Is Artificial Intelligence an empirical or a priori science?

Abstract
This essay concerns the nature of Artificial Intelligence. In 1976 Allen Newell and Herbert A. Simon proposed that philosophy is empirical in nature, akin to chemistry or physics. They saw AI as the generation and testing of hypotheses, using the same scientific paradigm we see in the natural sciences. However in 1989 Andre Kukla argued that this is not the case, and that AI is in fact an a priori science, more similar to mathematics than chemistry. He posited that computer programs, the bread and butter of AI, are a priori endeavours, therefore AI is a priori in nature. This essay discusses both arguments, and the evidence each side uses to support their claim. The conclusion addresses which line of reasoning comes out on top, and whether the authors are really addressing the same issues.

Key words: Artificial Intelligence; Philosophy of AI; nature of AI; empirical science; a priori science
Is Artificial Intelligence an empirical or a priori science?

In 1968 Marvin Minsky asserted that Artificial Intelligence is “the science of making machines do things that would require intelligence if done by men”. Searle (1980) later expanded on this definition, distinguishing between different types of AI by introducing the idea of strong versus weak AI. Strong AI hinges on the hypothesis that an Artificial Intelligence system could possibly be created that could actually think and possess a mind (in the philosophical sense). Weak AI on the other hand “merely” hypothesises that an Artificial Intelligence system could imitate thought but would not actually be in possession of a mind. Strong AI is similar to the type of AI we see in Hollywood blockbusters, where the robot doesn’t just walk and talk like a human, they can think like a human too. Weak AI is more of a form of “theoretical psychology” (Longuet-Higgins, 1981), a way of testing psychological theories without the inconveniences associated with human subjects such as ethical treatment (Buchanan, 1994).

Newel and Simon (1976) posit that both forms of AI come under the umbrella of empirical science. The Oxford English Dictionary defines empirical as “based on, concerned with, or verifiable by observation or experience rather than theory or pure logic”. Biology for example is an empirical science. We know what a cell looks like because we have seen it. We know what it does because we’ve (in a scientific manner of course) prodded and poked it, and seen what happens under different conditions. Kukla (1989) however argues that AI is an a priori science (a priori is defined by the same dictionary as “based on theoretical deduction rather than empirical observation”), like mathematics. In mathematics, we do not physically have to cut a pie in half to figure out that two halves make a whole. By all means we could. It would certainly make mathematics a more rewarding enterprise if it directly involved delectable bakes goods. But it’s not necessary. We won’t learn any more than pure logic could have worked out. Clearly, these are two radically different views on the scientific underpinnings of Artificial Intelligence. Before we begin to answer the question of who is right, let us start with how the whole argument came about.
It began with a lecture given upon their presentation with the Turing Award in 1975, Allen Newel and Herbert A. Simon argued that the study of computers is an empirical discipline. They said that “Each new program that is built is an experiment”. As they envisage it, the program poses a question, it physically embodies the hypothesis, and the resulting “behaviour” as they put it, helps us to answer the hypothesis. In some ways, they have a point. Most computer programs are far too long and complicated to know exactly what they are going to do before they actually do it. As Simon wrote in 1995, “The moment of truth is running a program”. Furthermore he wrote that an Artificial Intelligence system produces data which can be empirically studied using the same scientific methods used in studying a natural system. He argues that when we design an AI program and note changes in its “behaviour” when we change the design, we are essentially carrying out an experiment. Furthermore, much of our knowledge about Artificial Intelligence has been gathered in this way. This leads him to echo his earlier speech with Newel, and conclude that Artificial Intelligence is absolutely an empirical science.

Kukla (1989) completely and utterly disagrees. He argues that Artificial Intelligence is not an empirical science because we know, or theoretically know, what a program will do before it does it, because “we wrote it!” (Walmsley, 2012). If we’re not sure about the actual resulting actions of the program, it’s not because we can’t know (Kukla, 1989). We might be limited by our memory or attention or programming knowledge. But let’s say someone with a particularly good memory and excellent programming skills comes along and takes a look at the program, they might be able to figure it out. Therefore it cannot have been an empirical undertaking, because the result will simply be what logic has already concluded. Kukla refutes Simon’s idea that the trial and error nature of Artificial Intelligence is experimental in approach, and therefore AI is empirical, by saying that the same happens in mathematics. A mathematician searching for a proof will try numerous methods until they find one that works, but no-one would call this empirical testing. He admits that the end goal in AI is less clear than in mathematics, but maintains that this does not undermine its status as an a priori science.

Just as a pen and paper are merely tools in the study of mathematics, computers are tools (albeit very important tools) in the study of Artificial Intelligence, and do not constitute experiments (Kukla, 1989). Kukla believes they are “incidental” to the study of AI, and their role is merely to speed things up a bit, so to speak. Dennett (1978) agrees that in principle,
computers are irrelevant to the study of Artificial Intelligence, in the sense that blackboards are irrelevant to the teaching of geometry. From a practical point of view, they are almost but and this is crucial, not quite, essential. He sees AI programs as thought experiments in material form, not empirical experiments.

To bolster their argument, Newel and Simon compare Artificial Intelligence to the study of information processing in psychology, whereby experiments are carried out to investigate human behaviour at certain tasks, and from the results of these experiments, a theoretical model is constructed. They believe that in a similar way, observations of human activity lead to hypotheses about how they are completing the task and inform the writing of computer programs which attempt to simulate the activity. An example of this is Newel and Simon’s (1961) General Problem Solver (GPS). They created the GPS program with the intention that it would function as a universal problem solving program. In theory, any formalised symbolic problem can be solved by the GPS. According to its creators, the basic principles of GPS were sourced from detailed analysis of transcriptions of human subjects completing the task while vocalizing their thought processes. They claim that this parallel with research psychology characterizes the empirical nature of AI (Newell and Simon, 1976).

But once again, Kukla disagrees. He argues that psychological research differs completely from research in AI. He holds that in psychology, one starts with an observation or some available data about human performance in a particular area, comes up with a theory which accounts for the structures underlying the activity, establishes a new hypothesis which tests the theory, and designs and completes an experiment which either supports or contradicts the hypothesis. But in AI, the underlying structures are already known (Kukla, 1994). It would be like doing tests on the brain to find out how it worked, if we’d already made it ourselves!

Simon (1995) claims that we can learn new information from Artificial Intelligence systems, in spite of Kukla’s objection that we have already designed the system and so it is a priori, by saying that the natural laws operate on complex systems the same way they do on natural systems, and because we are greatly ignorant of natural law and its effects on complex artificial systems, there is room for surprise as it were. He adds that by trying to deduce complex systems from their first principles (for example, just looking at the program as opposed to actually running it) we capture merely a glimpse of the richness and complexity (Simon, 1995).
But I think he is missing the point. Kukla doesn’t say that Artificial Intelligence can’t be compatible with empirical research. He argues that even though empirical methods are possible in AI, perhaps even vital in a pragmatic as opposed to philosophical sense, the distinction between a priori and empirical is asymmetrical (2001). A scientific problem is empirical if it can only be solved by empirical methods, but a problem which can be solved by a priori methods of logic and deduction, remains a priori whether an empirical solution exists or not.

Buchanan (1994) argues that even if we concede that the output produced by a program is completely determined by the input we bestowed it with, we are still left with the interesting question of whether the output could be considered intelligent, and if so to which elements of the system do we attribute the intelligence? This is not a question we can answer by a priori means (or so he claims). It leads one to wonder whether Kukla and our two bastions of empiricism, Newell and Simon, are really looking at the same thing. In principle at least, it seems Kukla is correct about AI being an a priori science, if we consider purely the relationship between the program and its behavioural output. However, if we look at the entire AI process, from empirical testing with humans, forming hypotheses, testing these hypotheses with specifically built programs, evaluating the results, rejigging the hypotheses, the test or the program, running it again, evaluating the results and so on (Buchanan, 1994) then it seems unlikely that a priori methods could replace empirical ones, especially if our aim is to evaluate whether the programs output is intelligent or not.

I am inclined to agree with Schank (1990), who views AI as an array of enterprises, some of which are a priori and some of which are empirical in nature. Depending on our aims, a priori methods may be more suitable than empirical ones, and vice versa. However, Kukla (1994) emphasises that it doesn’t matter whether in practice empirical methods are more suitable, if it can or could be achieved using a priori means, then it is a priori. Be that as it may, what are the implications, one way or another? What does it matter whether AI as a whole, or individual elements of it, is an a priori or empirical science? Wilks (1990) argues that it is matter of description. He states that philosophical discussion about the epistemic nature of AI is not intended to condemn empirical activity within AI, but that there should be consistency and transparency in how work in the field is described. Part of that is deciding
what is empirical and what is not. Kukla asserts that there are two grounds upon which an activity could be seen as empirical;

A) A project which attempts to establish contingent truths, not necessary truths, could be seen to be empirical. The distinction between necessary and contingent truths dates back to Leibniz. Necessary truths (or truths of reason) are truths which are unconditionally true in all possible worlds and whose denial implies a contradiction (Shand, 1993). Take a square, a four sided object with sides of equal length. A square is a square no matter what imaginary world you put it in, and if it’s not, then it’s not a square! Whereas to deny a contingent truth (a truth of fact), does not imply this kind of contradiction. For instance, to deny that “All old men have beards”, is not to deny the existence of old men or of beards, merely that all old men have beards. We can undertake empirical research to see if all old men do indeed have beards, but we cannot empirically investigate whether squares are squares. We could measure the sides a four sided object to check if it were square, but this is then a contingent fact because the necessary fact remains that a square is a four sided object with sides of equal length.

B) The second ground upon which a project may be seen as empirical, is the methods used. A posteriori methods such as observation or experimentation are the hallmarks of empirical research, as opposed to the previously mentioned a priori methods of logic and deduction.

However, Kukla (1994) says we have a problem in that a project may be seen as empirical on just one of these grounds, and not the other. Kukla cites running a program as an ideal example. The procedure may be empirical if we must run the program and observe what happens, to evaluate the result, but the result will still be a necessary truth because the program was written to perform in that way, and if it wasn’t written that way, then it wouldn’t have performed that way! I believe this comes back to the idea that Kukla and his supporters, and Newel and Simon and their followers, are essentially looking at different sides of the same coin. In psychology, my usual field of study, we have a concept known as the Dodo Bird Verdict (that all psychotherapies produce relatively similar outcomes), which I think is quite fitting here. In ‘Alice and Wonderland’ the Dodo Bird announces after the race that “Everybody has won and all must have prizes” (Carroll, 1865). I believe in the case of Newel and Simon versus Kukla on the matter of AI’s epistemic nature, both are correct and may all revel in their prize of rightness.
References


Walmsley, 2012. *In press*