

# Artificial Intelligence

As a theory in the philosophy of mind, artificial intelligence (or AI) is the view that human cognitive mental states can be duplicated in computing machinery. Accordingly, an intelligent system is nothing but an information processing system. Discussions of AI commonly draw a distinction between weak and strong AI. Weak AI holds that suitably programmed machines can simulate human cognition. Strong AI, by contrast, maintains that suitably programmed machines are capable of cognitive mental states. The weak claim is unproblematic, since a machine which merely simulates human cognition need not have conscious mental states. It is the strong claim, though, that has generated the most discussion, since this does entail that a computer can have cognitive mental states. In addition to the weak/strong distinction, it is also helpful to distinguish between other related notions. First, cognitive *simulation* is when a device such as a computer simply has the same the same input and output as a human. Second, cognitive *replication* occurs when the same internal causal relations are involved in a computational device as compared with a human brain. Third, cognitive *emulation* occurs when a computational device has the same causal relations *and* is made of the same stuff as a human brain. This condition clearly precludes silicon-based computing machines from emulating human cognition. Proponents of weak AI commit themselves only to the first condition, namely cognitive simulation. Proponents of strong AI, by contrast, commit themselves to the second condition, namely cognitive replication, but not the third condition. Proponents of strong AI are split between two camps: (a) classical computationalists, and (b) connectionists. According to classical computationalism, computer intelligence involves central processing units operating on symbolic representations. That is, information in the form of symbols is processed serially (one datum after another) through a central processing unit. Daniel Dennett, a key proponent of classical computationalism, holds to a top-down progressive decomposition of mental activity. That is, more complex systems break down into more simple ones, which end in binary on-off switches. There is no homunculi, or tiny person inside a cognitive system which does the thinking. Several criticisms have been launched against the classical computationalist position. First, Dennett's theory, in particular, shows only that digital computers do not have homunculi. It is less clear that human cognition can be broken down into such subsystems. Second, there is no evidence for saying that cognition *is* computational in its structure, rather than saying that it is *like* computation. Since we do not find computational systems in the natural world, it is more safe to presume that human thinking is only *like* computational processes. Third, human cognition seems to involves a global understanding of one's environment, and this is not so of computational processes. Given these problems,

critics contend that human thinking seems to be functionally different than digital or serial programming.

The other school of strong AI is connectionism which contends that cognition is distributed across a number of neural nets, or interconnective nodes. On this view, there is no central processing unit, symbols are not as important, and information is diverse and redundant. Perhaps most importantly, it is consistent with what we know about neurological arrangement. Unlike computational devices, devices made in the neural net fashion can execute commonsense tasks, recognize patterns efficiently, and learn. For example, by presenting a device with a series of male and female pictures, the device picks up on patterns and can correctly identify new pictures as male or female. In spite of these advantages, several criticisms have been launched against connectionism. First, in teaching the device to recognize patterns, it takes too many training sessions, sometimes numbering in the thousands. Human children, by contrast, learn to recognize some patterns after a single exposure. Second, critics point out that neural net devices are not good at rule-based processing higher level reasoning, such as learning language. These tasks are better accomplished by symbolic computation in serial computers. A third criticism is offered by Fodor who maintains that connectionism is presented with a dilemma concerning mental representation;

1. Mental representation is cognitive
2. If it is cognitive, then it is systematic (e.g., picking out one color or shape over another)
3. If it is systematic, then it is syntactic, like language, and consequently, it is algorithmic
4. However, if it is syntactic, then it is just the same old computationalism
5. If it is not syntactic, then it is not true cognition

But connectionists may defend themselves against Fodor's attack in at least two ways. First, they may object to premise two and claim that cognitive representation is not systematic, but, instead, is pictorial or holistic. Second, connectionists can point out that the same dilemma applies to human cognition. Since, presumably, we would want to deny (4) and (5) as pertains to humans, then we must reject the reasoning that leads to it.

The most well known attack on strong AI, whether classical or connectionist, is John Searle's Chinese Room thought experiment. Searle's target is a computer program which allegedly interprets stories the way humans can by reading between the lines and drawing inferences about events in the story which we draw from our life experience. Proponents of strong AI say that the program in question (1) understands stories, and (2) explains human ability to understand stories (i.e., provides the sufficient conditions for "understanding"). In response, Searle offers the following thought experiment. Suppose that a non-Chinese speaking person is put in a room and given three sets of Chinese characters (a script, a story, and questions about the story). He also receives a set of rules in English which allow him to correlate the three sets of characters with each other (i.e., a program).

Although the man does not know the meaning of the Chinese symbols, he gets so good at manipulating symbols that from the outside no one can tell if he is Chinese or not Chinese. For Searle, this goes against both of the above two claims of strong AI. Critics of Searle contend that the Chinese Room thought experiment does not offer a systematic exposition of the problems with strong AI, but instead is more like an expression of a religious conviction which the believer immediately "sees" and the disbeliever does not see.