Descartes and Spinoza in the 21st Century: the advantage of reductionist architectures for Artificial Intelligence and how new advances in holistic methodologies may be incorporated.

David Hui

8th November 2016

Word Count: 4106

Artificial Intelligence has been described as our last invention. [1] Whether that is because it would do all tasks for us or eradicate humanity, it is a topic increasingly relevant to a contemporary society driven by and dependent on technological advances occurring at an exponential rate. [2] It is therefore imperative that a technique for creating Artificial Intelligence, or AI, balances security with efficiency. It is the ultimate goal of researchers and industrialists to create Artificial General Intelligence, or AGI, a machine capable of exhibiting intelligence equal or superior to that of a human, given any task. [3] In this essay, I will use holistic and reductionist approaches to classify techniques used to make AGI. Holistic theorems claim that intelligence is a complex emergent property of the system as a whole, generated from the interaction of its constituents [4], and reductional theories believe that there exists a fundamental mathematical or physical rule responsible for and explaining intelligent actions. [5] Although reductionism generated many discoveries from the 1960s to 1980s, they have since been replaced in favour of holism. [6] The two different approaches represent a fundamental difference in the way in which AI may be realised, and changing viewpoints may be the difference between fulfilling the possibility of AI or proving the fears about AI. During the 1980s, reductionism, was known as "GOFAI", standing for "Good Old-Fashioned Artificial Intelligence" [7] and holism was known to researchers as connectionism [8]. Then, articles were discussed about the relative merits of both schools of thought [9] [10]. Although reductional GOFAI techniques have not changed much over the last 30 years, holistic connectionist techniques have changed drastically, warranting an update of the review [11]. In this essay, the meaning of AI and methods of creating it will be discussed using a historical-philosophical context. Then, the efficiency and safety of the methods will be compared in the new context of holism versus reductionism, as well as reasons for the increasing relevance of reductionism.

'Artificial Intelligence' is a misnomer. Intelligence is a property, which may be artificial, but it is ultimately an attribute of an entity. [12] Entities which exhibit intelligence are named agents, and may be biological or mechanical. [13] Since the 17th century, philosophers, beginning with Descartes, have wondered about the possibility of reducing thinking and reasoning into a list of rules and instructions. [14] If written as a computer program, the list is known as a procedure. [15] Procedures which terminate given any input form a special operating class called algorithms. [16] An automaton is an agent which implements procedures and, according to Descartes, is what every system, including the mind and the brain, can be reduced to. [17] Three hundred years later, speculation about the possibility of 'thinking machines' were renewed by a paper published by Alan Turing. This 1936 paper explored the computability of numbers and theorems, forming the basis of modern-day computers. [18] In addition, Turing published another paper in 1950 establishing a metric of intelligence, based on the comparison of performances between a human and a computer. If a machine could fool a human evaluator into thinking that it was human through tests, such as conversing in natural language, the machine could be classified as being intelligent. [19] 70 years later, the 'Turing Test', as it has become known, has not yet been passed, with the best agent only fooling 30% of evaluators. [20] However, machine intelligence has proved to be superior over humanity in chess [21], the Jeopardy quiz contest [22] and recently Go [23]. It is believed that there will come a time where machines will be able to repeatedly self-improve, creating an intelligence explosion dwarfing that of humanity. Such an event is referred to as the Singularity. [24]

There are two types of agent: strong AI, which are intelligent [25], and weak AI [26], which are a simulation of intelligence. Although the Turing Test can recognise intelligence, it cannot distinguish between a weak or strong agent and thus it will fail, as a simulation of intelligence is not actually intelligent itself. An example of failure is given by the Chinese Room thought experiment, proposed by John Searle, Professor of Philosophy at the University of California. In Searle's argument, the Turing Test is conducted in Chinese with an agent with no knowledge of Chinese. The agent is isolated inside a sealed room, and communication is achieved only by passing printed notes through a cat flap. The agent has a large database of input-output pairs for every possible input, constructed such that the output makes perfect sense given the input. When an input is given to the agent in the room, it looks up the input argument in the database of pairs and outputs the other value in the pair, passing the test. [27] Searle elaborates his argument further and claims that it is impossible for a strong AI to be procedurally created, because the meanings, or semantics, of words cannot be encoded within the rules, or syntax, of programs. Machines thus have no knowledge of what they are outputting, and are only simulating intelligence. [28] However, this argument is not well founded because syntax must be based on semantics in order for the procedure to output responses that make sense within the context of a conversation. [29] Thus Searle's argument is shown to be incorrect. So in order for an agent to be strong, it must have a method of acquiring its own knowledge by means of encoding semantics within syntax. The subfield of Artificial Intelligence which deals with this problem is Machine Learning. [30] It is an important field as it enables a machine to learn autonomously without the aid of a human, and also enables a machine to adapt to different situations and discover previously unknown techniques. Using Machine Learning, an agent can be trained to become a strong AI, and thus the problem of designing AI can be transferred to designing Machine Learning algorithms. As with AI, Machine Learning can be categorised into holistic and reductional procedures.

The existence of a reductional strategy, or architecture, for AI hinges on whether a set of rules governing all intelligent behaviours and actions exist. The rules, once found by Machine Learning, would then be applied to a variety of environments in order to find the best strategy for achieving goals. Reductional architectures are model-based and agents contain an internal model. [31] The agent decides which action to take based on the results of experiments on its internal model. The earliest and perhaps most famous reductional model is Cartesian dualism. Invented and named after Descartes, the theory hypothesizes that intelligent agents can be reduced into two automata, the mind and the body. [32] The body performs two functions, firstly, it sends a perception of the surrounding environment to the mind, and secondly, it implements actions and responses received from the mind. The mind only performs one function: it finds strategies based on what the body has perceived to achieve goals. In 1976, the neuroscientist David Marr gave a name to each of these three functions. The body's functions were known as the 'computational layer' and 'physical layer', and the mind's function was called the 'algorithmic layer'. These three layers formed what Marr referred to as a 'tri-level' reductional model of AI, and each layer could be represented by a different rule. [33] With this sound framework in place, early research focused on how this reductional architecture could be realised by investigating procedural implementations on these three layers. [34] In general, there are two main problems with reductional architectures: the problem of acquiring new knowledge that may be helpful for problem solving, and the problem of evaluating how close a solution is towards a goal.

If it is possible to summarise intelligence into a single statement, then it may be likely that the statement can be encoded into an equation. A useful metric of intelligence is the ability of an agent to adapt to a wide range of environments. [35] This can be quantified by measuring the rate at which an agent achieves goals in a wide range of environments, as a more intelligent agent would achieve goals at a faster rate. In order to be classified as intelligent, an agent is forced, or constrained to achieve goals. Once the goal is satisfied, the constraint is then lifted. Thus an intelligent agent is able to reduce the number of constraints on it quickly, which is, in other words, maximising its future freedom of action. This statement has been formalized into an equation which is the basis of a 'sapient software' called Entropica. [36] Developed by Alexander Wissner-Gross of Harvard University, Entropica is an agent that 'maximises its future freedom of action'. Entropica is so named because it is based on thermodynamics and information theory, the two most universal theories to describe systems in physics. Every system contains information, and every system obeys the second law of thermodynamics: overall entropy increases, given enough time. Currently, using the same equation and a reductional approach, Entropica has been used to discover a more efficient trade network, earn money on the stock exchange, and to find new solutions to cooperation puzzles. [37]

Contrary to reductional architectures, holistic architectures have no common structure and are not model-based. By eliminating structure, designers consider the system as a whole comprised of components. Components interact in complex ways, generating emergent behaviours, which would perhaps be even more powerful than if the system was split up into rules. Behaviours can be analysed by grouping actions by complexity. Initially, the components interact with each other, forming simple actions which can be grouped together into a layer. Then, these actions can interact with each other, increasing complexity and forming the next layer. Increasingly higher layers are increasingly abstract, enabling complex behaviours to be produced. [38] A contemporary of Descartes, Spinoza, opposed 'mind-body duality', and instead formulated a holistic theory of mind centralized around the interaction of pain, pleasure and desire. As the fundamental components of the human body are muscles, the interactions of muscles can be grouped together to form actions, which themselves form behaviours. Behaviours themselves are characterised into those that maximise pleasure and those that minimise pain, which Spinoza argues is equivalent to maximising desire. [39] This principle forms the basis of reinforcement learning. An agent is placed in an environment, and, through experimentation, is rewarded if certain goals are achieved. Eventually, the agent learns how to behave optimally inside an environment in order to maximise desire obtained by achieving goals, resulting in complex phenomena. Although the behaviours learnt holistically are more powerful than the actions generated reductionally, there are two problems with holistic approaches. Firstly, a holistic architecture is more complicated. It may be made up of thousands of individual components, all of which would have to be trained. Secondly, it is difficult for a holistic architecture to adapt to different environments. [40] Whilst reductional architectures are designed with transferability and universality in mind, holistic architectures learn specific features found in environments the agent has been exposed to.

Reinforcement learning is used to train algorithm called a neural network, the algorithm which currently achieves the best results. [41] Though they are used holistically in reinforcement learning, they can also be reductionally used as a layer in Marr's tri-level model. An artificial neural network is morphologically analogous to that of a biological neural network. [42] Both networks consist of interconnecting neurons, but artificial neurons mathematically abstract the function of a neuron, rather than modelling the neuron on a chemical scale. [43] If the inputs of the mathematical neuron satisfy an 'activating function', the neuron fires and signals 1. This signal can then be linked up to other neurons. Generally, the structure of the network can be abstracted into layers, each containing an ensemble of neurons. In a feedforward network, the outputs of each neuron can only be connected to subsequent layers. If the network is designed to choose between n outputs, the last layer in the network would have n neurons. This last output layer returns a list of 0s and 1s. Once the network is trained, only one element in that would be 1 given any input. All the other outputs would be 0. Each output is represented by a neuron, and the output of the whole network is taken to be the output represented by the neuron which returns 1. The network is trained by the backpropagation of errors algorithm. Initially, the error of a network is computed by taking the difference between the observed and expected output of the network. Then, the parameters of the 'activating function' for each neuron in the output layer are adjusted such that the observed output matches the expected output. All of the parameters of the activating functions are adjusted, going backwards layer by layer through the network until the first, or input layer of neurons is reached. [44] The network is able to organise itself into producing complex emergent phenomena by updating the network as a whole. Although neural networks usually yield good results, it is uncertain what the optimal shape or parameter values that govern the network should be. What activating function to use or how much backpropagation should change the network are unknown. Usually, researchers experiment with all of these factors, only selecting the best model after exhaustively searching through all possible options. [45] Rather than holistically experimenting, a reductional approach could be used to find rules governing the behaviour of neural networks, to save time.

These experiments have spawned a variety of different neurons and network structures, which has led to neural networks becoming increasingly complicated. [46] Although neural networks were invented in the 1980s, the recent rise in complication has spawned a new field of study in neural networks called Deep Learning. Deep learning generalises neural networks by adding more layers, neurons and parameters but removing rules about connections and structure of layers. A major difference is that neurons can now output continuous values. There are also new types of neural network, collectively titled Deep Nets. In a convolutional neural network, some layers may be an ensemble of networks, rather than neurons. Recurrent neural networks are the most structurally complex. There are no restrictions to how neurons can be connected, and a neuron may even feed back into itself. Although significantly more complex than before, Deep Learning is dwarfed by the size and complexity of the human brain. [47] The most significant result in Deep Learning is reinforcement learning, a holistic architecture which teaches an agent the optimal behaviours for all inputs. Deep Nets have been taught old Atari games, inventing new techniques surprising even their developers. However, each network can only play a specific game, and cannot be trained to play multiple games unlike a human brain. [48] On the contrary, the human brain is model-based, as rules about the game are formulated and subsequently manipulated in order to achieve a goal. It would seem that the human brain is reductional.

Due to similar reasons, researchers in the 1960s believed that AI was reductional, and that the algorithmic layer of Marr's tri-level model was mainly a searching algorithm. [49] As with all reductional techniques, the agent contains an accurate model of the environment. There are two important states in the model: the current state and the desired goal state. The agent searches for a solution from the current state, by testing all possible actions one by one to see whether the goal can be reached by one action. If not, then sequences of actions are tried. All possible combinations of two actions are tested, then all possible combinations of three and so on, until the goal state is reached. [50] To reduce searching time, searching can be done in opposite directions from the starting state and goal state until both searches have found each other. [51] This technique, known as bidirectional search, was independently improved by Alan Turing and Claude Shannon in 1950 to create an algorithm used to play two-player games such as chess. Originally called the Turing-Shannon procedure, it is now known as the minimax algorithm and is based on the principle that an agent is trying to maximise the same quantity that that an adversary is trying to minimise. [52] If playing chess, each state of the chessboard is assigned a desire value based on how well the agent is performing as opposed to the opponent using the points value of the pieces on the board. From the current state, all possible actions are simulated in turn to generate many new states, then, from each new state, all possible actions of the opponent are taken, which generates even more states. This cycle is repeated until a pre-programmed limit. The optimal strategy can be found by considering the last move taken. If the move was executed by the agent, then the action that guarantees the maximum return value is assumed to be taken. If the move was executed by the adversary, then the action that guarantees the minimum return value is assumed to be taken. Eventually, going back to the first move yields the best action. [53] Minimax was implemented in 1996 on Deep Blue by IBM, which was famously used to beat the then world champion, Gary Kasparov, at chess. [54] However, minimax is severely limited by the complexity of the game that it is trying to solve. Games that have a greater total number of possible moves such as Go, an ancient Chinese game, require more memory and time to compute the next move, and is subsequently infeasible. [55] The two greatest problems facing minimax and reductional architectures are how new knowledge can be acquired, and how all possible states can be assigned with a desire value, or how the states can be evaluated. Although reductionism was responsible for much of the early successes in AI, there were problems that needed to be fixed.

The effort led to improvements on the 'computational layer' of the tri-stage hypothesis, such that it could be used to evaluate the state of an environment and acquire new knowledge. In the 1970s, rule-based expert systems were developed which could deduce states using rules and data. [56] By the 1980s, such algorithms were commercialised and could detect what illnesses a patient had [57], teach students electronics [58], and act as mission control on board the Space Shuttle [59]. However, the problem of these systems was that rule-based algorithms were exceptionally sensitive to noise. The presence of noise, or an incorrect observation of the environment could lead to unfounded deductions [60]. To combat random error, or noise, rules became probabilistic and expert systems became fuzzy-logic inference models, which could only express probabilistic deductions. However, the problem of how an agent could acquire new knowledge about a world was still persistent. [61] The problem of evaluating states were equally difficult. So far, new rules had been added only by humans, not by machine learning, which was time-intensive. A famous example is Cyc, an ambitious Artificial Intelligence startup. [62] The company, founded by Douglas Lenat, famously claimed that "Intelligence [is] ten thousand rules" [63], and sought to codify all of human knowledge into expert system rules. The project failed partly because it was difficult to codify abstract concepts, partly because rules were subjective to opinion, but mostly because the 'ten thousand rules' presented a problem far too complex for an algorithmic layer to search. [64] By the 1990s, no more research in expert systems were conducted, having been replaced by other algorithms. [65]

Having observed the pitfalls of reductional agents, Rodney Brooks of MIT created the first holistic architecture called subsumption architecture. [66] In order to overcome the problems of new knowledge acquisition and state evaluation, subsumption architecture directly evaluates raw environmental inputs, making an internal model redundant and the structure of a 'tri-level' model unnecessary. The four principles of subsumption were situatedness, embodiment, intelligence and emergence. [67] The agent had to be embodied and situated in an environment. Instead of experimenting with an internal model, the agent experiments in the environment. Emergent behaviour that displayed intelligence would then be rewarded, enabling the agent to learn. Subsumption architecture is used in the iRobot Roomba vacuum cleaner and lawnmower. Increasingly abstract concepts were observed in the robots, from avoiding objects, moving around the environment, and then exploring the environment whilst vacuuming or mowing. [68] Although the problems of reductional architectures were circumvented, subsumption produced two new problems. Despite achieving goals to a high standard, it was difficult for the robots to adapt to different environments, and to be given different goals from what they were trained to do. [69] Due to a lack of an internal model and 'algorithmic layer', the robots had no goal seeking behaviour. However, subsumption architecture had shown to be a viable alternative to fuzzy logic in terms of sensing the environment, and could be used to solve the problem of new knowledge acquisition.

A famous agent which uses a holistic architecture for the 'computational layer' is AlphaGo, an agent trained by Google DeepMind which beat the world champion, Lee Sedol in Go. [70] The 'computational layer' was a deep net called the policy network, trained to acquire new knowledge to output the three most popular moves from any arbitrary state. Evaluating environmental states was achieved by using another deep net called the value network. These two networks approximate the minimax algorithm and reduces the number of states to search. [71] Thus an overall reductional architecture utilising segments of holistic techniques seem like a promising route towards AGI.

Although the two main problems hindering a reductional approach to AGI have been overcome, there are still three key problems. Firstly, there is an insufficient amount of training data for 'general intelligence', secondly, the training time is too long [72], and thirdly, current hardware has insufficient memory and computational power to process all the data, even if it existed [73]. With reductional architectures, the training time and the amount of data needed is less because the problem has been reduced to merely recognising the current state, evaluating whether the state is favourable and acquiring new knowledge about the environment. With holistic architectures, the behaviour given all states in all possible environments needs to be determined. However, there is a mathematical bound on the performance of all agents. The No Free Lunch Theorem states that it is impossible to design an agent that performs optimally in all possible environments. If the agent is procedural, then an environment can be designed which deliberately ensures that whatever the agent's procedure returns, what it thinks is the 'optimal action', is, in fact, suboptimal. The second clause of the theorem states that all algorithms perform equally poorly when faced with all possible unknown situations. [74] Thus universal intelligence is impossible. On the contrary, it is possible to build an agent which performs well for all environments that the human mind can conceive, because this is a finite subset of all environments. It may be possible to find a single equation governing intelligent actions in this environment. Although finding this reductional rule may be difficult, it may be even more difficult to train and retrain a holistic architecture on every possible environment.

In addition, a holistic architecture may be more dangerous. According to Nick Bostrom, Professor of Philosophy at St. Cross College, Oxford, we can only ensure humanity's survival before the singularity. [75] When AGI is at the cusp of self-improving by rewriting itself, researchers need to ensure that all future policies created by an AI would not include eradicating humanity. With a reductional architecture, it is easier to reason about the future behaviour of an AI system, whereas with a holistic architecture, it is harder to predict what emergent techniques may arise. It has been claimed that an AGI with Entropica at its core will not even consider exterminating humanity because that would be imposing a constraint on itself, contradicting the principle of not maximising its future freedom of action. [76]

In summary, a reductional architecture is still relevant towards current research because enables a greater control over AI, better adaptability towards different situation and is more functional as it is goal driven, rather than behaviourally oriented. In general, this makes it a more route towards creating AGI. As holistic approaches are better at learning specific goals, they may prove to be a useful replacement of some components of a reductional architecture. Even if a reductional technique is implemented, it can be argued that AI is ultimately holistic, because every operation in a computer can be expressed as 0 or 1. [77] Goals and actions would be merely an emergent, abstract, intelligent property. To complement computation's bottom-up holistic nature, a top-down reductional approach would be similar to a bidirectional search, reducing the search time. But whether if Artificial Intelligence eradicates humanity, enhances humanity, or is merely the product of hubris, I think that it would be beneficial to have a single equation concerning intelligence. It would provide a remarkable insight into our own minds, and, perhaps, could be used to help humanity in times of need.

References

[1] Barrat, J., *Our Final Invention: Artificial Intelligence and the End of the Human Era*, Thomas Dunne Books, New York, 1st Edition, 2013.

[2] Moore, G., *Cramming More Components onto Integrated Circuits*, Electronics. Retrieved 8th November 2016 from `http://www.cs.utexas.edu/~fussell/courses/cs352h/papers/moore.pdf`.

[3] `http://www.agi-society.org/`, retrieved 8th November 2016.

[4] Smuts, J.. Holism and Evolution 2nd Edition. Macmillian and Co, 1927.

[5] Ney, A. *Reductionism.* Internet Encyclopedia of Philosophy, University of Tennesse, retrieved 8th November 2016 from `http://www.iep.utm.edu/red-ism/`.

[6] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 2nd Edition, 2003, p 25.

[7] Ibid, p 951

[8] Ibid, p 25.

[9] Bersini, H., *Connectionism vs Gofai: A Brief Critical Analysis*, Springer, Berlin, 1989.

[10] Crick, F. *The recent excitement about neural networks.* Nature, 1989.

[11] Nielsen, M. *Neural Networks and Deep Learning*, Determination Press, 2015, retrieved 8th November 2016 from `http://neuralnetworksanddeeplearning.com/about.html`.

[12] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, p 34.

[13] Ibid, p 46.

[14] Walter, S. *Epiphenomenalism.* Internet Encyclopedia of Philosophy, University of Bielefeld, retrieved 8th November 2016 from `http://www.iep.utm.edu/epipheno/#H2`.

[15] Cormen, T., Leisterson, C., Rivest, R., Stein, C., *Introduction to Algorithms*, MIT Press, Massachusetts, 3rd Edition, 2009. p 6.

[16] Ibid, p 5.

[17] Kim, J., *Philosophy of Mind*, Westview Press, Perseus Books Group, 2006, pp 40 - 41

[18] Turing, A., *On Computable Numbers, with an Application to the Entscheidungsproblem*, Proceedings of the London Mathematical Society, 1936, published 1937, pp. 23065, Retrieved 8th November 2016 from `http://plms.oxfordjournals.org/content/s2-42/1/230`.

[19] Turing, A., *Computing Machinery and Intelligence*, Mind, 1950, pp 433460, Retrieved 8th November 2016 from `http://mind.oxfordjournals.org/content/LIX/236/433`.

[20] `http://www.aisb.org.uk/events/loebner-prize#finals2016`, retrieved 8th November 2016.

[21] `http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/`, retrieved 8th November 2016.

[22] *Computer crushes the competition on 'Jeopardy!'.* Associated Press. 2011. Archived from the original on February 19, 2011, retrieved 8th November 2016 from `https://web.archive.org/web/20110219023019/https://www.google.com/hostednews/ap/article/ALeqM5jwVBxDQvVKEwk_czuv8Q4jxdU1Sg?docId=2e3e918f552b4599b013b4cc473d96af`

[23] Byford, S. *"DeepMind's AlphaGo secures victory over Lee Se-dol with third straight win"*, 2016, The Verge, retrieved 8th November 2016 from `http://www.theverge.com/2016/3/12/11210650/alphago-deepmind-go-match-3-result`

[24] Markoff, J. *When Is the Singularity? Probably Not in Your Lifetime.*, The New York Times, The New York Times Company, 2016, retrieved 8th November 2016 from `http://www.nytimes.com/2016/04/07/science/artificial-intelligence-when-is-the-singularity.html`

[25] Kurzweil, R., *The Singularity is Near*, Viking Press, New York, 2005, p 260.

[26] Ibid. f

[27] Searle, J. *The Rediscovery of the Mind*, M.I.T. Press, Massachusetts, 1992, p 44.

[28] Searle, J., *Minds, Brains and Programs*, Behavioral and Brain Sciences, 1980, retrieved 8th November 2016 from `http://web.archive.org/web/20071210043312/http://members.aol.com/NeoNoetics/MindsBrainsPrograms.html`

[29] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, pp 1032 - 1033.

[30] Bishop, C., *Pattern Recognition and Machine Learning*, Springer Science + Business Media, LLC, New York, 2006, pp 1 - 2.

[31] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, pp 48 - 58.

[32] Ibid, p 2017.

[33] McClamrock, R., *Marr's Three Levels: A Re-evaluation*, University at Albany, Minds and Machines, 1991, retrieved 8th November 2016 from `http://www.albany.edu/~ron/papers/marrlevl.html`

[34] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, pp 18 - 20.

[35] Legg, S. and Hutter, M. In Proc. 1st Annual artificial general intelligence workshop, editor, Goertzel, B., 2006, retrieved 8th November 2016 from `http://www.vetta.org/documents/A-Collection-of-Definitions-of-Intelligence.pdf`

[36] Wissner-Gross, A., Freer, C., *Causal Entropic Forces*, Physical Review Letters, Americal Physical Society, 2013, retrieved 8th November 2016 from `http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.110.168702`

[37] `http://www.entropica.com/`, retrieved 8th November 2016.

[38] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, p 1003.

[39] Schmitter, A., *Spinoza on the Emotions*, Stanford Encyclopedia of Philosophy, 2010, retrieved 8th November 2016 from `http://plato.stanford.edu/entries/emotions-17th18th/LD5Spinoza.html`

[40] `https://openai.com/blog/openai-gym-beta/`, retrieved 8th November 2016.

[41] Nielsen, M. *Neural Networks and Deep Learning*, Determination Press, 2015, retrieved 8th November 2016 from `http://neuralnetworksanddeeplearning.com/index.html`.

[42] Winston, P., *Artifiical Intelligence*, Addison-Wesley, Massachussetts, 3rd Edition, pp 443 - 446.

[43] McCulloch, W., Pitts, W., *A Logical Calculus of Ideas Immanent in Nervous Activity*, Bulletin of Mathematical Biophysics, 1942, retrieved 8th November 2016 from url-http://link.springer.com/article/10.1007%2FBF02478259

[44] Nielsen, M. *Neural Networks and Deep Learning*, Determination Press, 2015, retrieved 8th November 2016 from `http://neuralnetworksanddeeplearning.com/chap2.html`.

[45] ttp://www.pyimagesearch.com/2016/08/15/how-to-tune-hyperparameters-with-python-and-scikit-learn/ retrieved 8th November 2016.

[46] Nielsen, M. *Neural Networks and Deep Learning*, Determination Press, 2015, retrieved 8th November 2016 from http://neuralnetworksanddeeplearning.com/chap2.html.

[47] Goodfellow, I., Bengio, Y., Courville, A., Book in preparation for M.I.T. Press, unpublished, retrieved 8th November 2016 from http://www.deeplearningbook.org/contents/intro.html.

[48] https://www.youtube.com/watch?v=0X-NdPtFKq0, retrieved 8th November 2016.

[49] Winston, P., *Artifiical Intelligence*, Addison-Wesley, Massachussetts, 3rd Edition, p 79.

[50] Ibid, pp 63 - 79.

[51] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, pp 90 - 92.

[52] Franchi, S., Guezeldere, G., *Mechanical Bodies, Computational Minds*, M.I.T. Press, Massachussetts, 2005, pp 48 - 51.

[53] Winston, P., *Artifiical Intelligence*, Addison-Wesley, Massachussetts, 3rd Edition, pp 103 - 104

[54] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, pp 29.

[55] Ibid, p 186.

[56] Ibid, pp 22 - 23.

[57] Buchanan, B., Shortliffe, E., *Rule-Based Expert Systems*, out of print, retrieved 8th November 2016 from http://people.dbmi.columbia.edu/~ehs7001/Buchanan-Shortliffe-1984/MYCIN%20Book.htm.

[58] *http://files.eric.ed.gov/fulltext/ED098963.pdf*, retrieved 8th November 2016.

[59] Rasmussen, A., Muratore, J., Heindel, T., *The INCO Expert System Project: CLIPS in Shuttle mission control* NTRS, 1990, retrieved 8th November 2016.

[60] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, p 480.

[61] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, p 1003.

[62] Stork, D., Lenat, D., *From 2001 to 2001: Common Sense and the Mind of HAL. Hals Legacy: 2001 as Dream and Reality.*, Cycorp Inc., 1996, retrieved 8th November 2016 from http://www.cyc.com/wp-content/uploads/2015/06/img-615153705-0001-1.pdf.

[63] Johnston, B., *The Collection of Physical Knowledge and Its Application in Intelligent Systems*, Artificial General Intelligence: 4th International Conference, AGI 2011, Mountain View, CA, USA, August 3-6, 2011, Proceedings, p 163.

[64] Mullins, J., *Whatever happened to machines that think?*, retrieved 8th November 2016 from https://www.newscientist.com/article/mg18624961-700-whatever-happened-to-machines-that-think/

[65] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, p 24.

[66] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, p 1003.

[67] Brooks, R., *Cambrian Intelligence: The Early History of the New AI.* Cambridge, Massachusetts, M.I.T. Press, 1999, pp. 165170, retrieved 8th November 2016 from http://dlia.ir/Scientific/e_book/Technology/Mechanical_Engineering_Machinery/TJ_210.2_211.47_Mechanical_Devices_Figures_/020957.pdf.

[68] https://www.ibm.com/developerworks/library/j-robots/, retrieved 8th November 2016.

[69] Russell, S. & Norvig, P., *Artificial Intelligence, A Modern Approach*, Prentice Hall, New Jersey, 3rd Edition, 2003, p 1004.

[70] Byford, S. *"DeepMind's AlphaGo secures victory over Lee Se-dol with third straight win"*, The Verge, 2016, retrieved 8th November 2016 from http://www.theverge.com/2016/3/12/11210650/alphago-deepmind-go-match-3-result.

[71] Silver, D., Huang, A., Maddison, C., Guez, A., Sifre, L., Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., *Mastering the game of Go with deep neural networks and tree search*, Nature, 2016, retrieved 8th November 2016 from http://www.nature.com/nature/journal/v529/n7587/full/nature16961.html.

[72] Simonite, T., *Algorithms That Learn with Less Data Could Expand AIs Power*, M.I.T. Technology Review, https://www.technologyreview.com/s/601551/algorithms-that-learn-with-less-data-could-expand-ais-power/.

[73] https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-cus html, retrieved 8th November 2016.

[74] Wolpert, D., Macready, W. *No Free Lunch Theorems for Optimization*, IEEE Transactions on Evolutionary Computation, 1997, retrieved 8th November 2016 from http://ti.arc.nasa.gov/m/profile/dhw/papers/78.pdf.

[75] Bostrom, N., *Superintelligence*, Oxford University Press, Oxford, 2014, pp 78 - 83.

[76] https://www.youtube.com/watch?v=ue2ZEmTJ_Xo, retrieved 8th November 2016.

[77] Mullins, R., *Raspberry Pi Tutorials*, https://www.cl.cam.ac.uk/projects/raspberrypi/tutorials/turing-machine/one.html.