**Distribution Agreement**



Robert Raymond Lusardi II April 11, 2017

Rational Machinery: Can Computers have Aristotelian Intellectual Virtues?

by

Robbie Lusardi

Mark Risjord
Adviser

Department of Philosophy

Mark Risjord

Adviser

Marta Jimenez

Committee Member

Elena Antoniadou

Committee Member

2017

Rational Machinery: Can Computers have Aristotelian Intellectual Virtues?

By

Robbie Lusardi

Mark Risjord

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Philosophy

2017

Abstract

Rational Machinery: Can Computers have Aristotelian Intellectual Virtues?
By Robbie Lusardi

This thesis aims to discuss intelligent machinery in a different way than that of traditional works in the Philosophy of Artificial Intelligence (AI). In this thesis, I explore whether computers currently possess the five Aristotelian intellectual virtues, *episteme*, *techne*, *nous*, *sophia*, and *phronesis*. In cases where computers do not meet Aristotle's criteria for possession of a specific intellectual virtue, I instead discuss what computers would need to accomplish to possess it. When discussing Aristotle, I focus on his analysis of the intellectual virtues in the *Nicomachean Ethics*, *Metaphysics* and *Posterior Analytics*, and incorporate some of *De Anima* as well. When discussing AI, I draw from past debates in history AI as well as modern developments in the field.

The thesis concludes with a discussion of future questions regarding genuinely intelligent machinery and recommends areas for further research.

# Rational Machinery: Can Computers have Aristotelian Intellectual Virtues?

By

Robbie Lusardi

Mark Risjord

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Philosophy

2017

Table of Contents

Section I: Introduction.............................. 1

Section II: Setting the Stage...................... 4

Section III: *Episteme*.................................. 8

Section IV: *Techne*..................................... 13

Section V: *Nous*......................................... 19

Section VI: *Sophia* & *Phronesis*.................. 28

Section VII: Conclusion.............................. 36

Section VIII: Works Cited........................... 38

## I. Introduction

Machines take me by surprise with great frequency.

– Alan Turing, *Computing Machinery and Intelligence*, 1950

In his seminal 1950 paper "Computing Machinery and Intelligence," Computer Scientist Alan Turing theorized about under what conditions a machine can be said to be thinking, specifically aiming to answer the question "can machines think?" Turing admitted this question was too broad to answer adequately, and it its stead, proposed an imitation game as a litmus test for genuine machine intelligence. He argued if a machine could win at his game, now colloquially known as passing the Turing Test, it could be said to have the same mental capacity as a human.[1]

The Turing Test has proven highly controversial – I will discuss some specific objections to it in sections V and VII of this paper – but regardless, Turing's 1950 paper serves as a milestone for the field of Philosophy of Artificial Intelligence (AI). Turing, in his paper, claims that attempting to answer the question "can machines think?" directly is too Herculean a task, a sentiment I agree with. So, much like Turing, I will sidestep that question and instead replace it with a similar discussion, specifically whether, from an Aristotelian perspective, machines can be said to be rational beings, or at least perform competently some of the functions that Aristotle attributes to reason.

The primary aim of Aristotle's *Nicomachean Ethics* is to discuss the human *telos* (end, goal, or purpose). This *telos* as Aristotle understands it is partly the activity of the best part of our souls, or, in modern terms, the proper actualization of what it is to be human. Reason for

[1] Turing, Alan, "Computing Machinery and Intelligence." 1950, *Mind 49:* 433-460.

Aristotle is the defining and central aspect of our humanity (i.e. reason is what separates us from animals or plants) and so, embedded in his discussion of the human *telos* is a discussion of human rationality. Specifically, Book VI of the *Nicomachean Ethics* explains the different functions and parts of human reason, which when used properly make us good humans. So, in my discussion of potentially rational machinery, I will primarily use Aristotle's analysis of the different aspects of rationality from *Nicomachean Ethics* VI. I will also incorporate some of Aristotle's other works, specifically the *Posterior Analytics*, for Aristotle's discussion of *episteme*, *De Anima*, for Aristotle's discussion of the soul, and the *Metaphysics*, for Aristotle's discussion of *techne* and *sophia*.

Obviously, when composing the *Ethics* or any of his other works, Aristotle did not have computers in mind; yet, he deals with many questions which map surprisingly well onto both classic and modern debates and research in artificial intelligence. Some debates which I will touch on in this paper are the distinction of semantic understanding versus syntactic understanding, as discussed by John Searle in his Chinese Room thought experiment, and machine emotionality, as presented by Geoffrey Jefferson. Some current research in the field I will discuss is the development of machine intuition, the potential for abstract reasoning in machines, and the potential for emotions in machines.

As Aristotle's discussion of human rationality parallels the history of the Philosophy of AI and modern developments in AI, I will use it to develop a response about the possibility of genuinely rational machinery. To do this, I will discuss whether various forms of AI can or could possibly satisfy Aristotle's criteria for rationality, specifically if they can possess Aristotelian intellectual virtues. Further, if AI fails in this regard, I will discuss both why this failure occurred and how AI could be adapted to meet the Aristotelian definition of a rational being. Hopefully, this paper can provide an alternative framework for the discussion of genuinely intelligent machinery and can further the discussion around it.

## II. Setting the Stage

In Book VI of the *Nicomachean Ethics*, Aristotle separates the virtues of the soul into two categories, "virtues of character and [virtues] of thought,"[2] and begins his consideration of the latter. Aristotle specifies five distinct virtues of thought, also known as intellectual virtues – *episteme techne*, *nous*, *sophia*, *phronesis* (I will discuss them in this order) – which encompass

> a wide range of cognitive states that enable an agent to reason correctly and to judge truly, sometimes about what is the case, sometimes about best to make something, sometimes about how best to act.[3]

The virtues of thought have a wide range of functions. Some are logical or pragmatic, while others are used to determine moral virtues, or virtues of character, such as temperance, courage, or justice. Further:

> Virtues [of thought] appear to be states of character relative to the kinds of activities in which we engage and to the kinds of states that characterize us when we deliberate or reason or plan or develop theories — when we conduct ourselves, in other words, as thinking agents.[4]

In brief, these five virtues of thought are the states that allow the human capacity for rationality, the rational soul, to flourish or operate in the best way possible. An individual who occupies these five states simultaneously is using their rationality virtuously, that is to its highest capacity. The five virtues of thought, in brief, are as follows:

*Episteme*, most frequently translated as "scientific knowledge," is a theoretical and deductive knowledge of relations between principles and universals, or of basic facts and greater scientific necessary truths. Further, *episteme* is knowledge that is transferable from or

---

[2] Aristotle. *Nicomachean Ethics*. Trans. Terence Irwin. Indianapolis: Hackett Publishing, 1999. Print. p. 86, 1139a 2
[3] Kosman, Aryeh. "Aristotle on the Virtues of Thought." *Virtues of Thought*, Harvard University Press, 2014, pp. 281, www.jstor.org/stable/j.ctt6wpmjx.18.
[4] Ibid.

demonstrable by one to another, say, through a book or lecture, and is knowledge that pertains to a specific area. Aristotle often uses geometric proof as an example of *episteme* – one begins with axioms and then, through valid reasoning, can connect or relate them to reach greater conclusions.

*Techne*, commonly translated as "craft skill," "craftsmanship," or "art," is the practical application of reason primarily concerned with production of physical objects such as chairs, sculptures, or buildings. In short, *techne* is the knowledge of how to make, shape, or transform things. Idiomatically, it is "know-how."

*Nous*, "intellect" or "mind," much like the modern definition, is a type of intuition or perception, specifically an understanding of things that cannot be known explicitly or demonstrated, such as the first principles that are the basis for *episteme*. More accessibly, *nous* is a type of intelligent insight.[5] A seasoned Chess grandmaster, for example, would possess considerable *nous*, as although when playing, he cannot deduce every single possible outcome of every possible move, he could, through his intuition, still play the game of Chess well.

*Sophia*, translated as "wisdom," is an ethereal hybrid of *nous* and *episteme* concerned with the "scientific knowledge and understanding about the things that are by nature most honorable."[6] *Sophia*, a theoretical form of wisdom, is concerned with universals and higher knowledge often outside the realm of human comprehension. *Sophia* is not intrinsic, but rather must be acquired through painstaking study.

*Phronesis*, often translated today to "practical wisdom" or "prudence," is a disposition concerned with virtuous action. A *phronimos*, one who possesses *phronesis*, has, through a

[5] David D. Corey, "Vogelin and Aristotle on Nous: What is Noetic Political Science?." *The Review of Politics*, Vol. 64, No. 1, 2002, pp. 57-79
[6] Aristotle. *Nicomachean Ethics*. Trans. Terence Irwin. Indianapolis: Hackett Publishing, 1999. Print. p. 91, 1141b 3

lifetime of experience and reflection, come to understand all virtues of character and will always act accordingly. For example, one with *phronesis* would exhibit the appropriate degree of courage when the situation calls for it (he would be neither cowardly nor reckless), or would show a sufficient level of generosity to those in need. For one with *phronesis,* virtuous actions are habit[7] – that is, rather than carefully deliberating about what course of action to take, the *phronimos* has learned to always do the right thing to the right degree instinctually. Further, the *phronimos* gains a sort of pleasure from acting virtuously, although importantly, he does not act virtuously to gain this sense of pleasure, he acts virtuously because it is the right thing to do.[8]

In combination, *episteme, techne, nous, sophia,* and *phronesis* encompass the entire ability of rationality and form the rational soul.[9] The capacity for these virtues is unique to rational agents – plants and animals, for example, cannot possess any of these virtues of thought – thus meaning rationality, according to Aristotle, is a uniquely human property. In *De Anima*, however, Aristotle argues that the rational soul exists in a nested hierarchy within the other types of souls,[10] the nutritive soul, which belongs to all living things and is responsible for life and growth, and the sensitive soul, which belongs to animals and humans and is responsible for things such as locomotion and sensations. Further, Aristotle argues that, as the rational soul exists in this nested hierarchy, it is inseparable from the other two souls[11] (some evidence suggests that *nous* might be separable,[12] but this is highly controversial and so is a discussion outside the scope of this paper) and cannot exist in a vacuum. This entails, trivially, that

---

[7] Demos, Raphael. "Some Remarks on Aristotle's Doctrine of Practical Reason." *Philosophy and Phenomenological Research*, vol. 22, no. 2, 1961, pp. 153–162., www.jstor.org/stable/2104836.
[8] Ibid.
[9] Aristotle. *Nicomachean Ethics*. Trans. Terence Irwin. Indianapolis: Hackett Publishing, 1999. Print. p. 86, 1139a 5
[10] Miller, Fred. "Aristotle's Philosophy of Soul." *The Review of Metaphysics*, Vol. 53, No. 2, 1999. pp. 309-337 http://www.jstor.org/stable/20131355
[11] Aristotle. *De Anima*. Trans. Hugh Lawson-Tancred. New York: Penguin Books, 1986. Print. p. 158, 413a 3
[12] Miller, Fred. "Aristotle on the Separability of Mind." *The Oxford Handbook of Aristotle*. August 2012.

machines cannot have a rational soul and so cannot be rational beings as they are not alive. Clearly, if taken at face value, this claim would invalidate the entirety of this paper. But, as this paper is designed to engender discussion for the rational capabilities of AI, rather than be a literal analysis of Aristotle, I feel I can ignore this issue. I will leave the validity of this move up to the reader.

## III. Episteme

*Episteme*, although usually translated as "scientific knowledge," is specific type of theoretical knowledge distinct from a contemporary understanding of science. The modern scientific inductive method, which uses trial and error experiments to support or reject hypotheses, did not exist in Aristotle's time. *Episteme*, rather, is a deductive, rather than inductive, knowledge of how basic premises and axioms lead to necessary but demonstrable truths. Accordingly, one possesses *episteme* if

> one knows the appropriate explanation of [something] and knows that the [some]thing cannot be otherwise... To scientifically know [in the sense of *episteme*] that P is to demonstrate P which amounts to explaining P.[13]

A vehicle through which to understand *episteme* is geometry. To know geometry in the sense of *episteme* is to be able to demonstrate convincingly various formal geometric proofs, for example the proof of the Pythagorean theorem, from basic known geometric axioms. One with *episteme* in this instance knows and can demonstrate deductively that the conclusion of the Pythagorean theorem, the square of the hypotenuse equals the square of the other two sides, is undoubtedly true; for, starting from indisputable premises (e.g. basic properties of various shapes), he knows and can demonstrate the truth of each individual step in the proof. Further, for possession of *episteme*, one does not need to know the premises themselves, as by definition, they are not demonstrable but rather are invariably true.[14] *Nous*, which I will discuss in a later section, is the virtue responsible for this apprehension of initial premises.

---

[13] Aydede, Murat. "Aristotle on *Episteme* and *Nous*: The *Posterior Analytics*." Southern Journal of Philosophy (1998), Vol. 36, No. 1, Blackwell Publishing Ltd, pp. 15–46. http://faculty.arts.ubc.ca/maydede/Aristotle.pdf
[14] Ibid.

Accordingly, for a machine to possess *episteme*, it would have to use deductive reason in a similar manner to the way Aristotle describes and would have to be able to demonstrate its use of this reason. Conveniently, any computer functions using exclusively deductive logic, thus almost trivially satisfying the first requirement of *episteme*. Early predecessors to computers, such as the theoretical Turing machine used to solve the *entscheidungsproblem* (The Decision Problem), were purely deductive symbolic logic machines.[15] Even today's computer programs are nothing more than logical syllogisms[16] – a programmer specifies rules in lines of code, and the computer follows these rules precisely. Take, for example, Aristotle's famous syllogism:

> Socrates is a man
> All men are mortal
> Therefore, Socrates is mortal

Deductively, this argument is valid. Starting from the basic premises, Socrates is a man, all men are mortal, the argument always leads to a necessary conclusion: Socrates is mortal.

Computer programs work in the same way as this deduction does. In computer programming language, the above logical syllogism would be expressed as follows:

> "Socrates" = "Man"
> "Man" = "mortal"
> IF Socrates THEN mortal

Clearly, the two syllogisms function in the same manner – the second syllogism is an arguably clearer expression of the way deductive logic works – and aside from some small syntactic changes do not differ in any way. A computer began with the same basic premises, "Socrates" = "Man," "Man" = "mortal," and has reached the same necessary conclusion: Socrates is mortal. All working programs function in this exact fashion. All programs begin with basic premises,

---

[15] Oaksford, Mike. *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford University Press: Oxford, 2007. p. 49
[16] Ibid. p. 48

build on these premises using lines of code, and eventually reach a necessary conclusion. Computers follow these lines of code deductively and rigidly. Further, no working program is deductively invalid (i.e. the premises do not logically entail the conclusion), for if it were, the program would not compile, meaning a computer could not run it. So, with some exceptions (computers/programs that can reason inductively, which I will discuss later) all computers must always use deductively valid logic. Trivially, by being written lines of code, it seems programs satisfy the demonstrability requirement for *episteme* as well.

Although I will not delve further into Aristotelian logic, deductively reasoning along these lines – using logical deductions to move from basic premises (Socrates is a man, men are mortal) to a greater conclusion that follows necessarily (Socrates is mortal) – is a clear example of *episteme*. Every computer is essentially a combination of thousands of these deductive syllogisms, in the form of programs, operating simultaneously. Accordingly, by being literal deductive inference machines, it seems that all computers trivially possess *episteme*.

If the above argument does not satisfy the skeptic, take, for a specific example, the first computer to solve a mathematical proof, that of the four-color map theorem. The theorem states that any two-dimensional map can be, using four colors, colored in a way such that all adjacent regions in the map do not share the same color (e.g. no blue area touches another blue area). This, I argue, aligns exactly with Aristotle's frequent discussion of geometric proof, which he uses as an example of *episteme*.

The eventual method to prove this theorem – a proof by contradiction using a technique called minimal counter examples[17] – is intuitively simple. A minimal counter example, also

[17] Gonthier, Georges. "Formal Proof – The Four-Color Theorem." Georges Gonthier. Notices of the American Mathematical Society. 2008 Vol. 55. p. 1382–1393

called a minimal criminal, is the smallest possible example that shows the proof to be false. In this instance, a minimal counter example would be a map that requires at least five separate colors, but if one country were removed, could be colored with only four colors. So, to prove the validity of the theorem, one needs only to show that no minimal counter examples exist. To do so, however, would require one to devise and check a near infinite number of different configurations of maps, a task unaccomplishable by a human. Accordingly, although it was first proposed in 1852, the problem remained unsolved until 1976.

The proof of this theorem was undoable by hand. So, University of Illinois mathematicians Kenneth Appel and Wolfgang Haken used a computer. This computer checked approximately two thousand reducible configurations – an arrangement of regions that cannot occur in a minimal counter example – one by one to show that a minimal counter example to the four-color theorem could not exist, thus through brute force proving the theorem by contradiction. Although this proof, as the first proof completed by a computer, was highly scrutinized at the time of its completion, it has withstood any challenges since.[18] Today, computerized proofs of mathematical theorems are both widely accepted and routinely utilized.[19]

This computer clearly meets a stricter criteria for possession of *episteme*. In addition to running via the use deductive syllogisms in the form of programs, it solves mathematical proofs in a way indicative of *episteme*. Much as with Aristotle's geometer, the four-color theorem computer began from basic indisputable first premises, properties of various shapes and their boundaries, and managed to prove deductively a greater conclusion, in this instance the four-color theorem. Further, the computer's proof is demonstrable as it could, in theory, teach an eager student the proof line by line by breaking down each of the two thousand cases it checked.

---

[18] Ibid.
[19] Ibid.

In practice, due to the difficulty of checking each case – it took the computer one thousand hours in total – it may not be feasible, but this is a limit of the pupil, not the computer, and so does not seem to matter for the purposes of *episteme*. This result seems generalizable to any computer that can solve mathematical proofs.

## IV. Techne

*Techne* is a highly pragmatic intellectual virtue concerned with ends. Specifically, *techne* is the skill for "making something into something it is not."[20] The purpose of, say, a craftsman is to make timber into sturdy and elegant pieces of furniture. Likewise, the purpose of a doctor is to make sick patients well, via diagnosing diseases correctly and performing the correct treatment consistently. Accordingly, the value of these positions, and for that matter, all activities requiring *techne*, is primarily generated by their end.[21] Despite the end being the primary concern of *techne*, the process by which this end is generated matters as well. As Aristotle writes in the *Metaphysics*:

> we think… that the master-workers in each craft are more honourable and know in a truer sense and are wiser than the manual workers, because they know the causes of the things that are done (we think the manual workers are like certain lifeless things which act indeed, but act without knowing what they do, as fire burns,—but while the lifeless things perform each of their functions by a natural tendency, the labourers perform them through habit)[22]

So, *techne* seems to be composed of two parts, the first being a purely mechanical element, and the second being a deeper theoretical knowledge. Returning to the example of the craftsman, a manual laborer, who satisfies only the first mechanical aspect of *techne*, can build a sturdy chair. This craftsman, however, does not satisfy the second knowledge aspect of *techne*, as he does not understand the qualities that make the chair sturdy or elegant (qualities such as the grain and fineness of the wood or the intricacies of the joints). Accordingly, although the manual craftsman possesses *techne*, as being able to assemble a chair clearly requires some craft skill, it is in a very

---

[20] Meagher, Robert. "Technê." *Perspecta*, vol. 24, 1988, pp. 159–164., www.jstor.org/stable/1567132.
[21] Ibid.
[22] Aristotle. *Metaphysics*. W.D. Ross, trans. Oxford: Clarendon Press, 1908. Reprint. Stillwell, KS: Digireads 2009 p. 3

limited way. Only the master craftsman, who understands both the mechanical aspect and the theoretical aspect, possess *techne* in the fullest and most virtuous sense. So, for a machine to satisfy *techne* in the mechanical and limited sense, it would only have to be able to manufacture some product from raw parts as the manual laborer does – understanding is not necessary. For a machine to possess *techne* in the fullest sense, however, understanding of the causes (why a chair is study or elegant, or why a patient is sick) becomes mandatory.

Conveniently, many machines today perform tasks that satisfy the mechanical component of *techne* incredibly capably, thus easily possessing it in the more limited sense. Take, for example, the 3D printer. In the 3D printing process, also known as additive manufacturing, a computer, via following a digital blueprint, stacks layers of material, thus building an object level by level. This process is similar to the steps used to build a pyramid: first, the bottom layer is set. Next, a second layer is placed on top of the bottom layer. Then, a third layer is placed on top of the second layer, and so on until the top layer has been completed. A 3D printer, however, is not limited to producing only pyramid shaped objects. For example, recent items that have been made through the additive manufacturing process include functional guns,[23] replica human skulls,[24] and even hamburger meat.[25] Much as with the manual laborer, a 3D printer, produces a well-made, working version of what it has been tasked to build. Likewise, as the laborer converts timber into a chair, a 3D printer converts something – metal, plastic, or in the case of the hamburger meat, a culture of live cells – into something it is not. Clearly, however, a 3D printer,

---

[23] Bilton, Nick. "The Rise of 3-D Printed Guns." *The New York Times*. August 13th, 2014. https://nyti.ms/2nD3bye. Accessed April 3rd, 2017

[24] "Surgeons perform 'world's first' implant of entire 3D printed plastic skull dome." *The Russian Times*. March 28th, 2014. Accessed April 3rd, 2017

[25] Fountain, Henry. "A Lab-Grown Burger Gets a Taste Test." *The New York Times*. August 15th, 2013. https://nyti.ms/2m8xy2u. Accessed April 3rd, 2017.

like the manual laborer, does not understand the underlying causes of what it is producing. So, like the manual laborer, a 3D printer only possesses *techne* in a limited sense.

*Techne* in the second, fullest sense undoubtedly sets a much higher standard that machines must satisfy. Manufacturing machines today, such as the 3D printer, do not need any sort of understanding to perform their roles well; accordingly, manufacturing machines function purely mechanically (although there are some recent developments in this area).[26] It seems that today's machines could not produce a chair in the way the master craftsman does, namely with understanding, because economically there has been no need for them to do so.

Medicine, however, another profession which Aristotle argues requires significant *techne*, may be more promising on this front for machines. A master doctor with *techne*, for example, would possess considerable skill at diagnosing and treating diseases, in addition to understanding the cause of these diseases and how best to treat them. In Aristotelian terms, a master doctor possesses *techne* as he can turn a patient's sickness into health, and can understand the causes of why the patient is sick. So, I argue, if a computer could diagnose and treat diseases in a similar fashion, with a sort of understanding of the causes and functions of these diseases, it could be said to have *techne* in the strongest sense.

A machine that could satisfy this second, more stringent requirement for *techne* is IBM's supercomputer Watson. Most famous for soundly beating two ex-champions at the trivia game show *Jeopardy!* in 2011, Watson today is used in a few specialized hospitals in the United States as a digital doctor, and aims to diagnose patients and recommend personalized treatments for patients with various rare diseases. In effect, Watson is a medicine-specific hybrid of search

[26] Thorsten Wuest, Daniel Weimer, Christopher Irgens & Klaus-Dieter Thoben. "Machine learning in manufacturing: advantages, challenges, and applications." *Production & Manufacturing Research*. 2016. pp. 23-45. http://dx.doi.org/10.1080/21693277.2016.1192517

engine and encyclopedia – doctors input a patient's symptoms into Watson and the supercomputer trawls its internal database containing millions of medical records to determine what the disease is and how to go about treating it.[27] Watson's methods, made possible through machine learning methods (which I will explain in more detail in the subsequent section), have proven incredibly effective in both providing diagnoses and recommending appropriate treatment. In one instance, doctors at the University of Tokyo were unable to treat a woman with cancer for months. Watson found both the correct diagnosis, a rare form of leukemia, and the most effective form of treatment in under ten minutes.[28] Clearly, Watson's immense skill at diagnosing and recommending treatments for patients, turning sickness into health, seems to indicate it at the very least has possession of the mechanical aspect of *techne*.

I argue that unlike the 3D printer, however, Watson, in its process of diagnosing diseases, can be said to have an understanding of what these diseases are, thus also satisfying the second, stronger component of *techne*. Watson operates using a sophisticated form of natural language processing, called DeepQA, which analyzes natural language and allows computer systems to "deeply analyze the breadth of relevant content to more precisely answer and justify answers to user's natural language questions."[29] Perhaps the best example of DeepQA in action is Watson's performance on *Jeopardy!*, which propelled the AI into the national spotlight. *Jeopardy!* is a popular quiz show which, uniquely, gives its contestants an answer and asks them to formulate a question in response (e.g. A sample answer would be "In 1864, General Sherman and his Confederates burned this city to the ground," and the correct response by the contestant would be

---

[27] IBMWatsonSolutions. "IBM Watson: How It Works." *YouTube*. YouTube, 07 Oct. 2014. Web. 04 Apr. 2017.
[28] Otake, Tomoko. "IBM big data used for rapid diagnosis of rare leukemia case in Japan." *Japan Times*. August 11th, 2016. http://www.japantimes.co.jp/news/2016/08/11/national/science-health/ibm-big-data-used-for-rapid-diagnosis-of-rare-leukemia-case-in-japan/#.WOM0ThLytE4. Accessed April 3rd, 2017
[29] Ferruci, David et al. "Building Watson: an Overview of the DeepQA Project" *AI Magazine*, 2010. Vol. 31. pp. 59-79

"What is Atlanta?"). Further, the show often poses short questions involving complicated word play, such as puns, riddles, or rhymes. Accordingly, a computer's winning in this show would require the capability to:

> answer… rich natural language questions over a very broad domain of topics… A computer system that could compete at human champion levels at this game would need to produce exact answers to often complex natural language questions with high precision and speed and have a reliable confidence in its answers, such that it could answer roughly 70 percent of the questions asked with greater than 80 percent precision in 3 seconds or less.[30]

Clearly, despite most of its questions being fact based, *Jeopardy*!'s complex structure, tricky use of natural language, and fast paced nature, means a pure search engine, although a promising start, is inadequate to play the game well. Watson's DeepQA (QA fittingly stands for Question & Answer) methods bridges this gap. DeepQA is a complex series of algorithms (which I will not discuss in detail, as I feel their workings are not relevant to my discussion of *techne*) which break down these natural language questions into searchable forms using probabilistic techniques. Once this searchable form is achieved, Watson harnesses its processing power to access its encyclopedic memory banks of data and selects what it deems to be the most likely answer. This method proved incredibly successful, as in 2011 Watson managed to best the two best human contestants in the game show's history by a significant margin.

Although today Watson, repurposed for the medical community, uses a slightly different form of DeepQA, the workings of the current DeepQA do not seem to differ in any significant way from the workings of the DeepQA used in *Jeopardy!*. Watson is first given a list of the patient's symptoms expressed in natural language by doctors and then proceeds to break down this list of symptoms into searchable form. With this information, Watson then trawls its vast

[30] Ibid.

databank of hundreds of millions of medical records and papers to determine what the patient's most likely condition is. Further, the symptoms need not always be expressed in natural language; Watson, in certain instances, can diagnose patients using images of their symptoms as well.[31]

The way Watson operates seems to be indicative of it having *techne* in the second stricter sense. Unlike the manual laborer or mechanical 3D printer, Watson is not merely following some blueprint provided by others, as clearly this would be inadequate to play *Jeopardy!* well or diagnose diseases well. Rather, it seems Watson functions in a manner reminiscent of the way a human does, albeit in a very robotic way and with some extra luxuries (namely heaps of data and processing power). Watson first "understands" (I am using this term loosely) the semantic meaning of the quiz question (in the case of *Jeopardy!*) or a list or image of symptoms (in the case of medicine) using probabilistic methods and heuristics. Then, Watson searches its current stores of information to reach a conclusion.

Watson seems to operate in a parallel manner to how a master doctor would operate: first, the doctor would need to understand and interpret the symptoms, and second, the doctor would, based on past experiences and current knowledge, reach a diagnosis in a probabilistic way. As evidenced by its aforementioned rapid diagnosis at the University of Tokyo, Watson in many instances is more skilled at its craft than leading doctors today. Further, Watson, thanks to its sophisticated natural language processing, searching algorithms, and image recognition abilities, does seem to have an understanding of the cause of sicknesses. In brief, Watson performs its craft well and in a manner indicative of knowledge of the craft. Accordingly, it seems Watson satisfies both *techne* in both senses.

[31] IBM. *Watson Health Medical Imaging Collaborative Expands to 24 Members; IBM Debuts Watson Imaging Clinical Review, the First Cognitive Imaging Offering*. *IBM*. IBM, 20 Feb. 2017. Web. 3 Apr. 2017.

## V. Nous

*Nous*, unlike *episteme*, requires a non-deductive form of reasoning.[32] *Nous*, in the strictest sense, is an intuition regarding the first principles that are the basis of *episteme*.[33] These axiomatic principles are unprovable and undeducible, and thus cannot be known via *episteme*, yet are invariably true. Thus *nous*, allowing for apprehension of these principles, is a more ethereal intellectual virtue than the formal and deductive *episteme,* and mirrors what is today known as intuition.[34] Accordingly, although most types of computers seem to satisfy requirements for possession for *episteme*, the way in which most computers think is inadequate for intuition and thus possession of *nous*. John Searle's famous Chinese Room thought experiment, originally presented in his 1980 essay *Minds, Brains*, and *Programs* as a response to the Turing test, demonstrates the reason for this inadequacy.

Searle's thought experiment is as follows: suppose there is a man locked in a room who has initially been given a piece of paper containing a large text of Chinese characters forming a question, and is asked to write down a response in Chinese. Save for a slit in the wall, through which the man receives the characters he must respond to and through which the man must pass his answers, he is isolated from the outside world. Further, the man speaks only English, and does not understand Chinese in the slightest – he only sees the Chinese characters as drawings or combinations of squiggles and cannot assign any meaning to the characters. To aid him with his translation, he is given a tome which has a comprehensive set of rules, written in English, which describes, purely syntactically, how to respond. For example, one rule might say, "if you see this

[32] Aydede, Murat. "Aristotle on *Episteme* and *Nous*: The *Posterior Analytics*." Southern Journal of Philosophy (1998), Vol. 36, No. 1, Blackwell Publishing Ltd, pp. 15–46. http://faculty.arts.ubc.ca/maydede/Aristotle.pdf
[33] Ibid.
[34] Ibid.

*squiggle-squaggle* followed by a *squaggle-squiggle*, write down a *squaggle-squaggle* in response." By systematically following his rulebook, the man can respond to any sequence of Chinese characters in the exact way a native Chinese speaker would. As Searle says, thanks to the tome,

> [The man's] answers to the questions are absolutely indistinguishable from those of native Chinese speakers. Nobody just looking at [the] answers can tell that [the man doesn't] speak a word of Chinese.[35]

So, the man, via strictly following his rule set, can answer any given question in perfect Chinese; however, Searle's key claim is that, from this process, the man does not gain any semantic understanding of Chinese whatsoever. The tome has not taught the man what the meaning of the Chinese characters are, but rather has only taught him how to manipulate them in the correct way. For example, if the given question were to say "你吃了吗?" ("have you eaten," a common Chinese greeting), the man, by following the tome, would correctly respond with "我吃了" (I've eaten). But, despite this correct response, the man still sees all Chinese characters as collections of squiggles, and does not understand he is having a conversation about eating. So as before, he cannot assign any meaning at all to them. He has syntactic understanding, but no semantic understanding.

So, Searle claims, like the man in the room, a classic computer program, the type which functions only deductively as outlined in my discussion of *episteme*, cannot gain any understanding of the meaning of its code; it too is merely manipulating symbols.[36] Take this very basic two-line Chinese speaking (faux) program:

```
IF "你吃了吗"
THEN print "我吃了"
```

[35] Searle, John R. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (1980): 417-457.
[36] Ibid.

The way this program operates, Searle argues, is exactly analogous to the man in the Chinese room.[37] The program is given an input – a string of Chinese characters – and, via strictly following given rules, in this case its code, produces the correct output of Chinese characters. But much like the man, the computer does not gain any understanding of Chinese through this process, as it is merely following rules. For example, it does not and cannot understand that "吃" means "to eat" or "吗" is a particle which when appended to a sentence signifies it is a yes or no question. From this, it follows that, even if the computer program could answer any Chinese question, despite being able to speak Chinese as a native would, the program still does not understand it. Although learning Chinese is a niche case, this argument for lack of understanding can be extended to any old-school AI program – these programs only manipulate symbols formally, and so acquire no semantic knowledge of what they are manipulating.

Returning to *nous*, recall the example I provided previously of a chess player. Chess is technically a solvable game,[38] that is, in theory, via an *episteme*-like deduction, it is possible to play a strategy that always wins. By checking every possible move and every possible response, and every possible response to that response, *et cetera*, one can deduce this always winning optimal strategy. No human, obviously, has the memory or mental ability to play chess in such a fashion. Accordingly, aside from openings, which are largely formulaic, high level chess players rely on their *nous*, or intuition, rather than brute force calculation both to understand chess and to play chess well. So, I argue, if a computer were to possess *nous* in this chess playing scenario, it would have to play chess both well and in an intuitive way.

---

[37] Ibid.

[38] Schwalbe, Ulrich & Walker, Paul. "Zermelo and the Early History of Game Theory." August, 1997. http://www.math.harvard.edu/~elkies/FS23j.03/zermelo.pdf

An example of a program that plays chess well in an *episteme*-like way but not an intuitive way indicative of *nous* is IBM's Deep Blue computer. In 1997, IBM's supercomputer Deep Blue bested then world champion Garry Kasparov in a game of chess, an incredible accomplishment at the time (today, your average laptop computer's built-in chess software can beat any player on earth). Despite this accomplishment, I claim Deep Blue still cannot be said to have *nous,* as although it clearly could play chess well, the way in which it played was not using anything resembling intuition. When questioned whether Deep Blue played via intuition, or something resembling it, the supercomputer's programmers responded as follows:

> The short answer is "no." Earlier computer designs that tried to mimic human thinking weren't very good at it. No formula exists for intuition.[39]

Instead, Deep Blue, via exploiting massive computational power, played using "brute force aplenty,[40]" calculating 200 million moves per second and selecting from this set the move that maximized its probability of winning (or minimized its probability of losing). Even Deep Blue's seemingly clever moves, which when played by a human would require significant *nous*, such as its famous Nxe6 (Knight to the E6 square) sacrifice against Kasparov, were in fact largely done by brute force. In this case, this knight sacrifice, which forgoes material in exchange for more favorable attacking position later in the game, was pre-programmed by chess grandmasters into Deep Blue rather than thought of by the computer independently.[41] Deep Blue only played this move because this specific sacrifice, which would put it in a commanding position to win, was stored in its list of openings – if left to its own devices, Deep Blue would likely not have even considered the gambit.[42] Simply, intuition was something Deep Blue did not use to play; rather,

---

[39] "IBM Research | Deep Blue | Overview." *IBM Research | Deep Blue | Overview*. N.p., Februrary 23rd, 2001. Accessed April 3rd, 2017. https://www.research.ibm.com/deepblue/meet/html/d.3.3a.shtml
[40] Ibid.
[41] Newborn, Monty. *Deep Blue: an artificial intelligence milestone*. New York: Springer, 2003. p. 198
[42] Ibid.

it played using brute force deduction, something more along the lines of *episteme*. So, Deep Blue could not be said to have an understanding of chess as it was merely robotically crunching numbers and determining moves in a deductive fashion. It follows that Deep Blue, even though it could play chess at a superior level to a grandmaster such as Garry Kasparov, could not play chess in the intuitive way necessary to demonstrate possession of *nous*.

Modern artificial intelligence, however, is more promising on this front. Although computers still do rely on raw power for many tasks, artificial intelligence, thanks to improved programming techniques, can successfully recreate human intuition, thus demonstrating a capacity for *nous*.

The best example of machine intuition is from Google's software AlphaGo, which in 2016 impressively bested 18-time world champion Lee Sedol in the complex east-Asian board game Go. Go requires strong pattern recognition skills, creativity, intuition, and strategic thinking to be played well; accordingly, as these qualities are not replicable through sheer computational brute force, artificial intelligence has until recently been very weak in these areas. Correspondingly, mastering Go has traditionally been seen as an immensely difficult challenge for those in the field of AI, significantly more difficult than mastering even chess. As British mathematician I.J. Good stated:

> In order to programme a computer to play a reasonable game of Go, rather than merely a legal game – it is necessary to formalize the principles of good strategy, or to design a learning programme. The principles are more qualitative and mysterious than in chess, and depend more on judgment. So, I think it will be even more difficult to programme a computer to play a reasonable game of Go than of chess.[43]

[43] Good, I.J. "The Mystery of Go." January 21st, 1965. New Scientist. http://www.chilton-computing.org.uk/acl/literature/reports/p019.htm Accessed April 3rd, 2017

Unlike chess, a computer could not brute-force its way in an *episteme*-like fashion to victory against a strong player – the game of Go is much too abstract and complex for a Deep Blue style computer to succeed. Instead, to succeed in Go, a program must be able to deliberate and play using intuition.

In AlphaGo, this intuition was recreated using machine learning methods. Unlike the old-school AI, which, as the Chinese Room experiment showed, rigidly follows lines of code, current machine learning techniques (I will use modern AI interchangeably) are designed to be adaptable and flexible. Machine learning, in brief, aims to give computers the ability to learn or perform a task, such as play Go, without explicitly being programed to do so.[44] Although there are quite literally thousands of machine learning methods and algorithms, all function in roughly the same way. As Pedro Domingos, a Computer Scientist at the University of Washington succinctly puts it, "[Machine] Learning = Representation + Evaluation + Optimization."[45] First, the data must be represented in a suitable algorithmic form for the program to process. Second, the algorithm is run, and an evaluation of its performance is made. Third, the algorithm is tuned in response to any error – this can either be done by the program itself or a human – and then the process is repeated until acceptable results are achieved.

The primary machine learning method used in AlphaGo is called a neural network.[46] As the name implies, neural networks are artificial small-scale attempts to model the neural structures and processes of a mammalian brain.[47] These neural networks are composed of a series of interconnected artificial neurons and aim to solve problems in the same way a human

---

[44] Munoz, Andres. "Machine Learning and Optimization." 2014. https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf. Accessed April 3rd, 2017

[45] Domingos, Pedro. "A Few Useful Things to Know about Machine Learning." 2012. https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf. Accessed April 3rd, 2017

[46] Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search" N.p., 27 Jan. 2016. Web. 3 Apr. 2017. http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html.

[47] Ibid.

brain would. Neural network methods have had the most success in visual processing, excelling in tasks such as face recognition, image classification, and even simple video game playing.[48] AlphaGo builds on these methods.

A note: the neural network methods implemented in AlphaGo are highly technical, so to that end, when describing them I will be as broad as possible while preserving aspects relevant for my discussion of *nous* and trying to stay faithful to the way these methods work.

AlphaGo's playing method is (roughly) as follows: an image of the current state of the board is given as an input, and this input is then processed through 12 layers of neural networks, each containing millions of artificial neuron connections. Two layers of the neural network are of relevance for this paper. First is the "policy network" layer, which is the neural network responsible for selecting the next move to play. Second is the "value network" layer, which predicts the probability of a win from this given state.[49]

Although, in brief, much like Deep Blue, AlphaGo attempts to play the move which maximizes its estimated probability of winning, unlike Deep Blue, which calculated these probabilities through either brute force simulation of future game states or via having them preprogrammed in, AlphaGo "learned" these probabilities much in the same way a human Go player would learn the optimal move to play – through careful study and practice of the game.

AlphaGo did not start the learning process by playing Go at a beginner level; rather, as the aim of the program was to master Go, AlphaGo first was taught to predict specific moves master level human players would play.[50] To accomplish this task, AlphaGo's policy network, the network that selects which move to play, was trained from a dataset of 30 million moves

---

[48] Ibid.
[49] Ibid.
[50] Ibid.

previously played by expert human players. After being fed this dataset, AlphaGo was able to predict the next move a master level human would play at a 57% level of accuracy, eclipsing the best previous machine effort by 13%.[51] Even a 1% improvement in predictive power represents a massive increase in a program's playing skill; accordingly, a 13% improvement meant that AlphaGo could play Go at a skill level that exponentially dwarfed that of its nearest software competitor.[52]

The next step in AlphaGo's training was more reminiscent of the human method of learning; it began to play the game as much as possible. Unlike a human player, who would require someone else or something else to play against AlphaGo had the luxury of being able to play against itself as well as outside opponents. It did so thousands of times. To hone its play, AlphaGo used a process called reinforcement learning. Reinforcement learning, in brief, is "a paradigm of [machine] learning by trial-and-error, solely from rewards or punishments,"[53] and attempts to model very basic forms of human or animal learning.[54] In brief, proper outcomes (in this case, wins or strong moves) are rewarded whereas bad outcomes (losses or bad moves) are punished. As AlphaGo clearly cannot be rewarded or punished in the same manner as a human, these rewards and punishments for AlphaGo were represented algorithmically. Through this reinforcement learning method, AlphaGo managed to improve its overall skill and implement new strategies by itself (i.e. they were not preprogrammed) with each game.

After completing its training process, after learning to predict moves and playing thousands of different games, AlphaGo was unleashed onto outside competition. In a

---

[51] Ibid.
[52] Ibid.
[53] Silver, David. "Deep Reinforcement Learning." Google DeepMind. https://deepmind.com/blog/deep-reinforcement-learning/. Accessed April 3rd, 2017
[54] Ibid.

competition against all other Go playing programs, AlphaGo dominated, winning 499 out of 500 possible contests. Even most expert human players stood no chance; in October 2015, AlphaGo shutout European Champion Fan Hui by a score of 5 to 0. The following March, as mentioned previously, AlphaGo beat Lee Sedol, one of the best Go players in history.

Today AlphaGo, much as Deep Blue, can easily best any human competitor. However, whereas I argued that, despite its skill, Deep Blue did not possess *nous*, as it did not play chess in an intuitive fashion, I think it is quite evident that AlphaGo has successfully recreated human intuition, at least that which pertains to Go, and thus does at least possess *nous* in a limited sense. AlphaGo cannot (at least currently) do anything besides play Go; accordingly, I cannot make stronger conclusion.

Admittedly, AlphaGo, by virtue of being a supercomputer, has many luxuries a human does not, such as being able to process millions of possible moves at once or being able to play against itself. However, despite this, fundamentally, the way AlphaGo learned how to play Go and the way AlphaGo currently plays Go is the same as that of a human. A child learns a game by first watching others play to acquire basic knowledge and strategies. AlphaGo too began its learning process via watching others play, attempting to predict not only what move they would play but also why they would play it much as in the same way a human would. Again, thanks to reinforcement learning, AlphaGo mastered the game of Go through playing games repeatedly and learning from mistakes. The way a master human player has learned Go is the same. Likewise, a human playing Go relies nearly entirely on intuition acquired from experience. AlphaGo plays in the exact same way. A human Go player undoubtedly possess significant *nous*. As there is no appreciable difference between the way a human or AlphaGo learns and plays the game of Go, it follows that AlphaGo possess *nous* in this area as well.

## VI. Sophia & Phronesis

In the following section, I argue that today's computers can possess neither of *sophia* or *phronesis*. These two virtues require a wide range of cognitive abilities far too advanced and complicated for any modern computer to satisfy. So, in this section, after introducing each of these virtues, I will discuss some reasons why computers cannot possess either. Subsequently, I will discuss what a computer would have to do, or the way in which a computer would have to work, to satisfy some of the most basic requirements for *sophia* and *phronesis*. Attempting to determine under what conditions computers could possess either, or both, virtues in their entirety is too difficult and speculative of a task for this paper; so instead, I will discuss some areas where computers are currently making improvements, or are poised to make improvements in the near future, that are relevant to *sophia* and *phronesis*. Specifically, I will discuss internal goals, which is relevant to both *sophia* and *phronesis*, and emotions, which is relevant to only *phronesis*. Hopefully, this section can serve as a jumping off point for future discussion in this area.

*Sophia*, theoretical wisdom, is a combination of *episteme* and *nous*; however, it is more general in scope than either. Unlike *episteme*, *sophia* is not a specific, deductive form of knowledge and unlike *nous*, *sophia* is not only concerned with understanding un-deducible first premises. *Sophia*, rather, is a theoretical, holistic, and universal knowledge of first principles and causes.[55] Specifically, as Aristotle writes in the *Metaphysics*, the truly wise man, the man with *sophia*, is one that:

[55] Berti, Enrico. "Gadamer and the Reception of Aristotle's Intellectual Virtues." *Revista Portuguesa De Filosofia*, vol. 56, no. 3/4, 2000, pp. 345–360., www.jstor.org/stable/40337581.

> has in the highest degree universal knowledge; for he knows in a sense all the instances that fall under the universal. And these things, the most universal, are on the whole the hardest for men to know; for they are farthest from the senses.[56]

Further, Aristotle lists two qualities a man with *sophia* possesses. Reminiscent of *nous*, the man with *sophia*, Aristotle claims, intuitively knows these causes and principles despite being unable to comprehend each and every detail of them – instead, he understands each part of the whole and how they fit together.[57] Likewise, reminiscent of *episteme*, the man with *sophia* can capably teach others these concepts. Further, one with *sophia* seeks knowledge only for its own sake, rather than for some other end; the first principles and causes, Aristotle claims, cannot be known otherwise.[58]

*Phronesis*, practical wisdom, separates itself from each of *episteme*, *nous*, and *sophia*, in that it is not merely a rational state.[59] Rather *phronesis* is the intellectual virtue which allows for the proper determination and application of moral virtues and so involves a highly experiential component. These moral virtues, for Aristotle, are dispositions,[60] that is they are states one occupies which allows us to choose a specific moral action in every unique situation. Further, these moral virtues are mean states, in that they fall somewhere between vices of excess and deficiency. Courage, for example, falls somewhere in between cowardice and recklessness, although not precisely in the middle. As mentioned previously, for one to become a *phronimos*, these moral virtues must also be habituated by the individual. A *phronimos* does not act morally

---

[56] Aristotle. *Metaphysics*. W.D. Ross, trans. Oxford: Clarendon Press, 1908. Reprint. Stillwell, KS: Digireads 2009 p. 4

[57] Berti, Enrico. "Gadamer and the Reception of Aristotle's Intellectual Virtues." *Revista Portuguesa De Filosofia*, vol. 56, no. 3/4, 2000, pp. 345–360., www.jstor.org/stable/40337581.

[58] Aristotle. *Metaphysics*. W.D. Ross, trans. Oxford: Clarendon Press, 1908. Reprint. Stillwell, KS: Digireads 2009 p. 3

[59] Gottlieb, Paula. "Aristotle on Dividing the Soul and Uniting the Virtues." *Phronesis*, vol. 39, no. 3, 1994, pp. 275–290., www.jstor.org/stable/4182478.

[60] Moss, Jessica. "'Virtue Makes the Goal Right': Virtue and 'Phronesis' in Aristotle's Ethics." *Phronesis*, vol. 56, no. 3, 2011, pp. 204–261., www.jstor.org/stable/23056294.

after careful deliberation, but rather does so immediately through instinct. Further, part of this habituation process involves developing the appropriate feelings towards the individual virtues. The *phronimos* must learn to take joy in performing a virtuous action, and must learn to feel pain when performing a non-virtuous action.[61]

A central reason why I argue today's computers cannot possess either of *sophia* or *phronesis* is that modern computers cannot set goals for themselves. Aristotle explicitly states that one with *sophia* seeks knowledge for its own sake as the result of some internal drive. One clearly is not seeking knowledge for its own sake if it is only doing so thanks to some external force.

*Phronesis* is a similar case. Although in a sense, the aim of *phronesis* is to habituate each moral virtue, as mentioned above, clearly this habituation must occur through experience; moral virtues cannot be taught or expressed as a series of rules. Instead, the *phronimos* must discover these virtues through his own personal experiences and volition. So, although moral virtues are the desired goal of *phronesis*, *phronesis* is responsible for determining what these goals are.[62] *Phronesis,* for example, determines what courage is in every relevant scenario, or determines what generosity is in every relevant scenario. Clearly then, these are not goals that can be set by some external being, as the individual must discover them himself.

Invariably, computers cannot and do not set their own goals; rather, they can only perform the tasks that their programmers have designed them to do. Rephrased, computers must follow externally given goals, and they are incapable of creating or following their own goals. So, a computer can neither seek knowledge for its own sake nor discover for itself what moral

[61] Demos, Raphael. "Some Remarks on Aristotle's Doctrine of Practical Reason." *Philosophy and Phenomenological Research*, vol. 22, no. 2, 1961, pp. 153–162., www.jstor.org/stable/2104836.
[62] Moss, Jessica. "'Virtue Makes the Goal Right': Virtue and 'Phronesis' in Aristotle's Ethics." *Phronesis*, vol. 56, no. 3, 2011, pp. 204–261., www.jstor.org/stable/23056294.

virtues are. So, a clear current limitation of modern machines is that they lack this internal drive necessary for both *sophia* and *phronesis*.

Even machines that use sophisticated machine learning methods fall short in this regard. For example, AlphaGo, although it may possess intuition/*nous*, can use this only to learn Go quickly and play go well. It cannot repurpose its intuition for other aims, unless its programmers, a la IBM's Watson, explicitly rewire it do so. So, again, it does not seek knowledge for its own end, as clearly this is a goal that cannot be set externally. Further, as with all other computers, it does not set its own goals the way the *phronimos* has done. AlphaGo, and any other machine today, can only fulfill a specific purpose that has been externally determined. Machines cannot set their own goals, and currently cannot function in the abstract and general way (i.e. outside of the explicit aims of programmers) necessary to do so. So, at the very least, *sophia* and *phronesis* would require of a machine this ability to reason abstractly.

There is currently some modest progress being made in developing abstract reasoning machines. Neural Engineer Chris Eliasmith has had some success in building some general reasoning machines by creating virtual simulations of certain areas of the human brain. Eliasmith's SPAUN, a small-scale 2.5 million neuron simulation of the human brain, can successfully complete eight diverse, albeit very limited tasks by mimicking the firing patterns of the neurons in the human brain.[63] The eight tasks are as follows: copy drawing, image recognition, list recollection, counting, basic question answering, variable creation, and fluid reasoning.[64] To perform these tasks, in each instance SPAUN is first presented a visual image and draws its responses using a mechanical arm. For example, when performing the first task,

---

[63] Eliasmith, Chris et al. "A Large-Scale Model of the Functioning Brain." *Science*. November 30th, 2012. Vol. 338. pp. 1202-1205. http://science.sciencemag.org/content/338/6111/1202
[64] Ibid.

copy drawing, SPAUN is "given a randomly chosen handwritten digit [and] should produce the same digit written in the same style as the handwriting."[65] SPAUN performs the other seven tasks in this type of fashion as well.

Importantly, unlike other AI such as AlphaGo, SPAUN has not been designed to perform only some specific task; rather, SPAUN has been designed to mimic human behavior. To clarify, although the builders of SPAUN aimed to create an artificial entity that can reason generally, they did not explicitly design SPAUN to perform the eight specific tasks mentioned above. SPAUN performs these eight precisely because it can reason generally and in a similar fashion to that of a human. In fact, while performing these tasks, SPAUN's "connectivity and functional ascriptions to brain areas… are consistent with current empirical evidence."[66]

Clearly, SPAUN has quite some way to go, as there is a vast difference between being able to copy a handwritten '7' and being able to reason more broadly or being able to set general goals. There is an even larger gap between SPAUN's current capabilities and the ability to desire knowledge or to determine what the moral virtues are. However, SPAUN's being able to perform even rote tasks without being designed to do so explicitly represents an exponential improvement versus other attempts to recreate general reasoning, and so is a promising base from which to build on for the purposes of *sophia* and *phronesis*. Further, it is possible that rapid improvements in SPAUN's power may soon be made, as currently it is only a simulation of 2.5 million neurons (the human brain has approximately 100 billion).[67] Expanding this neural simulation could lead to impressive results in the area of abstract thought in the future and so could perhaps serve as a key building block for machine possession of *sophia* and *phronesis*.

---

[65] Ibid.
[66] Ibid.
[67] Ibid.

Another limitation of today's computers which prevents them from possessing *phronesis* specifically is that computers lack emotions. This inability of computers to feel emotions has been a common criticism in the history of AI. Famously, Geoffrey Jefferson's 1949 oration, presented by Alan Turing as a response questioning the conclusion of his Turing Test, expresses the limitations of machines as so:

> "Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain-that is, not only write it but know that it had written it. No mechanism could feel (and not merely artificially signal, an easy contrivance) pleasure at its successes, grief when its valves fuse, be warmed by flattery, be made miserable by its mistakes, be charmed by sex, be angry or depressed when it cannot get what it wants."[68]

Jefferson's argument, in its most basic form, is that emotions are an inseparable component of consciousness and cognition – the latter cannot exist without the former. Machines today, inarguably, do not have emotions. Therefore, via the logic of this argument, they cannot have cognition.

Similarly, the reason why modern computers cannot possess *phronesis* is that they lack emotions. As mentioned above, emotionality, specifically the proper feeling of pleasure and pain, is a central aspect of *phronesis*. As moral virtues are dispositions, rather than just merely qualities, for an individual to possess some virtue he must not only understand and habituate it, but also feel appropriately towards it. Accordingly, *phronesis* requires of its possessor a proper orientation of emotions; specifically, the *phronimos* must feel a sort of pleasure from performing the right, moral action and pain from performing the wrong one, a key step in the habituation process. Clearly, modern computers do not feel pleasure and pain, let alone any emotions at all, thus failing this requirement.

---

[68] Jefferson, Geoffrey. "The Mind of Mechanical Man." *British Medical Journal* 1.4616. 1949. p. 1110.

Modern machines clearly do not possess emotions. However, Marvin Minsky, the former head of MIT's Artificial Intelligence laboratory, argues that in the future, machine emotionality is entirely possible. Currently, Minsky notes, mechanical processes do a capable job of modeling basic instincts. Surprisingly large amounts of human infant or animal behavior can be modeled as nothing more than a collection of what Minsky calls if-do rules.[69] For example, an if-do rule would be "*If* you are too hot, *Move* into the shade," or "*If* you are hungry, *Find* something to eat."[70] Clearly, rules such as these could not model the more sophisticated realms of human thought; however, Minsky argues, in certain combinations these rules can model many sophisticated behaviors of animals and fish, including mating, parenting, nesting, and defending territory.[71] Accordingly, if one were to build a robot that could interact with the outside world in a way reminiscent of an animal (e.g. it could move around, it could defend itself, it could use sensors to see objects or sense temperatures), then by following only if-do rules, this "rules-based reaction-machine," or "instinct-machine," could perform these instinct behaviors as well.[72]

So, although basic instincts can be modeled by machines, what is currently preventing machines from possessing genuine emotions, Minsky argues, is lack of human knowledge on the subject. That is, humans do not understand either how emotions work or the relationships between emotions to the degree necessary to implement them in machines, and the lack of machine emotionality is not due to any inherent limitation of machines themselves. Further, for Minsky, the way emotions are commonly perceived today is mistaken. As he says:

> The meanings of words like "feelings" [or] "emotions" seem so natural, clear, and direct to us that we cannot see how to start thinking about them. However… none of those popular Psychology words refers to any single, definite process; instead

---

69 Minsky, Marvin. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster, 2006. Print.
70 Ibid.
71 Ibid.
72 Ibid.

> each of those words attempts to describe the effects of large networks of processes inside our brain.[73]

Emotions, for Minsky, are a complicated series of processes, rather than a simple state. Further, he argues, emotions only arise from how specific parts of the body interact with and the "stuff" of which these parts are made is irrelevant.[74] If our perception of emotions were re-oriented to be seen in such a way, Minsky claims, then much as how thought and intuition has to a degree successfully become mechanized in recent forms of AI (think of the neural nets in AlphaGo), perhaps emotions could be represented as a series of mechanical processes as well and perhaps it would be possible to "design machines that can feel."[75]

Minsky's book is highly speculative in nature; accordingly, although he is optimistic on machine emotionality, there is no guarantee such machines can be built. Even if machines could one day possess emotions, similarly to SPAUN, much more would need to be done for machine possession of *phronesis*. Feeling emotions is only a building block for later steps. To satisfy fully the emotionality requirement of *phronesis*, the machine would also have to be able to from, experience, orient its emotions in the proper way. Specifically, it must learn to feel pleasure at performing a virtuous act, and feel pain from performing a non-virtuous act. Beyond this, machines would also have to be able to habituate moral virtues as well, that is, act virtuously without deliberation thanks to the proper orientation of these feelings. Clearly, neither of these tasks are soon forthcoming or easily accomplishable.

---

[73] Ibid.
[74] Ibid.
[75] Ibid.

## VII. Conclusion

Currently, as evidenced by the discussion of *sophia* and *phronesis*, computers are simply too limited and inflexible to satisfy all of Aristotle's criteria of rationality. Likewise, even when computers do satisfy one of the intellectual virtues, aside from *episteme* (as it seems to be had by machines trivially), it is either only in a specific case (Watson with *techne*) or in a very narrow sense (AlphaGo with *nous*). Further, it does not seem as if, again *episteme* excluded, computers currently have the capability to possess multiple virtues simultaneously. Clearly, computers must improve significantly for them to be deemed rational beings in the Aristotelian sense.

I am, however, optimistic this improvement can occur. Although computers may be quite a ways from satisfying some, I feel I have shown there does not seem to be anything, in principle, that would prevent computers from possessing Aristotle's intellectual virtues. Requirements for three of the five virtues have already been met, and, if optimists such as Minsky or Eliasmith are to be believed, the remaining two virtues can be met eventually as well. Perhaps this will be sooner than we think, as computers have grown exponentially in sophistication, power, and ability year-by-year. Twenty years ago, for example, it was unthinkable that a machine could play Go competently, let alone at a world-champion level. Likewise, even five years ago, it would have been difficult to argue that computers could possess the second component of *techne*, let alone *nous* (even though it is in a limited sense). Hopefully, this rate of rapid advancement continues into the foreseeable future and hopefully machines will continue to surprise.

As computers continue grow, so too will the discussion around them. Some brief areas for further discussion include: what would it take for computers to experience sensation, which

Aristotle sees as another necessary component of pleasure and pain, how would Aristotle analyze self-driving cars (as they have locomotion and sensation of sorts via sensor), is *sophia* possible without knowledge of the first principles (e.g. the internet is a knowledge of all things, but does not arise from first principles), under what conditions would computers possess *nous* in a more general sense, whether computers can satisfy other requirements of *phronesis*, whether computers can possess other forms of Aristotelian souls, and whether the issue of the inseparability of the rational soul can be rectified. I hope attempts to answer questions such as these can serve as a guide for AI research in the future.

# VIII. Works Cited


Aristotle. *De Anima*. Trans. Hugh Lawson-Tancred. New York: Penguin Books, 1986.

Aristotle. *Metaphysics*. W.D. Ross, trans. Oxford: Clarendon Press, 1908. Reprint. Stillwell, KS: Digireads 2009

Aristotle. *Nicomachean Ethics*. Trans. Terence Irwin. Indianapolis: Hackett Publishing, 1999.
Aydede, Murat. "Aristotle on *Episteme* and *Nous*: The *Posterior Analytics*." Southern Journal of Philosophy (1998),

Vol. 36, No. 1, Blackwell Publishing Ltd, pp. 15–46, http://faculty.arts.ubc.ca/maydede/Aristotle.pdf

Berti, Enrico. "Gadamer and the Reception of Aristotle's Intellectual Virtues." *Revista Portuguesa De Filosofia*, vol. 56, no. 3/4, 2000, pp. 345–360., www.jstor.org/stable/40337581.

Bilton, Nick. "The Rise of 3-D Printed Guns." *The New York Times*. August 13th, 2014. https://nyti.ms/2nD3bye. Accessed April 3rd, 2017

Corey, David "Vogelin and Aristotle on Nous: What is Noetic Political Science?." *The Review of Politics*, Vol. 64, No. 1, 2002, pp. 57-79

Demos, Raphael. "Some Remarks on Aristotle's Doctrine of Practical Reason." *Philosophy and Phenomenological Research*, vol. 22, no. 2, 1961, www.jstor.org/stable/2104836.

Domingos, Pedro. "A Few Useful Things to Know about Machine Learning." 2012. https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf. Accessed April 3rd, 2017

Eliasmith, Chris et al. "A Large-Scale Model of the Functioning Brain." *Science*. November 30th, 2012. Vol. 338. pp. 1202-1205. http://science.sciencemag.org/content/338/6111/1202

Ferruci, David et al. "Building Watson: an Overview of the DeepQA Project" *AI Magazine*, 2010. Vol. 31. pp. 59-79

Fountain, Henry. "A Lab-Grown Burger Gets a Taste Test." *The New York Times*. August 15th, 2013. https://nyti.ms/2m8xy2u. Accessed April 3rd, 2017.

Gonthier, Georges. "Formal Proof – The Four-Color Theorem." Georges Gonthier. Notices of the American Mathematical Society. 2008 Vol. 55.

Good, I.J. "The Mystery of Go." January 21st, 1965. New Scientist. http://www.chilton-computing.org.uk/acl/literature/reports/p019.htm Accessed April 3rd, 2017

Gottlieb, Paula. "Aristotle on Dividing the Soul and Uniting the Virtues." *Phronesis*, vol. 39, no. 3, 1994, pp. 275–290, www.jstor.org/stable/4182478.

IBM. *Watson Health Medical Imaging Collaborative Expands to 24 Members; IBM Debuts Watson Imaging Clinical Review, the First Cognitive Imaging Offering*. *IBM*. IBM, 20 Feb. 2017. Web. 3 Apr. 2017.

IBMWatsonSolutions. "IBM Watson: How It Works." *YouTube*. YouTube, 07 Oct. 2014. Web. 04 Apr. 2017.

Jefferson, Geoffrey. "The Mind of Mechanical Man." *British Medical Journal* 1.4616. 1949.

Kosman, Aryeh. "Aristotle on the Virtues of Thought." *Virtues of Thought*, Harvard University Press, 2014, www.jstor.org/stable/j.ctt6wpmjx.18.

Meagher, Robert. "Technê." *Perspecta*, vol. 24, 1988, www.jstor.org/stable/1567132.

Miller, Fred. "Aristotle's Philosophy of Soul." *The Review of Metaphysics*, Vol. 53, No. 2, 1999. pp. 309-337 http://www.jstor.org/stable/20131355

Miller, Fred. "Aristotle on the Separability of Mind." *The Oxford Handbook of Aristotle*. August 2012.

Minsky, Marvin. *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*. New York: Simon & Schuster, 2006.

Moss, Jessica. "'Virtue Makes the Goal Right': Virtue and 'Phronesis' in Aristotle's Ethics." *Phronesis*, vol. 56, no. 3, 2011, pp. 204–261., www.jstor.org/stable/23056294.

Munoz, Andres. "Machine Learning and Optimization." 2014. https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf. Accessed April 3rd, 2017

Newborn, Monty. *Deep Blue: an artificial intelligence milestone*. New York: Springer, 2003.

Oaksford, Mike. *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford University Press: Oxford, 2007.

Otake, Tomoko. "IBM big data used for rapid diagnosis of rare leukemia case in Japan." *Japan Times*. August 11th, 2016. http://www.japantimes.co.jp/news/2016/08/11/national/science-health/ibm-big-data-used-for-rapid-diagnosis-of-rare-leukemia-case-in-japan/#.WOM0ThLytE4. Accessed April 3rd, 2017

Schwalbe, Ulrich & Walker, Paul. "Zermelo and the Early History of Game Theory." August, 1997. http://www.math.harvard.edu/~elkies/FS23j.03/zermelo.pdf

Searle, John R. "Minds, Brains, and Programs." 1950, *Behavioral and Brain Sciences* Vol. 3, pp. 417-457.

Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search" N.p., 27 Jan. 2016. Web. 3 Apr. 2017. http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html.

Silver, David. "Deep Reinforcement Learning." Google DeepMind. https://deepmind.com/blog/deep-reinforcement-learning/. Accessed April 3rd, 2017.

Thorsten Wuest, Daniel Weimer, Christopher Irgens & Klaus-Dieter Thoben. "Machine learning in manufacturing: advantages, challenges, and applications." *Production & Manufacturing Research*. 2016. pp. 23-45. http://dx.doi.org/10.1080/21693277.2016.1192517

Turing, Alan "Computing Machinery and Intelligence." 1950, *Mind 49*. pp. 433-460.

"Surgeons perform 'world's first' implant of entire 3D printed plastic skull dome." *The Russian Times*. March 28th, 2014. Accessed April 3rd, 2017

"IBM Research | Deep Blue | Overview." *IBM Research | Deep Blue | Overview*. N.p., Februrary 23rd, 2001. Accessed April 3rd, 2017. https://www.research.ibm.com/deepblue/meet/html/d.3.3a.shtml