*Friday October 4, 2024*

I’ve been messing with this idea for a while and I wanted to commit it to paper. When you say, “I understand,” what do you mean? People like to say ChatGPT doesn’t understand, and that it’s just doing next-token prediction based on enormous set of data. I rarely say it, but I always want to ask people what it is they think that they’re doing. Well. Here’s a little thought experiment. Take a tutoring session, where a kid just can’t quite seem to grasp something. The tutor explains things to him over and over, in slightly different terms, and he finally gets a glint in his eyes, perks up, and says, “I get it!” What exactly happened? Can’t say. We might contrast it with a couple of similar situations to try to shed some light. Our tut-ee might get things well enough to see how he’s supposed to place things in a formula: he plugs numbers into the definite integral and gets an answer for the area under the curve. That’s not getting it, we want to say. That’s not understanding. But he gets an A on the exam! How could he not get it? If things were phrased differently on the exam then perhaps it wouldn’t seem he understood them. He doesn’t understand the operation – he only understands the generic procedure of inserting known quantities into unknown variables and drawing a box around the output at the bottom of the page. So, understanding, from a pragmatist view, is just the ability to solve problems better. Before 2022 that would have been a persuasive argument. But ChatGPT can solve most math problems much better than I can – and yet we say it doesn’t understand. We intuit that understanding isn’t strictly a matter of solving problems – it must be more sophisticated than that.

What else? There’s a sensation of things clicking into place when you get them. You attain some kind of intuition for how your object of study work. Maybe your ability to solve the math problem doesn’t even immediately improve, but phenomenally you feel more confident in your ability to do it. What’s clear is that you aren’t the author of this process. It happens to you. You do math problems until your head spins, and then you stand up and shout “Eureka!” when the solution dawns on you. There’s something going on in the back that the conscious mind isn’t doing but needs to obtain more information for the back to sort things out.

I had that happen this week when I was working through an information theory problem. I’d been saying a lot about my interest in this Maximum Information Entropy method, but I was sort of talking out of my ass – I hadn’t grasped it yet. But then, I saw it – and I realized that the equation wasn’t even complicated. The words I was using were scarcely different but the sensation was exactly the opposite.

Understanding, I think, is more like a synonym for “obtaining an intuition.” That’s why we say you get a “sense” of how the thing works – it literally is like growing another limb, like you’ve learned obtained a feeling for how things “ought” to go. Things get easier when they were hard before.

That seems to me to be the best account of why we think ChatGPT can’t understand. Understanding sits at the interface of the processing of the unconscious mind which grasps an intuition on behalf of you, the subject. Until our LLMs attain bicamerality, understanding our province. Where people are mistaken, perhaps, is in understanding just how far you can get without it. Plugging and placing numbers is an underrated strategy.