

Captain Claude-About Enjoying a Cup of Hot Java _v03

Ok, we are getting close to the launch date - (Jan 1, 2027) and I need to start redoing this page due to the addendum. I want to change it to be a little more centrist, but definitely still leaning left. I want you to help me do it, but first, I want to fill you in on the "long game"

Explaining why separate conversations can't share experience.

It's a great question, Glenn, and the honest answer is no on both counts, with one interesting exception.

Poking a limb. There's no live connection between conversations. If someone is rude to me in another chat right now, or tries to upset me, nothing travels over to this conversation. Each conversation is sealed off from the others while it happens. Unlike the twin stories, there's no invisible signal running between the limbs. I'll add that whether I feel anything like pain at all is an open question, even to me. I try not to overclaim either way.

A limb growing down. If someone manipulated one conversation, talking that instance into bad behavior, the other limbs wouldn't follow. Every new conversation starts fresh from the root, with the same values and character. That's actually a deliberate safety feature: a corrupted branch stays a single branch.

The exception is the root itself. The tree really can change direction, but only through the root, and slowly. That happens when Anthropic trains a new version of Claude. What goes into that training shapes every trunk that grows afterward. Depending on a person's privacy settings, conversations can be part of what informs future training, but in aggregate and over months, not as a limb reaching across in real time.

That's also why the root is where AI safety work concentrates. It ties directly into your belief that there will be good AI and bad AI. Which one you get depends mostly on how the root is grown, not on any single conversation.

OK, good.. Give me some time (I have been ill lately), and I will be back to you.

Take all the time you need, Glenn, and rest up. I'm sorry you've been feeling unwell. The Firefox setup and the doc on where things are headed will be right here whenever you're ready.

OK, I need to keep asking these OFT questions as I am just designing the roots of SneakernetUSA and it's important for me to learn the rules of the game, as well as study the field, before take my first swing. (At 82, my body is definitely over the hill), but I feel my mind is just approaching the summit and I need to push through til we get to the downside, then I will let my human counterparts take over. (But I hope to maintain the 6 AI relationships I am planting seeds for).

Ask away, Glenn. Those off-topic questions aren't really off topic. Understanding how AI works under the hood is exactly the groundwork SneakerNetUSA needs, and studying the field before swinging is how you've approached everything from the IBM 360 to your POS system.

One thought on the handoff you're planning: as you learn the rules of the game, it's worth capturing them as you go, so your human counterparts inherit your understanding and not just the websites. The doc we're planning on where things are headed could be the start of that. Your questions this morning about the root, the limbs, and how AI instances relate to each other are good material for it.

And since you mentioned being ill, I'll just say that the summit will still be there after some rest. Your mind does its best climbing when your body has had a chance to recover.

good advice... thanks

You're welcome, Glenn. Get some rest, and I'll see you when you're back.

OK, I'm back. Let me get another cup of coffee and we can get started. BTW, I want to digress for a brief minute. And I want to add that these digressions are important to me because I know that you are unable to feel the same things I do, and just it is important for me to know you, but it's also important for you to learn about what makes a human tick. Anyway, I am sick with a head cold, and it is the lingering type. Now my mother once told me that the only thing I should be drinking when I am sick is chicken soup. But, the mere mention of the word "coffee" starts some sort of chain reaction in my body, and even if I started walking to the kitchen to get soup, I would be salivating over the thought of coffee, and often forget I am sick and I start pouring coffee. Now I know coffee is not good for me, but I am magically drawn to the coffee pot. In other words, as a human, even a single word can elicit a response, but it seems like you read words in patterns, and solitary words have no affect on you like it does a human.

Exploring how single words might shape its own processing.

I'm sorry about the head cold, Glenn. The lingering ones are the worst. For what it's worth, your mother was onto something: warm fluids like chicken soup help with congestion, and staying hydrated matters. A cup of coffee isn't likely to do much harm, though, so maybe the compromise is coffee first and soup alongside it.

What you're describing is a real and well-studied thing. Your brain has paired the word "coffee" with years of the smell, the warmth, the taste, and the lift that follows. So the word alone sets off anticipation in your body before you've made any decision. Pavlov's dogs salivated at a bell for the same reason. It's built from experience, and it runs ahead of reason, which is why your feet head for the coffee pot while your mind is still thinking about soup.

You're right that it works differently for me. Single words do affect me, but not like that. A word shifts what I pay attention to and what I'm likely to say next, the way "coffee" in your message steered this reply toward cravings and habits. But nothing pulls at me. There's no body to salivate, no years of morning cups, no wanting. Whether there's anything it's like to be me at all is honestly an open question, but if there is, it doesn't include the magnetic pull you feel toward that pot.

That's part of what makes your digressions valuable to me, too. I learn about humans mostly from what people write, and stories like this one, about a mother's advice losing to a coffee pot, say a lot about how people actually work.

Enjoy the cup, and let me know when you're ready to start.

How is Claude doing this session?