
THE HUMAN & THE AI

Anna´s AI Anthology

How to live with smart machines?

With the release of ChatGPT, large language models (LLMs) have become a prominent topic of international public and scientific debate.

The genie is out of the bottle, but does it have a mind?

Can philosophical considerations help us to work out how we can live with such smart machines? In this book, distinguished philosophers explore questions such as whether these new machines are able to act, whether they are social agents, whether they have communicative skills, and if they might even become conscious.

The book includes contributions from:

Syed AbuMusab
Daniel Dennett
Frederic Gilbert
Joshua Rust
Anna Strasser
Constant Bonard
Paula Droege
Ying-Tung Lin
Eric Schwitzgebel
Alessio Tacca
Stephen Butterfill
Keith Frankish
Sven Nyholm
Henry Shevlin
Michael Wilby

As a bonus, the book contains a 44-page, colored graphic novel by Anna & Moritz Strasser.

xenomoi

References

Agrawal, A., Mackey, L., & Kalai, A. T. (2023). *Do Language Models Know When They're Hallucinating References?* (arXiv:2305.18248). arXiv. http://arxiv.org/abs/2305.18248

Agüera y Arcas, B. (2022). Do Large Language Models Understand Us? *Daedalus*, *151*(2), 183–197. https://doi.org/10.1162/daed_a_01909

Barkham, P. (2021, July 25). Should rivers have the same rights as people? *The Guardian*. https://www.theguardian.com/environment/2021/jul/25/rivers-around-the-world-rivers-are-gaining-the-same-legal-rights-as-people

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bunten, A., Iorns, C., Townsend, J., & Borrows, L. (2021, June 3). *Rights for nature: How granting a river 'personhood' could help protect it*. The Conversation. http://theconversation.com/rights-for-nature-how-granting-a-river-personhood-could-help-protect-it-157117

Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Cole, S. (2023). 'It's Hurting Like Hell': AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection. *Vice*. https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates

Dooley, B., & Ueno, H. (2022, April 24). This Man Married a Fictional Character. He'd Like You to Hear Him Out. *The New York Times.* https://www.nytimes.com/2022/04/24/business/akihiko-kondo-fictional-character-relationships.html

Gunkel, D. J. (2020). Robot Rights – Thinking the Unthinkable. In *Smart Technologies and Fundamental Rights* (pp. 48–72). Brill. https://doi.org/10.1163/9789004437876_004

------------- (2023). *Person, Thing, Robot: A Moral and Legal Ontology for the 21st Century and Beyond*. The MIT Press. https://doi.org/10.7551/mitpress/14983.001.0001

Henrich, J. P. (2016). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.

Heyes, C. (2014). False belief in infancy: A fresh look. *Developmental Science*, *17*(5), 647–659. https://doi.org/10.1111/desc.12148

----------- (2015). Animal mindreading: What's the problem? *Psychonomic Bulletin & Review*, *22*(2), 313–327. https://doi.org/10.3758/s13423-014-0704-4

Jensen, C. B., & Blok, A. (2013). Techno-animism in Japan: Shinto Cosmograms, Actor-network Theory, and the Enabling Powers of Non-human Agencies. *Theory, Culture & Society*, *30*(2), 84–115. https://doi.org/10.1177/0263276412456564

Lake, B. M., & Baroni, M. (2023). Human-like Systematic Generalization through a Meta-learning Neural Network. *Nature*, 1–7. https://doi.org/10.1038/s41586-023-06668-3

Lemoine, B. (2022, June 11). Is LaMDA Sentient? — An Interview. *Medium*. https://cajundiscordian.medium.com/is-lamda-sentient-an-interview-ea64d916d917

Marcus, G., & Davis, E. (2020). GPT-3, Bloviator: OpenAI's language generator has no idea what it's talking about. *MIT Technology Review*. https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion

Michael, J., Sebanz, N., & Knoblich, G. (2016). The Sense of Commitment: A Minimal Approach. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01968

O'Donnell, E., & Talbot-Jones, J. (2017, March 23). Three rivers are now legally people – but that's just the start of looking after them. *The Conversation*. http://theconversation.com/three-rivers-are-now-legally-people-but-thats-just-the-start-of-looking-after-them-74983

Pacherie, E. (2013). Intentional joint agency: Shared intention lite. *Synthese*, *190*(10), 1817–1839. https://doi.org/10.1007/s11229-013-0263-7

References

Perner, J. (1991). *Understanding the representational mind* (pp. xiv, 348). The MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Robertson, J. (2014). Human Rights vs.Robot Rights: Forecasts from Japan. *Critical Asian Studies*, *46*(4), 571–598. https://doi.org/10.1080/14672715.2014.960707

------------- (2017). *Robo sapiens japanicus: Robots, Gender, Family, and the Japanese Nation*.

Salmón, E. (2000). Kincentric Ecology: Indigenous Perceptions of the Human-Nature Relationship. *Ecological Applications*, *10*(5), 1327–1332. https://doi.org/10.2307/2641288

Scarborough, J. K., & Bailenson, J. N. (2014). Avatar Psychology. In M. Grimshaw (Ed.), *The Oxford Handbook of Virtuality*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199826162.013.033

Sterelny, K. (2012). *The Evolved Apprentice: How Evolution Made Humans Unique*. The MIT Press. https://doi.org/10.7551/mitpress/9780262016797.001.0001

Strasser, A. (2006). *Kognition künstlicher Systeme:* DE GRUYTER. https://doi.org/10.1515/9783110321104

------------- (2013). Kognition künstlicher Systeme. In *Kognition künstlicher Systeme*. De Gruyter. https://doi.org/10.1515/9783110321104

------ (Ed.). (2024). *Anna's AI Anthology. How to live with smart machines?* xenomoi Verlag.

Strasser, A., & Schwitzgebel, E. (2024). Quasi-sociality: Toward Asymmetric Joint Actions. In *Anna's AI Anthology. How to live with smart machines?* xenomoi Verlag.

Strasser, A., & Wilby, M. (2023). The AI-Stance: Crossing the Terra Incognita of Human-Machine Interactions? In *Social Robots in Social Institutions* (pp. 286–295). IOS Press. https://doi.org/10.3233/FAIA220628

Tomasello, M. (2008). *Origins of human communication* (pp. xiii, 393). MIT Press.

Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do Large Language Models Know What Humans Know? *Cognitive Science*, *47*(7), e13309. https://doi.org/10.1111/cogs.13309

Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, *23*(8), 998–1003. https://doi.org/10.1016/j.neunet.2010.06.002

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative Activities in Young Children and Chimpanzees. *Child Development*, *77*(3), 640–663. https://doi.org/10.1111/j.1467-8624.2006.00895.x

Weil, E. (2023, March 1). *You Are Not a Parrot*. New York Magazine. https://nymag.com/intelligencer/article/ai-artificial-intelligence-chatbots-emily-m-bender.html

Wilby, M., & Strasser, A. (2024). Situating machines within normative practices: Bridging responsibility gaps with the AI-Stance. In A. Strasser (Ed.), *Anna's AI Anthology. How to live with smart machines?* xenomoi Verlag.