On the Early History of Artificial Intelligence and Its Philosophical Significance

Andrew H. Zhang

University of Toronto St. George

COG250Y1: Introduction to Cognitive Science

Dr. J. Vervaeke

Last Revised: January 31, 2022

Arising in the wake of the scientific revolution, the idea of the equivalence between cognition and computation is elucidated by the arguments between Thomas Hobbes and René Descartes regarding whether meaning and purpose can spontaneously emerge from the interactions between inert matter. The brilliance of the artificial general intelligence hypothesis originates from the ability of constructing an autonomous reasoner to resolve this argument and the profound philosophical implication it has. The various technical hurdles that plagues the project of creating artificial general intelligence draws our attention to the deeper philosophical problem underlying the Homuncular nature of artificial general intelligence under the framework established by the scientific revolution. However, one can reasonably argue that these issues can be addressed by clarifying the distinction between algorithmic play and algorithmic implementation and seeing some incomputable "qualia" as emergent properties.

Initially, proper treatment of the history of artificial general intelligence demands a more indepth consideration of the individual arguments of Thomas Hobbes and René Descartes in context of the scientific revolution. By defining matter as objectively "inert" and without purpose, the scientific revolution overthrows the previous paradigm defined by the Aristotelian world view that explains natural forces through how matter has a "preferred" position in space and time or that they have intrinsic "purpose". The scientific revolution also establishes the disconnection between the mind and the world with subjective experience as a highly unreliable medium, with mathematics being the sole channel between the two. The implication of this is that unquantifiable qualities, or "qualia" that are inoperable by mathematics does not have the same objective existence as inert matter in the paradigm as established by the scientific revolution. With this as premise, the English philosopher Thomas Hobbes proposes that under the

assumption that mathematics is the sole reliable channel between the mind and the objective world, human thought is therefore computation, and subsequently, if computation as a formal system can be reliably mechanized, so does the mental processes of human beings. However, Hobbe's contemporary French mathematician René Descartes refutes this argument with the following logic adhering to the principles of the scientific revolution, which follows that the framework established by the scientific revolution which regards matter as completely inert and therefore incapable of possessing "qualia" such as meaning and purpose. For an artificial reasoner constructed purely of inert matter, it will therefore by induction be incapable of possessing "qualia" such as meaning and purpose, which is essential in the human mental process to facilitate logical and scientific reasoning along with mathematics in the first place, and therefore this reasoner will not be able to achieve the same general level of cognition and intelligence as the human mind.

Further, the brilliance of the artificial general intelligence hypothesis is immediately apparent if we consider its role in settling the feud between Descartes and Hobbes along with the profound philosophical implications that it would have in doing so. If it is indeed possible to manufacture an automated general problem solver completely from inert matter as suggested by Hobbes, it would resolve the argument by showing the flaw within Descartes' argument. This has two possible implications. If we strictly adhere to the doctrines of the scientific revolution that "qualia" could not be possessed by inert matter nor recreated from their interactions, the existence of an artificial general intelligence and the successful recreation of meaning and purpose as qualia could imply that the meaning and purpose that drives our mental are merely just illusions that do not objectively exist. This nihilistic philosophical conclusion would

propound the notion that there is nothing intrinsically special about the human mind and consciousness. Alternatively, we can also consider that perhaps the assumption that meaning and purpose are qualia that underlies Descartes's argument is incorrect, namely that the seemingly mathematically inoperable abstract concepts "meaning" and "purpose" could somehow be represented by formal systems. This philosophical implication lends itself to the possibility that our mind could extend beyond the mortality of our bodies, given the medium independence and multiple realizability of formal systems.

However, despite the recent advances in narrow artificial intelligence, the technical hurdles encountered by the project of developing artificial general intelligence really draws our attention to how the very notion of artificial general intelligence could be paradoxical and Hommuncular in nature under the framework laid down by the scientific revolution. Consider the fundamental assumption of the scientific revolution that mathematics is the only means for us to accurately understand reality. If we take a step back and examine the forces behind our drive to do mathematics and scientific reasoning, we would find that it fundamentally distills down to qualia such as curiosity, meaning, and purpose. This is extremely problematic, as the scientific revolution that is centered around mathematics driven by these forces has negated their own objective existence. Arguably, however, this problem can be readily resolved if we further expand on the idea that these qualia could be fundamentally mathematically representable by formal system by considering them as a emergent behavior that could arise when enough instances of inert matter has amassed to behave in a certain pattern in unison, similar to how the concept of "wetness" is independent of the individual chemical properties of each molecule but emerges from the unified behavior of water molecules when enough has amassed under certain

conditions. Another problematic aspect of artificial general intelligence lies with how we could reliably tell if an artificially created agent appearing to behave as if generally intelligent is actually generally intelligent instead of just merely systematically mimicking the behavior of a being that is generally intelligent. This problem is elucidated by the Chinese room argument, which describes a scenario where a non-Chinese speaker in a confined room can systematically produce fluent Chinese sentences that leads a Chinese speaker to believe the display of apparent understanding of being able to speak Chinese despite the opposite being true. However, one can argue that this argument is flawed in that it confounds algorithmic implementation and algorithmic play. The non-Chinese speaker confined within the room operating its machinery, despite his various conscious actions, still exists on the algorithmic implementation level just as how neurons in our brains that fire algorithmically lacking a mind cannot be used to argue that we lack a mind. Therefore the emergent behavior of apparent Chinese understanding as a Gestalt does not exist on the level of the non-Chinese speaker, but rather on the level of him/her and the room that he/she is confined within, which exists on the level of "play" and cannot be dismissed as mere pretense by the lack of understanding of the non-Chinese speaker on the algorithmic implementation level. This is abstractly the same confoundment as mentioned by Dr. Vervaeke during the lecture regarding how the fixation upon neuroscience could obstruct the study of the mind.