LLMs and the Human Condition

Peter Wallis
pwallis@acm.org

Abstract

Theory based AI research has had a hard time recently and the aim here is to propose a model of what LLMs are actually doing when they impress us with their language skills. The model integrates three established theories of human decision-making from philosophy, sociology, and computer science. The paper starts with the collective understanding of reasoning from the early days of AI research - primarily because that model is how we humans think we think, and is the most accessible. It then describes what is commonly thought of as "reactive systems" which is the position taken by many philosophers and indeed many contemporary AI researchers. The third component to the proposed model is from sociology and, although not flattering to our modern ego, provides an explanation to a puzzle that for many years has occupied those of us working on conversational user interfaces.

In AI at one point researchers were classified as "neats" or "scruffies". The neats worked from theory, while the scruffy researchers took an extreme programming approach, making things that did something and then tinkering to improve it. Theory was generally based on a **reference model** of semantics in which symbols such as the word 'cat' has meaning by referring to a thing in the world. Deciding if that phenomena on the end of the bed is a member of the set of all cats might be problematic but, having done that, the expression "the cat" simply refers to something in the world, and it is a cat because that is in the set of things called cats. The symbol grounding problem (Harnad, 1990) is a problem, but it turns out to be the wrong problem. In contrast the scruffies generally didn't bother with theory and indeed Brooks in his seminal paper "Intelligence without Representation" (Brooks, 1991) claimed to be "just doing engineering" and disavowed any influence from Heidegger. The aim here is to look again at the theory, but perhaps if we understand what is happening when a transformer assembles a LLM from all the text in the world, we can suggest a better way forward for those 'just' engineering.

The paper starts with a recap of good old fashioned AI (GOFAI) and then describes what Brooks and many philosophers (and psychologists) have said. It then introduces a third mechanism that was touched on in the 1970s in AI research but which is an idea that is hundreds of years old in what is now sociology. Finally the model being developed is applied to a major issue in Natural Language Understanding (NLU) namely the phenomenon of apparent "mind reading".

1 The problem: people read minds

Working on NLU, one sooner or later comes across a phenomenon in which humans appear to do "mind reading". In what follows the focus is one example but the phenomenon itself is extremely common. The solution is often assumed to be a better (and more detailed) representation of the context and from that perspective, the proposed solution is to represent context, not as things, but as practices.

The example is from Mann (Mann, 1988) and is a transcript of a real conversation in a naturally occurring setting:

1	Child:	I'm hungry
2	Mother:	Did you do a good job of your geography homework?
3	Child:	Yeah.
4	Child:	What's to eat?
5	Mother:	Let me read it.
6	Mother:	What is the capital of Brazil?
7	Child:	Rio de Janeiro.
8	Mother:	Think about it.
9	Child:	It's Brasilia.
10	Child:	Can I eat now?
11	Mother:	I'll let you have something later.
12	Mother:	What is the capital of Venezuela?
13	Child:	Caracas.
14	Mother:	Fine.

15 Child: So what can I eat?
16 Mother: You want some cereal?
17 Child: Sure.

All English speakers *by definition* (Tomasello, 2008) will follow what is happening in this conversation and feel there is nothing extraordinary about it. The problem for a reference model of language understanding is that the mother says nothing related to the child's utterances until line 11. There is no overlap of semantic content - none at all. Of course, as English speakers, we can all say what both the mother and the child *want*.

This has historically lead researchers interested in conversational agents off into the realm of recognition of intent and modelling theory-of-mind (ToM) (Allen et al., 1995). Classically this would be modelled in terms of the beliefs, desire and intentions of an agent, the theory being that a rational agent will do what it believes is in its interests (Cohen and Levesque, 1990), all in terms of good old fashioned symbolic representations.

2 Machines

The reference model of meaning is generally how people see language as working, and the early AI researchers saw the problem as one of tidying the hap-hazard oddities found in natural languages. For centuries there has been the idea that meaning bottoms out at some kind of semantic primitives (Ishiguro, 1990) and in this tradition Roger Schank proposed that there was a workable set of primitive verbs or 'acts' from which the meaning of a sentence could be based (Schank, 1972). Examples are PTRANS for "physical transfer" and MTRANS for mental or knowledge transfer, and that machines could translate natural descriptions into (representations of) meaning and figure out consequences from there. The primitives approach to AI didn't really work but there was also the idea that a formal system of symbol manipulations might map isomorphically onto events in the world. In "Godel, Escher, Bach", Hofstader (Hofstader, 1979) gives a beautiful example with his pq- system which is a set of syntactic transforms on strings of 'p', 'q', and '-'. The system produces an infinite set of strings including "`-p--q---`" and "`--p--q----`" but not "`-p---q--`". The number of '-'s in produced strings just happens to map onto the number of apples in a box as you add more, or take them out. The symbol 'p' seems to mean "plus" and 'q', "equals" because they appear in the system of productions in an isomorphic relationship with our system of adding and subtracting. They have no meaning in themselves without the system in which they participate. Rather than meaning being built up by combining primitive features with innate meaning, the "primitives" get meaning from the structure in which they occur.

Computers are good at this kind of syntactic symbol manipulation, and between that and Schank's observation that it is hard to think of a thousand things you know a hundred things about, this kind of reasoning formed the basis of thinking about computers and meaning. Today a spreadsheet is a great way to automate reasoning about some things. It turns out there is a wide range of tasks that fail when handled this way with a computer. Counting apples in a box might seem easy, but a box might have 15 apples in it according to Waitrose, 25 according to Aldi, and a horse would eat the lot no matter how many worms and rotten ones there are. Often the issue is speed, so if the idea is to map a room before you vacuum it, things that move like cats and people tend to re-set the model. In the early days of robot football the machines would wait for the ball to stop before deciding what to do. Air traffic control has a similar problem. At the time of writing aircraft have been grounded while air traffic controllers sort out why an aircraft, identified by flight number, occurred in two locations. There is something wrong with the system's representation of the state of the world.

The feeling in the late 1980s was that no matter how powerful computers might become in the future, simply getting the data for symbolic representations was going to be problematic. There were projects set up to encode all the data in the world, the CYC project (Lenat et al., 1986) being the most explicit, but the scruffies were making progress and perhaps there was a better way. Rather than using sense-data to form a representation of the world, then reasoning about it before acting, functional systems could be built by connecting sensing to acting. Roomba vacuum cleaners do a random walk of a room to clean it and avoid chairs (and cats) by touching them and turning away. Since 1990 there has been a glacial shift in the computer science collective understanding of the problem and today there is wide spread acknowledgement (Wooldridge, 2023) that the nature of meaning is inseparably linked to our embodiment in the world. How this pans out for machines is still not clear and Brooks' robots are often dis-

missed as being just "insect level intelligence" with robot developers continuing to add symbolic representations in order to support higher level intelligence. Philosophers on the other hand continue to push the limits of what higher level behaviours can be explained in terms of reactive systems (Ward et al., 2017; Gallagher, 2020; Bar-On, 2021). The notion of a "reactive system" describes a mechanism used for AI problems, but it is not an interesting "algorithms and data-structures" solution. The ELIZA mechanism takes a set of if-then rules to map input to output, and applied to the DOCTOR script was found to be quite engaging. That same mechanism is used today for a myriad of chatbots but at the time and for many years after, AI researchers would consider it just a trick. From a philosopher's perspective, the system is more interesting when considered as just a part of the agent-environment system. ELIZA worked in a particular environment in exactly the same way as Roombas work in a particular environment and, the claim is, in exactly the same way we humans make most of our decisions about what to do next.

3 Humans

Deciding if I should buy that bottle of whiskey might involve reasoning with symbols representing money, but getting something to eat when I know where the refrigerator is - that is perhaps literally a no brainer. The radical enactivists (Ward et al., 2017) point out how agents and their environment are often set up so that the environment triggers; the environment "brings on", or the agent "directly perceives" something that causes the agent to behave in an appropriate way with no need for intermediate representation or indeed thinking. Like the cat, when hungry I automatically move to the refrigerator where the handle "calls out" to me to open the door. Having opened the door, a packet of gnocchi catches my eye (because I am hungry) and picking up the packet, the pan-handle and tap guide my next actions. Like the insect crossing a pebble beach[1] the behaviour looks complex. Indeed we might be tempted for whatever reason to ascribe beliefs and desires to the insect, but what unfolds is a "trajectory" *produced* by the agent but *directed* by the environment. Looking back on my actions when cooking gnocchi I will of course explain them in terms of beliefs and goals, but in the act of doing them, the mechanism does seem to be reactive. Unfortunately is is completely uninteresting to real computer scientists.

We humans are not entirely reactive agents because we can, on occasion, think about things like whether I can afford that whiskey using an expression such as "£38 - £44 = ..." but we also have another trick. We humans modify the environment to fit with us. In the cooking gnocchi for lunch scenario, none of the things I interact with are naturally occurring – someone made them. What is more, these things were put where I would find them. I might use insect level intelligence to eat lunch, but surely others must have had thoughts (with symbols) in order to create our benign environment. However it seems safe to assume other creatures do not.

Recently we have been talking with sociologists who look at human decision making at the macro level. The strong reductionist position is alive and well in sociology and many take the line that collective action is simply made up from the behaviour of individuals making their own decisions - basically that sociology is just psychology on a massive scale in the same way as drunk physicists will claim chemistry is just physics. Social simulation based on the behaviour of individuals is of course entirely possible with today's computers. There are however problems with this in that collective action does not fit with rational self interest of individuals. Why did Britain leave the EU? And why does Germany want to go to war with a nation that 2 years ago was a key trading partner? At the other end of the "structure and agency" debate (wikipedia) in sociology is the idea that society consists essentially of structures such as institutions, traditions, and norms, and that individuals are merely acting out their role in those structures. Post The Enlightenment we collectively like to think we can reason about things and make rational decisions that are better than the status quo and it is of course not very flattering to be told you are merely a cog in a machine. A key finding of AI research over the last 70 years is of course that rationality is not all it is cracked up to be. We humans do situated action using "insect level" intelligence in a benign environment, but setting up that benign environment is well beyond the capabilities of me or any other individual. A group of the esteemed homo economicus, sitting on a desert island are not going to invent agriculture, houses, trade, utility compa-

[1] I recall this example from undergraduate lectures but can't find a citation.

nies, the stock market and war. These are handed down to us from those who came before. The existence of these institutions looks designed but, the theory goes, they are actually evolved structures of successful societies. Rather than smart people having a meeting and inventing the stock-market, the stock-market, like the human eye, evolved from what came before. The stock-market might have been set up by a committee, but concepts such as trade and money already existed, as did the process of collective decision-making by committee. Whatever the mechanism of change, the current institutions have been selected for by evolutionary pressures. Society is an "ecology of practices" where the practices are the thing that survive and people are merely the actuators for roles. As our techie once said, "Cogito, ergo, I'm a cog." (Leon, 1986).

It is an *ecology* of practices because the practices fit together in some way to create a viable environment – the environment in which we live and thrive. These institutions do not, indeed cannot, stand alone. My going to the fridge when hungry works only because I was raised in a society where there are supermarkets and I do a weekly shop. I might, but usually do not, plan to eat gnocchi at one pm next Wednesday, but rather see gnocchi on the shelf and go "I'll have that". Having bought gnocchi and put it in the fridge, when I open the refrigerator door a few days later, there it is and cooking and eating practices take over.

And like an ecology in the natural environment, there is a certain amount of autopoiesis resulting in an ecology of practices being robust. In the animal kingdom, if one species is removed, the resources once used by that species are appropriated by neighbours in that space. Reduce the number of cats in Sydney, and brown snakes move in to eat the rats and mice (ABC , Australia). If the local supermarket closes, I do not starve, but engage another no doubt older hunter/gatherer practice. I wander about attentively until I find indicators of something to eat - an Aldi sign perhaps - which triggers the more modern practice of shopping.

The society's practices may or may not be visible to us mere mortals, and of course society's practices may not accord with the survival strategies of individual humans – war being a classic example. Social insects like ants and bees do not lay down their lives out of love for their fellow insect, but because they are wired that way. Indeed an individual ant does not (let's assume) have a strategy, or indeed even a practice. Instead the ant society has practices, including attacking intruders. The mechanism for implementing it is to produce ants that perform the role of soldier ants. As humans we may be tempted to claim that the ant nest has a strategy for protecting itself but that would suggest that the nest in some way has goals and representations of an enemy, war, and casualties. The notion of a practice is that it happens, with or without representations of the world or beliefs, with or without goals or desires, and with or without any sense of formulating an intention. Some have drawn parallels between termite mounds and cathedrals and suggested that a termite colony must have some form of "distributed intelligence" that enables (rational) decision making. Instead it seems we should look at Sagrada Familia and think of it as representing "insect intelligence" in humans. Gaudi undoubtedly thought about Sagrada Familia, but he did it in terms of the established practices in which individual stone masons and accountants, priests, engineers and parishioners, all fill roles.

We do have the ability to do rational choice, and we do it with symbolic representations of things in the world. But following the strong version of the linguistic relativism argument, l'sign is arbitare (Culler, 1976) and what we reason about is culturally constructed. But the things all humans want to talk about are within contingencies, contingencies that are shared across all cultures. The suggestion is that we animals behave in accordance with both the reactive framework in which the environment "causes" action, and a framework of practices which evolved through evolutionary pressures. The unique human skill is the practice of "thingifying" the stuff around us, giving it a label, and then reasoning symbolically with the label. Language is both a window on how that is done, and a mechanism for passing it on.

4 A 'pragmatics first' model of NLU

Whereas sociologists are interested at decision making at the macro level, the interest here is the mechanism for decision making at the agent level. Our aim is to find a plausible model of what goes on in heads. The position taken here is that language is indeed a window on the mind, but only the conscious part. We catch balls, and we catch a bus. We do not talk about the mechanism of eye-hand coordination directly, and we don't "see" the intricate

Figure 1: Sagrada Familia, Barcelona, vs Termite mounds in Western Australia

workings of the queue at the bus stop. For those of us interested in machines using language, we need to stop thinking about the meaning of words, and instead focus on the practices in which language is used to do things. Yes we do things with words (Austin, 1955), but the doing is embedded in practices.

In detail, returning to the problem of mind reading, the radical enacivists have claimed that in many cases what looks like reasoning about other minds and ToM is actually the "direct perception" of another's intent. (see discussion of Hutto in (Gallagher, 2020)) How might this work?

The proposal is that an agent starts with a catalogue of practices the agent knows how to do. Note that "knows how to do" is a skill rather than knowledge - possibly implemented as a boring old reactive (sub)system. At any point in time there will be a practice "in play" - the PiP - and when communication happens, it is based on a fit between a participating agent's actions and a shared PiP. The agent can predict what its conversational partner, or CP, will do next by simply remembering what happens in the PiP - no mind reading involved.

The interesting thing is when the CP does not respond as expected. Although in theory a human CP has the ability to say anything at all, in reality an utterance that is not part of the (shared) PiP will need to be accountable for as an opening to another practice from the catalogue. If it is not accountable for, the CP "risks sanction" (Wallis, 2005).

This is what is happening in the mother and child dialog above. Rather than thinking about being hungry and thinking of ways to satisfy it, the child's hunger in the presence of mother "triggers" his get-something-to-eat practice. The first step is to say (not think) "I'm hungry". The child's environment might respond with "want some cereal?". Resorting to mentalese, this is what the child "expects" or at least "hopes" will happen. And such a mechanism can be entirely Brooksean with the language produced and recognised by an ELIZA-style mechanism and the hunger and satisfaction there of being hard wired.

Of course in the example this does not happen. The child's environment (i.e. Mother) does something different to what is expected at step two. In the child's repertoire of practices however, mother's utterance is a fit with the practice of checking homework. The child then has a decision to make: continue with the practice of getting something to eat, or to switch the PiP to checking homework. It is here that the child gets to do "purposive human action" – to reason about what to do next, The thing reasoned about is *not* his choice of words however, but about his choice of practice to perform. This level of decision is not linguistic - it is not about syntax and semantics, but about all sorts of non-linguistic but very pragmatic factors such as the power relation between mother and child.

In making the decision, the child can predict what will happen next by remembering how each practice played out in the past. Presented with line 11, ("I'll let you have something later.") the child knows what is coming next if he doesn't stop niggling. We do not need to reason about other minds - the child does not need to reason about why mother is talking of homework - the child merely has to identify the practice in play and act out the appropriate role. Recognising the PiP is no more complex than recognising the words. Negotiating the PiP is where the interesting and I propose "conscious" decisions are being made.

Key to this mechanism is that the practices are largely shared by the community of speakers, and

that the practices are largely visible to others. The "system of practices" also needs to be robust as described above, which means they need to be learnable. To this end it is interesting to note that the child's role in the check-homework practice is passive - the child answers questions and waits for the next. When Mother stops asking questions, he (tentatively) reintroduces his own PiP. The suggestion is that, like the earlier example of hunting and gathering, this is a basic, fundamental, or possibly hardwired practice that a human can fall back on that enables learning new practices. The next time mother says "have you done a good job of your geography homework?" he knows what to expect. He will be able to "read mum's mind".

5 Representing practices

In retrospect our model is a modern take on Schank and Abelson's scripts (Schank, 1989). And indeed Voice XML (Voice XML) was an approach to spoken language interfaces based on the notion of forms being representatives of business practices. In the latter case the community was too focused on language understanding and filling in slots in forms to notice what the forms themselves represented. The problem Schank had was that he wanted his system to represent abstractions of events so that the system could recognise two actual events as the same (Schank, 1972). Going to The Curry Cabin on Saturday was an instance of the going-to-a-restaurant script. Today we can represent scripts not as abstractions, but more like recordings. The enactivists talk of first person intersubjectivity in which the subject experiences interacting with the world. Rather than mapping the world into things and recognising the current situation as the same as another based on the things, the idea is that we map the situation into action and then match on the actions - actions being something the agent has direct control over. In the terminology of this paper, the environment enables, or affords, a subset of the agent's known practices, and "same" means the same enabled practice(s). Counting apples is the same as counting oranges because the process of taking apples out of a box is the same for oranges. Apples and oranges are not the same when it comes to juicing them as the physical action for juicing is different.

The mechanism under development is to simply record sense and actuator data and, when a decision needs to be made, re-play the recorded data (without enacting it) to see which memory is the best fit with the current sense data. Predicting what happens next is then merely a case of remembering what happened in the past. This is of course very similar to case-based reasoning from good old AI, but without abstract representations of cases.

6 What LLMs are actually doing

In addition to "first person intersubjectivity" the enactivists also talk of "second person intersubjectivity" which is when an agent can put themselves in the shoes of a person interacting and "experience" the interaction vicariously. Observing someone cook gnocchi and then at some time in the future cooking it oneself, the conjectured mechanism does not remember pots, stoves and water, nor reason about why he put the lid on the pot, but simply "silently" copies the actions of the cook while the cook is cooking. These vicarious actions are then replayed when it comes time to cook gnocchi. The key is that abstraction of sense data to things such as pot lids such as it is , is done by "mirror neurons" or something similar. What is remembered is first person actions mirrored from the second person experience.

Applied to conversation the mapping from observed action of a second party to one own speech-acts is very straight-forward: one simply says the words that one heard someone else say in that situation. For disembodied conversation such as phone calls or text on the internet, all that is needed is recordings or text. This is of course what LLMs are trained on.

The proposal is that the things we write down are to a very large extent descriptions of practices. Sure there are references in there to things, but LLMs don't capture that. A Generative Predictive Transformer – or any mechanism for generative prediction – is not "understanding" that a cat has four legs, sharp teeth and fur, what it is learning is that we pat cats, we attend when they catch mice, and that they jump on window sills and knock things off. The things we notice about cats are all integrated into our own practices, and what we talk about are those practices.

What an LLM is doing to generalise from N descriptions of a practice to a rewriting of the practice selected and modified by the current prompt. Prompted by "I'm hungry" an LLM will "select" the practice of mother feeding a child - because this is a standard opening for a child to a mother -

and fill it in. If the prompt includes the statement that the child likes cereal, then "want some cereal" would be a very plausible magic response.

7 Conclusion

Although we can and do think about what other people are thinking, we generally don't bother. Much of what we do can be attributed to either reacting to stuff around us, or to the adoption of a role in a known practice. For social animals the practices are largely shared, and visible. That is, others can see what an individual is doing, and because the practice is shared, they know what the individual will do next. For humans, and the higher animals at least, practices can be learnt. The cat learns about the getting up in the morning routine; knows where the litter tray is and so on. People - language users - use language as part of practices, but language also empowers new practices. The symbolic nature of language reifies stuff into things so we can think about all cats when our experience is of just a few. We share these symbols - the sign and the signified - to invent (the practice of) counting and holding committee meetings, and we develop theories about "things" and how they interact. We can think about fairies, God, hermeneutics, and other people's intentions.

For the engineers wanting to build conversational user interfaces, it seems there might be potential in looking at language in use as being, not about syntax and semantics, and not a reference to things, but about negotiating a common practice between the speaker and his or her CP.

References

ABC (Australia). 2023. Sydney snake catchers inundated with calls. Https://www.abc.net.au/news/2023-09-27/sydney-snake-catchers-busy-hot-weather-habitat-shortage/102890204.

J. F. Allen, L. K. Schubert, G. Ferguson, P. Heeman, C. H. Hwang, T. Kato, M. Light, N. G. Martin, B. W. Miller, M. Poesio, and D. R. Traum. 1995. The TRAINS project: A case study in defining a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7(7):7–48.

J. L. Austin. 1955. *How to do Things with Words*. Clarendon Press, Oxford, UK.

Dorit Bar-On. 2021. How to do things with nonwords: Pragmatics, biosemantics, and origins of language in animal communication. *Biology and Philosophy*, 36(6):1–25.

Rodney A. Brooks. 1991. Intelligence without representation. *Artificial Intelligence*, 47:139–160.

P. R. Cohen and H. J. Levesque. 1990. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261.

Jonathan Culler. 1976. *Saussure*. Fontana Press.

Shaun Gallagher. 2020. *Action and Interaction*. Oxford University Press.

S. Harnad. 1990. The symbol grounding problem. *Physica D*, 42:335–346.

D. R. Hofstader. 1979. *Godel, Escher, Bach: An eternal golden braid*. The Harvester Press, Hassocks, Sussex.

Hide Ishiguro. 1990. *Leibniz's philosophy of logic and language*. Cambridge University Press.

D. Lenat, M. Prakash, and M. Shepherd. 1986. Cyc: Using common sense knowledge to overcome brittleness and knowledge acquisition bottlenecks. *The AI Magazine*.

Leon. 1986. (personal communication).

William C. Mann. 1988. Dialogue games: Conventions of human interaction. *Argumentation*, 2:511–532.

Roger C. Schank. 1972. Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology*, 3(4).

Roger C. Schank. 1989. An early work in cognitive science. A retrospective on Schank and Abelson's Scripts, plans, goals and understanding: an inquiry into human knowledge structures, 1977.

Michael Tomasello. 2008. *Origins of Human Communication*. The MIT Press, Cambridge, Massachusetts.

Voice XML. 2004. The voice xml standard. Http://www.w3.org/TR/voicexml20/.

Peter Wallis. 2005. Robust normative systems: What happens when a normative system fails? In *Abuse: the darker side of human-computer interaction*, Rome.

D. Ward, D. Silverman, and M. Villalobos. 2017. Introduction: The varieties of enactivism. *Topoi*, 36:365–375.

wikipedia. 2024. The structure and agency debate. Https://en.wikipedia.org/wiki/Structure_and_agency.

Michael Wooldridge. 2023. Royal institution christmas lectures.