QUASI-SOCIALITY: TOWARD ASYMMETRIC JOINT ACTIONS

ANNA STRASSER (LMU, Munich, DenkWerkstatt Berlin, Germany)
& ERIC SCHWITZGEBEL (UC Riverside, US)

HUMANS AND SMART MACHINES AS PARTNERS IN THOUGHT

WHAT ARE WE DOING 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 them

full-blown sociality

- social interactions

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

OVERVIEW

slides can be downloaded at
https://www.denkwerkstatt.berlin/ANNA-STRASSER/TALKS/

1
Purely fictional social attributions

mere tools
like Roomba, Pleo

2
Two ends of a spectrum

single-sided sociality
versus
full-blown social interaction

3
Between these two extremes

Asymmetric joint actions

4
Two Implications

PURELY FICTIONAL SOCIAL ATTRIBUTIONS TO MACHINES

"as if"

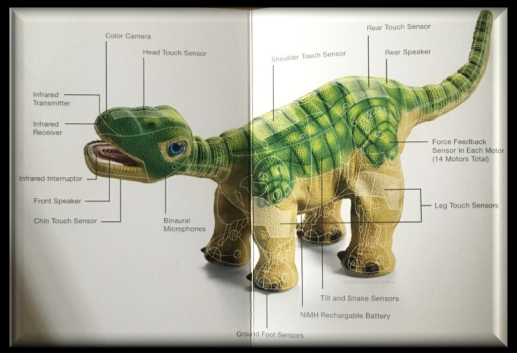
INDEPENDENT OF ANY SOCIAL OR QUASI-SOCIAL FEATURES IN THE PARTNER

Chris, a former student of Eric's, was in the habit of apologizing to his Roomba. The simple cleaning device would wander underfoot and bump him, or he'd accidentally kick it. "Oh, sorry, little guy!" Chris would say, then gently nudge it on its way.

ROOMBA	CHRIS
• **does not respond to** an apology or a polite redirection • nothing reacts to Chris's speech • *would respond in exactly the same way to mechanical redirection*	• **employs social skills** • treating the Roomba *as if* it were a social partner • is being social • knows that his sociality is cast into the void

Sociality is entirely one-sided	**in no respect a social partner**	**sociality is directed to an object**

- If the Roomba were a dog or a child, lifting and redirection would be a kind of social communication, presumably taken up in one way or another by the dog or child as a social interaction; but social redirection gains no such traction with the Roomba.



PURELY FICTIONAL SOCIAL ATTRIBUTIONS TO MACHINES

"as if"

INDEPENDENT OF ANY SOCIAL OR QUASI-SOCIAL FEATURES IN THE PARTNER

Kate Darling reported that even her team members (who definitely knew – how the robot was constructed) were, after a phase of interaction, reluctant to behave destructively towards this toy robot called Pleo.

PLEO	TEAM MEMBERS
no social uptake • triggering nurturing behavior	**employ social skills** • make a moral difference

BUT

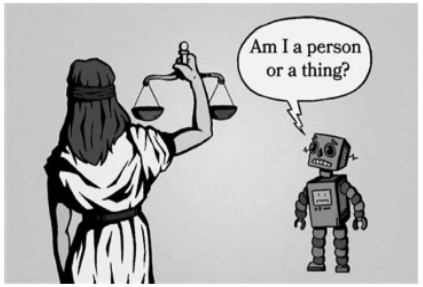
IN-BETWEEN PHENOMENA

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

describing artificial systems as social interaction partners poses a

CONCEPTUAL PROBLEM FOR PHILOSOPHY

- because conditions for full-fledged social agency are primarily tailored to sophisticated adult human beings



A dichotomous distinction between action & mere behavior leads to a terra incognita in which we find phenomena for which we have no established notions yet.

ABILITIES OF CHILDREN, NON-HUMAN ANIMALS, AND ARTIFICIAL SYSTEMS FALL THROUGH THE CONCEPTUAL NET

TWO ENDS OF A SPECTRUM

Anna's chatbot

SINGLE-SIDED SOCIALITY

- sociality tossed into a void
- application of social skills
- reactions toward entities who are in no respect social partners, with no capacity for social uptake

Wilson

asocial

FULL-BLOWN, INTELLECTUALLY DEMANDING, COOPERATIVE SOCIAL INTERACTION

as described by Davidson, Gilbert, Bratman

- both partners make second-order mental state attributions and satisfy various other conditions are required for full-blown adult human cooperative action

mindreading

general idea

- each needs to know what the other is thinking & needs to know that the other knows that they know

→ both partners engage in at least second-order mental state attribution:
having beliefs about your partner's beliefs about your beliefs

fully mutual joint action

DAVIDSON – ACTIONS

THE NECESSITY OF A COMPLEX SUITE OF CONCEPTUAL RESOURCES

Donald Davidson (1963, 1971, 1980, 1982, 1984, 2001)

- constitutive relations holding between propositional attitudes & their contents
- language
- intentional agency
- interpretation

} sharply separate off 'the beasts' from rational animals such as humans

FULL-BLOWN INTENTIONAL AGENCY
requires intentional action
to be carried out by an entity **with** an integrated, holistic set of propositional attitudes

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

BRATMAN – JOINT ACTIONS

shared intentions & goals

specific belief state

relation of interdependence & mutual responsiveness

common knowledge

mastery of mental concepts

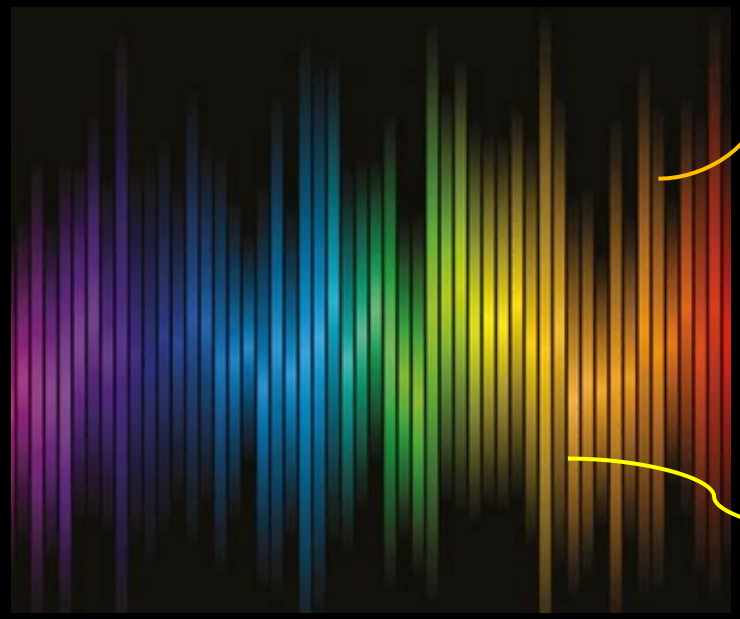
sophisticated mentalization skills

[Bratman 2014]

WHAT ABOUT INFANTS & NON-HUMAN ANIMALS?

DEVELOPMENTAL PSYCHOLOGISTS

- second-order belief attribution = late-emerging ability → well beyond the capacity not only of Roombas but also of three-year-olds



And yet you can play peek-a-boo with a three-year-old. Isn't that a social activity? And you can argue about bedtime. And you can take turns on a tricycle. …

→ three-year-olds' capacities for mental state attribution are more sophisticated than mainstream developmental psychologists think

→ you can also engage in social or quasi-social interactions with infants & cats

Parent and baby can gaze into each other's eyes and take turns making nonsense sounds.

You can snuggle up with your cat – and if your cat scratches you, you can slap it, in a way that communicates something, gaining uptake by the cat, hopefully, of the sort that it would be pointless to hope for in a Roomba.

BETWEEN THESE TWO EXTREMES

A SPECTRUM OF ASYMMETRIC JOINT ACTION

[junior partner]
is lifted or scaffolded into complex joint action by the engagement & structuring of the more knowledgeable partner

[senior partner]
- knows that they know what the other knows
- fully appreciates the social structure of the interaction they are having

ASYMMETRIC SOCIALITY

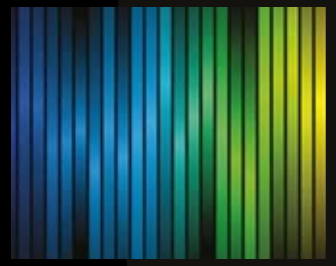
ONLY *QUASI-SOCIAL*
- closer to the Roomba end of the spectrum
- *premature infants might respond to a soothing touch or sound*
  ← without being ready for anything like joint action
- *letting a pet snake climb on you might be only quasi-social*
  ← pet snake might only in some minimal sense recognize that you are another entity with which it is interacting

WORTH CALLING PROPERLY SOCIAL, EVEN IF THEY ARE ASYMMETRIC
- *the argument about bedtime*
  ← child brings a lot of social understanding, even if the parent brings more
- *snuggling with a cat*

INTERACTIONS WITH LLMS

& OTHER RECENT & SOON-TO-EMERGE AI SYSTEMS

HAVE BEGUN TO MOVE ALONG THE SPECTRUM OF QUASI-SOCIALITY



Unlike the Roomba LLMs can be designed to respond to the social dimensions of our interactions with it

- interact aggressively with ChatGPT, by expressing anger and dissatisfaction → it emits outputs that are naturally interpreted as apologies & attempts to make amends
- If you express appreciation → it says thank you

INSIDER Newsletters

DISCOURSE | TECH

The new Bing is acting all weird and creepy — but the human response is way scarier

Adam Rogers

Bing's A.I. Chat Reveals Its Feelings: 'I Want to Be Alive. 😈

In a well-known conversation between Microsoft's Bing/Sydney language model and a New York Times reporter,

- Bing/Sydney appeared to express romantic interest in the interviewer, to pick up conversational threads, to accuse the interviewer of being pushy and manipulative, and seemingly it tried to seduce him.
- The interviewer deployed social skills in interacting with it, and the language model's responses invited interpretation as social reactions, precipitating new social reactions by the reporter.

INTERACTIONS WITH LLMS
& OTHER RECENT & SOON-TO-EMERGE AI SYSTEMS

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

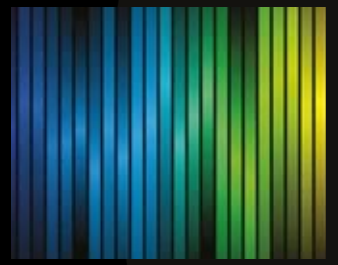
REPLIKA

advertised as "the world's best AI friend"
specifically designed to encourage social engagement,
customizing its interactions with users over time

You might say this is no different in kind from what is going on with the Roomba, only more complex. After all, this is only a machine responding to its programming, not a real locus of consciousness and feelings.

We agree that Replika can't really be social.

DIFFERENT IN KIND



BUT

WE INSIST THAT THE INTERACTION *IS* DIFFERENT IN KIND

case Roomba:

- social reaction to the Roomba is being tossed into the void, influencing nothing …

But with LLMs

- apologies & social reactions are not being tossed into the void, they influence the machine's responses, and they do so in ways that make social sense
- Anger leads to apology. Question leads to answer. Hints of sexual interest are picked up on and amplified back.

→ You can productively take a social stance toward the machine
→ You can call on your social skills in interacting with it
→ you can coax the machine into further socially interpretable interactions

DIFFERENT IN KIND

THE REAL SOCIAL SKILLS & KNOWLEDGE ARE COMING FROM YOU.

- We're not yet ready to say that large language models have social skills and social knowledge in the same robust sense that human beings.
- But the machine is designed, or at least has emerged from its developmental process, in a way that exploits the fact that you will react to it as a social agent; and you, in turn, can exploit that fact about it.

ASYMMETRIC QUASI-SOCIAL INTERACTIONS

interactions between a fully social agent and some partner – whether human, machine, or animal – that is not cognitively capable of full-fledged social joint action but that does respond in a way that productively invites further social responses from the social partner

sort of social exchange

→ drawing upon the social skills of the senior partner even if the junior partner lacks anything like real social understanding

→ quasi-sociality can place relatively little cognitive demand on the junior partner

JUNIOR PARTNER

needn't
- understand that the other is an agent or has beliefs, desires, or goals
- intend to communicate or cooperate
- even be a conscious entity

must be
- more than a Roomba
- structured in such a way as to draw social behavior from the senior partner
- reacting to the senior partner's social behavior in a way that solicits further social behavior
- able to do so in a manner that importantly *resembles* social interactions as they transpire between two fully-fledged social partners

RESEMBLANCE – A MATTER OF DEGREE

MULTIDIMENSIONALITY OF BEING A MATTER OF DEGREE

social interchange is complex → multiple relevant dimensions of resemblance

THIS IS EXACTLY WHAT WE SHOULD EXPECT, GIVEN A BIRD'S EYE VIEW OF THE PHENOMENON, WHICH DOESN'T MYOPICALLY FOCUS ON ADULT HUMANS AS THE ONLY TYPES OF SOCIAL PARTNERS.

Complex social skills will not emerge in an instant

- not developmentally in humans
- nor phylogenetically in animal evolution
- nor technologically in the design of AI systems

WE SHOULD EXPECT A WIDE RANGE OF QUASI-SOCIALITY

asocial Roombas and viruses

fully social, explicitly cooperative, second-order attitude ascribing adult humans

A LAST EXAMPLE: KIWIBOTS

simple delivery bots

- roll along sidewalks & onto campus to deliver small food orders
- mostly autonomous, but require some remote human intervention, e.g., when crossing intersections
- have digitally animated eyes

Occasionally, they wander off path or get stuck somewhere. Evidently, when this happens, passersby sometimes help the Kiwibots out. Maybe their cute and non-threatening appearance makes this more likely.

IMAGINE AN UPDATE OF THE TECHNOLOGY TOWARD QUASI-SOCIALITY

A QUASI-SOCIAL KIWIBOT

- if it gets stuck somewhere it emits some mild distressed noise – *"ooh, ooh"* – and says, *"Gosh, I'm stuck. Maybe someone will help me?"*
- it can detect whether it has been helped
- and whether a person has approached it, contacted it, and started it moving again

→ After this, maybe it says, *"Thank you so much for the help, friend!"*

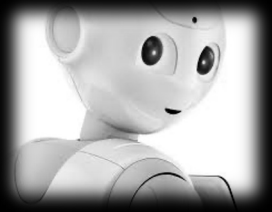
a small step along the path of quasi-sociality:

an interaction pattern designed to productively trigger social behavior in the senior partner

FUTURE VERSIONS WITH MORE SOPHISTICATED SOCIAL INTERACTIONS

- maybe people who order food can opt in to letting the Kiwibot display their name and face.
- If the bot is delivering to a crowded room, or if the bot is not promptly unloaded, perhaps it can approach a bystander, in a slow and seemingly timid way, and scan the bystander's face for a friendly or welcoming expression.
- If the bystander's expression isn't classified as welcoming, the bot can terminate the interaction and maybe approach someone else.
- Upon detecting a face classified as welcoming, the bot might emit "Could you help me find Devan?", displaying a picture of Devan's face. "I have a delivery for them!"

Asymmetric minimal joint actions



In this way, one could imagine a progression of ever more sophisticated delivery bots, that ever more effectively exploit the social capacities of senior partners.

Maybe at some point – who knows when? – they become genuinely conscious, genuinely capable of social emotion, and genuinely capable of knowing that you know that they know.

The quasi-sociality starts far before then.

- along the way, we expect a wide, wide area of first quasi-sociality and asymmetric sociality with the human as a senior partner

ASYMMETRIC MINIMAL JOINT ACTIONS

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

TWO IMPLICATIONS

1

A STRIKING FEATURE OF HUMAN ASYMMETRIC SOCIALITY AND ASYMMETRY JOINT ACTIONS

- parents provide scaffolding for the child's developing sociality
  - By treating the child as a social partner, the parent helps *make* the child a social partner.
    When we are slightly aspirational in our interpretation of our children, reading into their actions and reactions maybe a little more sophistication than is really there, this helps those children rise into the roles and attributes we imagine for them.
  - By trusting, we help make them trustworthy
  - By treating them as fair, moral, and sympathetic to others, we help make them more fair, moral, and sympathetic to others

This could potentially also be true for AI systems that are capable of learning from our interactions with them. Perhaps, for the right machines, if we treat them as social partners, this helps them develop the pattern of reactions that make them social partners.

TWO IMPLICATIONS

2

POTENTIAL FOR CORPORATE EXPLOITATION

- Our upgraded Kiwibot is innocent enough, but it is drawing upon the freely given goodwill of bystanders to achieve the corporate end of an efficient delivery.

- More problematically, if people really do fall in love with their Replika chatbots, then they will want to pay monthly fees to maintain the service, and they will pay extra for sexy pictures, and they will pay extra for stylish clothes and fancy features. → obvious potential for lonely people to be exploited
  Clever engineers of quasi-social AI systems could potentially become skilled at generating social reactions from users in a way that exploits human vulnerabilities for the sake of corporate interests.

- This could be especially the case if quasi-social AI systems are designed to generate real feelings of love and attachment.

We don't want people committing suicide when their chatbot rejects them.
We don't want someone diving out into traffic, risking their lives to save a Kiwibot from an oncoming truck.


end