

Cognitivism vs. Human Supremacy

In this paper I will be outlining the position of the well-known philosopher, Hubert Dreyfus, on computation and intelligence, after which I will assess the claims of his argument.

Dreyfus argues that humans inherently possess some sort of rich commonsense knowledge that is a unique to our minds. He claims that this knowledge is dependent on “a mixture of culture, context, upbringing, and bodily self-awareness” (Clark [1991], *Microcognition*, p. 25). This dependency obviously excludes machines and computers, and probably most animals, from possessing ‘intelligence’ in the special way that humans do.

To understand Dreyfus’ position, it is important to look at the context in which he was writing. At the time, robots and AI were being made and fine-tuned to complete very specific tasks, rather than the multiplicity of tasks that humans do both simultaneously and over long periods of time. The performance of these robots was sometimes impressive (eg: Winograd’s SHRDLU), but none was a true victory for AI, as they could not expand their repertoire to include all undertakings that a human would do normally. Every function performed was done so in a ‘microworld’. In other words, in an isolated environment where only one specific task was being evaluated and only certain kinds of stimuli being measured and applied. For example: robots have been designed and created to be amazing in playing the game of chess, but those same robots are not capable of reading human emotion, or

responding verbally to an individual. Its talent lies only in the realm of chess rules and strategies.

Based on these exhibitions, Dreyfus concluded that it would be impossible for AI to expand to the point of equaling human understanding. Even comprehending a simple statement or interaction requires the knowledge of multiple concepts, and the ability to use them together. Not to mention, concepts are hardly complete, unchanging entities. Humans develop concepts slowly over many experiences, and reflections of those experiences, so that those concepts are being continually built on and changed over a human's life. Dreyfus did not believe that there would ever be AI that would be capable of mastering the seemingly infinite number of microworlds that humans do on a day-to-day basis.

This very flexible, plastic characteristic of the human mind is a by-product of 'unconscious knowledge'. Unconscious knowledge is what humans use to intuit in many or all circumstances. In other words, the human mind draws from past experiences and memories by using associations or linking those past events to the current event, all of this happening without the human having to consciously go through the motions. This differs from the classical approach of computer programs and AI, as those analyze the input by trying to systematically match it up with a rule, and then delivering the same output every time. Due to this, Dreyfus argues that the flexibility and complexities in human reason cannot be paralleled by AI, as it is limited by stable and finite rules. From this description of the functions of both systems, the human mind

and AI, it is hard to see how a formal machine could gain the same kind of semantic understanding as a human does.

The idea of “frames” is brought up briefly in Chapter 2. Clark aptly describes them as “a data structure dealing with a stereotypical course of events in a given situation. It consists of a set of modes and relations with slots for specific details” (Clark, 26). Dreyfus argues that frames could hardly cover all possibilities, details, and subtleties of situations that humans are naturally inclined to deal with. He further points out that humans are also easily able to transition from one frame to another. For AI systems to do this, they would have to have rules governing when to change the frame (which they use to know what rules to apply to any given situation), and rules to determine when those rules must be used, ad infinitum, it would seem. This argument seems to effectively distance the classical conception of AI systems from that of humans.

Apart from his arguments against the theories of cognitivism, he also offers two separate accounts of human cognition. The first is the critical mass argument broken down into premises and conclusion form:

- P1 - You must know a certain amount about something (called a critical mass) before you know anything about it at all.
- P2 - The critical mass is arrived at through physical, social and cultural interactions with the world.
- C - Therefore, anything that lacks humans’ “socially situated and embodied advantages” (Clark, 28) cannot know anything.

This argument fails in a number of ways. First, the “critical mass” idea in itself is far too abstract and subjective to be applicable in any specific cases. There is no one that could determine how much ‘knowledge’ an individual must

ascertain in order to have 'knowledge', or what criteria must be filled on varying subject matter. It appears that Dreyfus did not think he would have to support or define his 'critical mass', and I doubt that he would be able to upon request. This takes away from his argument.

Second, we ascribe understanding to children and adults that are developmentally delayed although they by no means possess either entirely correct beliefs or a wealth of knowledge. I think that Dreyfus, and all people for that matter, need to be very careful in their arguments about the uniqueness of human intelligence. It would seem that there is a sliding scale of what we consider 'intelligent', and every time an animal or AI system approaches where humans sit, we search for a characteristic that distinguishes 'us' from 'them' and bump our little marker up the scale a bit further. I would posit that Dreyfus did just that, and failed to take into account the implications it would have on anyone besides the typical human adult.

His second argument in premise and conclusion form:

- P1 - Humans naturally identify the key features of a situation in order to respond.
- P2 - Humans do this by remembering past experiences, using a holistic method to link memories.
- P3 - AI systems respond to situations by employing a set of stored rules and propositions.
- C - Therefore, humans have a flexibility and ability to intuit that AI systems, by design, lack.

This argument seems better thought out and put together than his previous account for the uniqueness of human intelligence. By appearance only, it would seem that Dreyfus makes a simple, but good point. His premises seem to be true, but I do not believe that the truth of the premises guarantee the

validity of the conclusion. The context of Dreyfus' assertions seems to be important to me, as I think he is making assumptions that science did not, and perhaps has not yet, come to. It looks to me like Dreyfus is assuming that the holistic method that humans use to link memories and make associations to current situations is not translatable to a different kind of computational theory. Though humans have come very far in the realm of science, there is still much to learn, specifically in neuroscience. Due to this, I do not believe that it is logical or reasonable to leap to the assumption that there is no way in which an AI system could someday be capable of replicating the intricacies and complexities of human intelligence.

In conclusion, while Dreyfus may have had reason to make some assumptions regarding AI's limits - and the human brain's infiniteness - during the time of his earlier writing, his views are no longer defensible in the current realm of science. Regardless of the end-result, Dreyfus has contributed thought-provoking arguments to the discipline of philosophy of mind.