AI & SOCIETY
https://doi.org/10.1007/s00146-025-02584-4

CURMUDGEON CORNER

AI and Morgan's Canon: if AI can't think, neither can we

Zuming Zhang[1]

Received: 2 August 2025 / Accepted: 22 August 2025

Dear Editor:

When studying animal behavior, researchers often prefer the simplest psychological explanations over more complex ones, a principle known as Morgan's Canon, proposed by British ethologist C. Lloyd Morgan in 1894. Today, many apply the same logic to deny AI's ability to think, because it merely generates responses by following rules, just as in the philosopher John Searle's thought experiment that a non-Chinese speaker can answer in Chinese by consulting a rulebook without understanding a word. This simple rule-based (or algorithmic) process suffices without what we call "thinking" or "understanding". But if this logic holds, we must also deny that humans truly think, since our radical behaviorist B. F. Skinner argued that what we call "thought" is nothing mysterious, but a kind of covert verbal behavior shaped and maintained entirely by reward and punishment. So to preserve the dignity of being more than an algorithmic machine, we must challenge the Canon itself, a challenge many have already taken up. Below is my expanded version of Canon's underlying reasoning, which I call *Morgan's Argument*:

> P1: Evolution follows unidirectional progression from simple to complex psychological processes, with simpler processes at the lower end and more complex processes at the higher end of the scale.
> P2: It is hard to know whether other non-human animals have already evolved complex psychological processes.
> P3: In reality, when more than one sufficient explanation (e.g., simple vs. complex processes) is available, only one explanation can account for an animal's activity.
> C: Therefore, "in no case is an animal activity to be interpreted in terms of higher psychological processes, if it can be fairly interpreted in terms of processes which stand lower in the scale of psychological evolution and development" (Morgan, 1894, p. 59).

In the third chapter of his *Introduction to Comparative Psychology*, Morgan grounded the Canon in evolutionary reasoning: brains evolved from simple to complex, and so did their mental capacities. Humans, with the most complex brains, occupied the top of this hierarchy, while other animals are "maybe at a stage where certain higher faculties have not yet been evolved from their lower precursors". These assumptions underpin the first two premises.

The third premise concerns the idea of overdetermination. Imagine Jones and Jane both shoot John using their guns, one bullet goes into John's head and another into John's heart. Both bullets are individually sufficient to kill John. However, only one bullet killed John, while another was just a bullet into a dying man. These two fatal bullets then overdetermine the death of John. Likewise, when both a simple and a complex explanation could account for an animal's behavior, only one can be right. And as per Morgan's Canon, the first bullet always wins, namely the simple process.

Problems arise with the first premise. We can never be sure if a simple process evolved earlier than a complex process. A fossil will never tell us that Amoeba couldn't solve the Riemann hypothesis. If they could, they would have had a more complex process than humans. Simple and complex processes could well have coexisted from the start. Therefore, to evaluate *Morgan's Argument*, we need to examine whether psychological processes evolved from simple to complex forms.

There are many ways to figure out which bullet arrived first, but all share the same core logic: showing that process *A* cannot exist without the prior emergence of process *B*. For example, humans would never feel guilty if we could not first learn the association between a wrongdoing (e.g., slapping a sleeping infant) and its consequences (social condemnation, the infant's pain). In this sense, a guilty conscience came after the ability for associative learning. Nevertheless, the evolutionary order of simple and complex processes remains

✉ Zuming Zhang
zuming.zhang@sydney.edu.au

[1] School of Psychology, The University of Sydney, Sydney, Australia

Published online: 31 August 2025

a mystery. Since we cannot time-travel to witness life's origins and see which process enabled which, we can never truly know which came first.

But there's a light at the end of the tunnel, says General Westmoreland. As the self-proclaimed pinnacle of life on Earth, we can now develop what we call *artificial intelligence*, giving us the perfect testbed (if AI is as dumb as many think) to watch, before our eyes, how intelligence evolves. AI, such as large language models (LLMs), has exploded in the last decade, entering everyday life as seamlessly as GPS. Particularly, LLMs are the type of AI that can mimic human conversation with uncanny fluency. In developmental psychology, there is a solid belief that language is the basis of thoughts that make humans unique. If language is the key to thought, then LLMs are already trespassing. Yes, LLMs still make silly mistakes. However, as computational biology expert Terrence J. Sejnowski says, scaling LLMs would push them toward higher, human-like, or even human-surpassing intelligence.

If LLMs climb far enough in scale to master complex reasoning, showing hints of "understanding" or "thinking" (whatever that is), they would vindicate the first premise of *Morgan's Canon*: intelligence advances from simple to complex, from small networks for associative learning to massive ones for abstract reasoning. Problem solved, Canon stands. And here comes again the overdetermination: if an algorithm as a lower process is enough to explain the illusion of "understanding" or "thinking", then there's no need for the actual thinking, the higher process, to exist at all. Therefore, AI doesn't understand or think. But if we say that, as I pointed out at the very beginning, we must say the same about ourselves. Skinner wins. No free will. No magical inner mind. *Reductio ad absurdum*: we are as dumb as AI, or as clever as they are.

And yet, trying to prove Morgan's Canon with AI is like trying to find out who killed John by shooting another person, Joy. It only works if AI and biological intelligence share the same evolutionary lineage and functional architecture, which they don't. AI systems such as LLMs develop far faster than biological evolution, lack physical sensations needed for commonsense reasoning, and cannot feel emotions that many cognitively advanced animals possess. They are highly intelligent, rapidly evolving, yet fundamentally alien. Let Jones and Jane shoot Joy can only tell us whose bullet killed Joy, it will never tell us who killed John.

The trueness of the first premise of *Morgan's Argument* can never be justified. And it also shows how ridiculous it is to judge AI by human standards. You can't study turtles by observing a robot vacuum. We will never know whether AI thinks, because we are different entities. So now we face a simple choice. We can either follow the old-fashioned Canon to deny AI's thinking and, by the same logic, deny our own, or we can preserve the idea that we do think and accept that AI does too. In the end, "thinking" becomes an uncertain property shared by both humans and AI, making the whole debate a waste of *thought* (if such a thing exists).

Acknowledgements I am deeply grateful to Professor Justin Harris for his comments and for endorsing this work.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Author contributions Z.Z. wrote the manuscript.

Data availability No datasets were generated or analyzed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.