Marta Ziosi

The three worlds of AGI. Popper's theory of the three worlds applied to artificial general intelligence

Conference Item [Paper]

The Three Worlds of AGI

Popper's Theory of the Three Worlds Applied to Artificial General Intelligence

Marta Ziosi

Abstract This Capstone applies Popper's Three-worlds paradigm to the academic discourse on Artificial General Intelligence (AGI). It intends to assess how this paradigm can be used to frame the opinions of scientists and philosophers on Artificial General Intelligence (AGI) and what it reveals about the way the topic of AGI is approached from the fields of the Sciences and the Humanities. This has been achieved by means of a Literature Review reporting the opinions of main philosophers and scientists and by analysing two main projects – project CYC and project SOAR- advanced as possible ways to achieve AGI. As a result, most academics from the field of Science seem to better fit views on AGI interpreted through the lens of Popper's World 2, the world of the mind. On the contrary, most philosophers seem to better fit views on AGI interpreted through the lens of Popper's world 3, the world of the products of the human mind such as theories, knowledge and ideas. As a suggestion, this Thesis advocates the promotion of interdisciplinarity and discussion among the different academic fields.

1 INTRODUCTION

Back in 1965 the US psychologist Herbert Simon proclaimed that machines will be capable within 20 years to do any work a man can do (Simon, 1965). Nevertheless, the present state of affairs showcases how the promise has not withheld its foretelling. Why? It is a matter of timing? Or is it an illusionary idea which can avail itself solely of these empty '20 to 30-years' futurist prognoses? Opinions largely differ and many a times collide within people from different levels of expertise and belonging to different fields of research. Different opinions can be gathered from branches of Computer Science to Philosophy, from the Cognitive Sciences to Technology and Media Studies; more generally, from the fields of the Sciences to the ones of the Humanities.

Arguably, the question ought not to be of an 'all or nothing' nature but one about the approach we as humans should take towards General Intelligence. Plainly, the past years have witnessed an incredible confluence of storage of big data, probabilistic programming and sheer increase in computing power. However, computers are still not capable of engaging in some apparently easy tasks for humans. The approach should be in thinking about robots and AGIs – Artificial Intelligent Agents - not just as a technology which engages in physical and computational work. The key relies in thinking about them indeed as physical computational entities but in relation to humans[1]. Several researchers are already engaging with such an approach. The main questions which are being asked are of the kind, *'How can we and What does it mean to create an AGI which thinks?'* or '*What does it mean to create an AGI with a common sense of human society, knowledge and culture?'*.

I hypothesize that while researchers in the field of science tend to work on the first question, the ones in the humanities tend to focus on the latter. However, any potential answer to both questions fundamentally requires both computational capabilities or understanding of algorithms from the sciences and critical thinking or the heuristics of the humanities. Thus, if the intent is to reach a *generally* intelligent agent, the efforts ought to hail from an as encompassing as possible *interdisciplinary* background. To achieve that, we ought to agree on the question to ask. This is essential in order to avoid the carry-out of *miscommunication* under the illusion of *disagreement*.

This research will thus propose a framework to swiftly cut through the two different approaches in order to identify their differences in topic and purpose. The core-framework will be provided by Popper's theory of the Three Worlds. The question which instructs this Capstone is '*How can Popper's Three-worlds paradigm be applied to frame the opinions of scientists and philosophers on Artificial General Intelligence (AGI) and what does it reveal about the way the topic of AGI is approached?'.* The core information on the topic of AGI will be proffered by means of exposing the main ideas and opinions over AGI of mainly scientists and philosophers. Conceivably, a thorough analysis of these will be conducted by applying the chosen framework. The last word is left to the conclusion where a suggestion on how to deal with discrepancies in opinions will be advanced.

Finally, it is important to state that this thesis does not aim at predicting future scenarios and it aligns itself with Popper's claim that predicting technological innovation is impossible (Popper, 1979). Indeed, if humans could, they would already know how to implement it, thus leaving no logical space between the prediction and the realization of the technology. The intended relevance of this thesis is principally to provide a broader outlook on matters of AGI and it aims at breaching through the AGI discourse by Popper's toolbox of World 2 and World 3 in order to expose a potential thought-gap or

[1] I am aware and I will not deny the importance of the physical part of the process of computation. This sentence is solely aimed at emphasising the importance of thinking about this 'physical part' in relation to human capabilities, given that the goal is Artificial General Intelligence.

discrepancy of opinions between two chief-fields. A suggestion in favour of interdisciplinarity will be advanced at the end.

2 DEFINITIONS

ARTIFICIAL INTELLIGENCE

2a. Intelligence

To begin with, it is important to define the term 'intelligence' in the way in which it will be used in the paper. Intelligence is the '*computational* part of the ability to achieve goals in the world' (Stanford, 2017). There are varying kinds and degrees of intelligence which occur in people, in many animals and some machines. As it has not yet been decided which computational procedures ought to be called 'intelligent', it is also extremely difficult to frame a solid definition of intelligence which detaches itself from any reference to human intelligence as that is the only example at present. Thus, this definition ought not to be dogmatic throughout the Thesis but it mostly serves as a guideline.

2b. Artificial Intelligence

Artificial Intelligence (AI) is '*the science and engineering of making intelligent machines'* (Stanford, 2017). AI does not necessarily limit itself to biologically observable methods. Indeed, even though brain emulation[2] is an example of AI, there are several other approaches to AI such as ones working through probability or brute force algorithms (Goertzel, 2007).

ARTIFICIAL GENERAL INTELLIGENCE (AGI)

2c. General Intelligence

General Intelligence is the ability to achieve complex goals in complex environments (Goertzel, 2007). The plurality of the words 'goals' and 'environments' is crucial to explain how a single goal or a single environment would not account for the word 'general'. Indeed, a chess-playing program is not to be considered 'generally' intelligent as it can only carry-out one specific task. An agent possessing artificial intelligence ought to have the ability to carry-out a variety of tasks in diverse contexts, generalize from these contexts and to construct an understanding of itself and the world which is independent of context and specific tasks.

[2] The process of copying the brain of an individual, scanning its structure in nanoscopic detail, replicating its physical behaviour in an artificial substrate, and embodying the result in a humanoid form (aeon).

2d. AGI

A complete appreciation of the challenges encountered by the idea of 'general intelligence' in the field of AI requires a wide range of perspectives to be adopted. Correspondently, Artificial General Intelligence (AGI) is a highly interdisciplinary field. As it follows from the definition of AI, it could be said that AGI is 'the science and engineering of making *generally* intelligent machines'. As it follows from the definition of General Intelligence, AGIs are expected to solve a wide range of complex problems in several contexts. Additionally, they learn to solve problems whose solution was not presented to them as the stage of their creation. Currently, there are no existing examples of AGIs in the real world.

3. STATE OF AFFAIRS IN AI

The present section will acquaint the reader with a brief background on the history of AI and AGI (first sub-section), the approaches to AGI (second sub-section) and finally, projects and possible solutions (third sub-section).

3a. A bit of history of AI and AGI

In 1956, after the first programmable computer was invented, the genesis of a new field called 'Artificial Intelligence' was announced at a conference at Dartmouth College in New Hampshire (Brey, 2001). This field had the ambition to supply computers – by means of programming - with some sort of *intelligence*. Even before that, the scientist Vannevar Bush had already proposed a system which had the aim to amplify people's own knowledge and understanding (Bush, 1945). It was only five years later when the now celebrated Alan Turing wrote a paper centred around the idea of machines being able to simulate human beings and to carry out intelligent tasks, such as the playing of chess (Turing, 1950). As such, the idea of a machine which could encapsulate some sort of conception of intelligence can already find its space in that years.

3b. Current Approaches

Nowadays, there are two main views held in regard to algorithms. These two shape the different directions taken by approaches to AI. One is held by the proponents of *strong AI* and one by the ones of *weak AI.* The ones defending the former argue that an algorithm is a universal concept which is applicable to anything that works mechanically and thus, also the brain. They argue that human intelligence works through algorithmic processes just like computers. However, as the algorithmic processes regulating the brain are highly sophisticated, they do contend that there does not yet exist any man-made system comparable to it. Yet, it is only a matter of time. On the contrary, proponents of weak AI maintain that even though aspects of human thinking are algorithmic, there are critical aspects about the way humans are given to experience the world which do not

fit the algorithmic picture and probably never will. Humans experience the world from sensations. These two characteristic approaches to AI also shape any groundwork on AGI. Hence, they ought to be kept in mind throughout the Thesis to better grasp the subject matter.

3c. Projects

Apart from these two main approaches, there are several projects which have been attempted through the years and which are important to present in order to better understand the nature of the concerns and points advanced in the literature review. Two projects will hereby be presented. It is important to state that they differ in approach. These two projects are the CYC project and the SOAR project.

In the mid 80s, the CYC project began as an attempt to encode common-sense knowledge in first-order predicate logic (Goertzel, 2007). The encoding process was a large effort and it produced a useful knowledge database and a specialised and complex inference engine[3]. However, until today CYC does not *'solve problems whose solution was not presented to them at the stage of their creation'* (see AGI definition). Plainly, it does not come up with its own solutions; which is a defining feature of AGIs. CYC researchers have encoded in the system common-sense knowledge. However, this knowledge-filled database has resulted in an open-ended collection of data more than dynamic knowledge. By making use of declarative language by means of Lisp syntax[4], CYC features the ability to deduce concepts. However, differently from neural networks techniques, it still relies on humans inputting an 'unending' amount of data before outputting any result. This is one of the main critiques adduced to the CYC case. CYC enthusiasts have rushed in its defence by stating that CYC has the potential to be imported in future AI projects (Goertzel, 2007).

Adopting an opposite approach, Allen Newell's SOAR project is a problem-solving tool which is based on logic-style knowledge representation and mental activity figured as 'problem solving' expressed by a series of heuristics (Goertzel, 2007). The core of the effort behind the SOAR project is to investigate the architecture which underlies intelligent behaviour (Rosenbloom, Laird & Newell., 1993) and what constitutes intelligent action rather than knowledge. SOAR can be described as a sequence of three cognitive levels; the memory level, the decision level and the goal level. These are merely descriptive terms which are used to refer to the mechanism constitutive of the SOAR architecture (Rosenbloom, Laird & Newell, 1991). Even though it represents a great step in the AGI field, up until now the system is still a disembodied problem-solving tool lacking the autonomy and self-understanding which are expected in an AGI.

[3] More insights can be found on the site: www.cyc.com

[4] Lisp is the second-oldest high-level programming language favoured for research in Artificial Intelligence. It allows to interchangeably manipulate source codes as a data structure. His command line is called a *Read-Eval-Print-Loop* as it *reads* the entered expression, *evaluates* them and *prints* the result (Chisnall, 2011).

4. LITERATURE REVIEW

Notwithstanding the various pursuits for AGI implementation, the discipline was propelled chiefly from an idea. The present section will focus on the intellectual life and discourse surrounding AGI. This section lays the groundwork for the future analysis.

4a. Different Worlds

Through the following paragraphs, it is more specifically presented how, through the years, the expectations and what are considered the key factors on the way to AGI have differently developed on the side of the Humanities and on the side of the Sciences. The following paragraphs ought to elucidate this claim. Even though with a risk of redundancy, it is important to state that all the scholars and great minds presented in the paragraph 'The Stance of the Science World' come mostly from a scientific background, while the ones in 'The Stance of the Humanities' come mostly from a philosophy background. Some of them have also expertise in both fields. In that case, they are found in the section for which their background is stronger. The following paragraphs provide the content which will be subject to the application of the Theoretical Framework later in the paper.

4b. The Stance of the Science World

Influenced by the groundwork of Alan Turing, the 70s featured the creation of Putnam's 'mentalist project' (Dreyfus & Haugeland, 1974). The mentalist project was an endeavour to represent the rules that govern human behaviour and the mind by a Turing machine table that relates input and output states. Concurrently, the scientists Newell, Shaw and Simon who were in the 1950s considered the pioneers of Cognitive Simulation, announced that *'within ten years most theories in psychology will take the form of computer programs'* (Simon & Newell, 1957, p.8). George Miller himself, a distinguished psychologist at Harvard, asserted that the current developments in the study of man's understanding could be viewed as a system of *information processing* (Miller, Galanter & Pribram, 1960, p.57). The configuration of mental processes as computations was taken beyond a mere analogy.

A more critical stance towards the ability of re-creating certain mind-phenomena such as consciousness through algorithms is provided by the scientist Roger Penrose in his book, 'The Emperor's new Mind' (1989). On one hand, he claims that the mind understood as 'consciousness' cannot be computed. However, he contends that this is impossible only as long as the model is based on the idea of a Turing Machine, as the latter only *mimics* mental processes and does not progress towards any kind of 'understanding' for the machine. Even

though rejecting the Turing Machine's paradigm, as many other scientists he strongly defends that more generally mental activity is *'the carrying out of some well-defined series of operations'* (Penrose, 1989, p.17). He resorts to call these operations 'algorithms'. Penrose does convene that mental activity can be represented through algorithms. Additionally, he stresses that human mental processes result in our ability to 'understand' and that is what research ought to focus on. AGIs can improve their performance by experience through a sort of 'feedback system' for performance improvement. According to Penrose, this might account for some kind of 'understanding'.

Another scientist who widely confronted the assumptions underlying AGI implementation is Murray Shanahan[5]. Interestingly, as also Penrose proposed, he figures the main challenge on the road to AGI as a matter of endowing a system with '*common sense understanding'*. Howbeit, Shanahan considers '*common sense understanding*' to need to blend with *creativity*. He calls both these elements 'cognitive ingredients' and while describing AGI, he locates it in what he calls '*the space of possible minds'* (Shanahan, 2016). Thus, he adopts a mind-stance. In the space of possible minds, AGI can figure either by means of 'whole brain emulation'[6] or by constructing an artificial brain which matches a statistical description of a new-born's central nervous system. Even when Shanahan admits that the human brain is not necessarily the starting point on the path to AGI, he proposes different architectures such as brute force search algorithms and machine learning techniques which approach the problem on terms of computation (Shanahan, 2016). Indeed, he convenes that 'human thinking' can be instated through computation, whether they resemble the brain or not.

4c. The Stance of the Humanities

On the side of the humanities, the philosopher Hubert Dreyfus claims that AGI is based on a boastful epistemological assumption. This assumption implies that all knowledge is formalizable. Plainly, humans' thoughts and actions have produced a body of knowledge on which human reality feeds itself and stands on. Howbeit, AGI assumes that this body of knowledge can be expressed in context-independent formal definitions and rules (Brey, 2001, p.5). He asserts that while these formal rules can successfully describe human knowledge, they cannot be used to reproduce it. In fact, the application of these rules is actually context-dependent. Hence, he contends that there is a body of knowledge - constitutive of human reality - which ought to be acknowledged in AGI implementation. However, at the same time he stresses that this knowledge is too dependent on circumstances and on context to be successfully objectively formalized; this is where the main challenge lies.

[5] It is important to state that, even though Murray Shanahan is an expert in Cognitive Robotics, he also engaged in several philosophical work.

[6] The process of copying the brain of an individual, scanning its structure in nanoscopic detail, replicating its physical behaviour in an artificial substrate, and embodying the result in a humanoid form (aeon).

Another philosopher who adds a valuable contribution to the topic is John Searle. Dreyfus and Searle agree on the fact that (Strong) AI relies on another mistaken assumption. Strong AI figures intelligent systems as symbol processing systems (Brey, 2001, p.4; Searle, 1990, p.26). According to this view, thinking merely consists in symbol manipulation rather than meaning and human knowledge. Additionally, such an assumption furthers the idea that the mind stands to the brain as a program stands to the hardware. Searle however, strongly refutes this view. He claims that minds are not programs. In fact, programs are formal, syntactic and thus, they are sufficiently defined in terms of symbol manipulation. For example, a line of program can be 'if 01, then print 1'. In this case, a program does not need to understand or have *knowledge* of what '01' means in order to execute 'print 1' and to move from symbol '01' to symbol '1'. Differently, human minds have *mental contents* (Searle, 1990) and the linguistic understanding which happens between people who intend to share mental contents requires a semantic framework as provided by the net of human knowledge. This is what enables the conveying of meaning. As it is presently defined, strong AI appears to overlook this difference which is instead crucial when dealing with 'general intelligence'.

5 THEORETICAL FRAMEWORK

This section presents the theoretical framework which provides the lens through which the literature will subsequently be analysed.

5a. Core argument: 'Popper's 3 worlds'

Karl Raimund Popper was born in Vienna in 1902. He is one of the most prominent philosophers of Science. Karl Popper is more commonly associated with Critical Rationalism and his most acclaimed work is about Falsificationism and the evolution of objective knowledge in scientific inquiry. A special focus will hereby be dedicated to his pluralist view on reality.

Popper advocates a pluralist view of human reality. According to him, there exist three 'Worlds' or 'sub-universes' (Popper, 1979). World 1 consists of physical bodies. Plainly, elements of it are physical living and non-living objects such as stars, stones, animals and plants. World 2 is the world of conscious experience. It is the mental and psychological world filled with subjective experiences, mental states like pain and pleasure, perceptions and intentions. It is what humans think about the world as they try to map, represent, hypothesize or anticipate in order to maintain their existence in an ever-changing place. Finally, world 3 is the world of the products of the human mind. This broadly includes languages, songs, paintings, mathematical constructions, theories and even culture.

Popper strongly advocates not only the existence of the products of the human mind, but also their being real rather than fictitious. As far as these have a causal effect upon us, they ought to be real. Products of the human mind, for example

scientific theories, have proven to have an impact on the physical world by changing the way humans build things and utilize them. Popper believes that the causal impact of world 3 is more effective than scissors and screwdrivers (Popper, 1979). Furthermore, even though elements of World 3 are generally instantiated in a concrete object of World 1 – books, physical components of a computer… -, it is not a necessary condition that they be so expressed (Sloman, 1985).

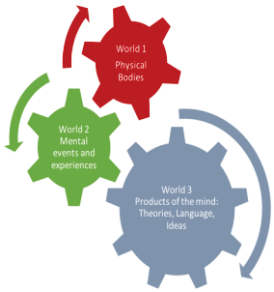

Figure 1. Popper's Three Worlds visualization

This simple above visualization suggests Popper's acknowledgement of the interaction between the three worlds. According to Popper, World 3 theories or plans always ought to be primarily understood by a mind in World 2 before they be operationalized. Withal, the theory itself and its operationalization have effects on World 1 physical objects. An example can be purported by Einstein's Theory of Relativity. The scientific community had to first subjectively grasp the content of the Theory of Relativity before this could be applied to change the physical reality. Hence, World 2 proves itself to be a necessary intermediary between World 3 and World 1. Likewise, as Einstein's special Theory of Relativity lead to the creation of the atomic bomb, World 3 impacts World 1.

Finally, both for the specific purpose of this research and to follow Popper's emphatic concern for this distinction, we ought to precisely differentiate between 'thought processes' and 'thought contents'. The former belong to World 2 while the latter to World 3 (Popper, 1979). Even though these two might appear to be interchangeable, they are fundamentally and foundationally different. It is paramount to understand that the process of thinking is unlike the knowledge which this process itself unveils. This distinction ought to be sheltered in the reader's mind as it gains momentum in the following paragraphs.

6 DISCUSSION

6a. Popper's Three Worlds

Programmatic processes – ex. Algorithms - and the data which they output and process act in interplay. For example, intelligent systems' internal algorithms are designed to deal with the data they are inputted with and the way they process these data consequently modifies the output. These processes – such as algorithms – and data – such as big packages of information – both ought to exist and co-exist in an AGI system and they have an impact the one on the other. While several algorithms in AGI aspire to imitate *thought processes*, the knowledge or data which they process and output can be thought of as the *content* which is the *product* of these processes. As Karl Popper stressed, *thought processes* – related to mental events and states - and *thought contents* – related to objective contents of thoughts – belong respectively to two different 'worlds' and hence, they are foundationally and fundamentally different (Popper, 1978). Indeed, the process of thinking is unlike the knowledge which this process itself unveils. Both concepts seem to unilaterally figure in the understanding and explanations of AGI, depending on from which field – Science or Humanities – the claim originates. Now, do they?

Both Penrose and Murray Shanahan build the foundations of their work on AGI on the conviction that the mind can be computed and specifically Penrose refers to AGI as a matter of 'mental processes' which manipulate information. On the other hand, philosophers such as Dreyfus claim that AGI systems ought to be deeply characterized by the character of the information which they manipulate and thus, they stress the role of World 3 *thought contents*. Popper's pluralist view helps to shed light on this subtle and yet fundamental distinction which appears to delimitate mainly the views of researchers in Philosophy and Scientists on the topic of AGI.

Arguably, if we read the topic of AGI under a World 2 lens, both subjective experience and mental tasks are key words (Popper, 1983). As per subjective experience, in the section 'Experience as Method' Popper addresses subjective empirical experience as the structured, logical description of only one world – the 'world of our experience' (Popper, 1983) - out of an infinite number of logically possible worlds. In the AGI case and for computers, the expression of their only 'world of experience' happens through binary logic and their 'mental tasks' are carried out through algorithms. Computer scientists and AI researchers adopt binary logic as their main tool and psychologists and neuroscientists primarily study mental tasks and subjective experience. Could this favour a reading of AI from a World 2 perspective?

On the other hand, in 'Epistemology without a knowing subject', Popper considers World 3's objective knowledge such as theories and ideas as something which does not need a knowing subject; as an entity independent of anybody's disposition or belief towards knowledge (Popper, 1972). Once we apply this to the context of AGI, Dreyfus would agree in the sense that we, as humans, rely on a body of knowledge that we have produced. That knowledge can be used to *describe* human behaviour. He claims that there is a body of knowledge that ought to be recognized in the implementation of AGI. Nevertheless, this last of Popper's formulations dissents with Dreyfus acknowledgement of the importance of *context* in matters of human knowledge. Indeed, Dreyfus contends that human knowledge is highly dependent of context and circumstances and henceforth, not independent of a subject.

Searle would also recognize the importance of a net of human knowledge from which to derive meaning. Nevertheless, he would also disagree in the sense that for him this knowledge is highly dependent of people's dispositions towards it. Thus, even though both philosophers would stress the importance of 'knowledge', Popper's world 3 does not exhaust what is important in their views.

6b. In the real World

The attempt to interpret the AGI discourse by means of the tension between World 2 and World 3, might advance a hypothesis on a possible reason why projects such as CYC and SOAR have not resulted to be successful (from section 'Projects and Possible Solutions'). On one hand, the CYC was started with the aim to encode all common knowledge. However, as it is a knowledge-filled database, it has resulted in the accumulation of data. On the other hand, the SOAR project was started with the aim to instantiate mental activity. However, as it reproduces 'intelligent action' by algorithms rather than knowledge, it has resulted in a disembodied problem-solving tool. It is clear how 'General Intelligence' cannot be reached unilaterally. While the endeavours of the CYC project might be better represented by World 3, SOAR's endeavours might be better represented by World 2. It ought to be acknowledged that in reality these two Worlds interact. Thus, it might be fruitful to think about a General Intelligent machine as something which can integrate both *though processes* and *thought contents*, the content of a theory and the processing of it.

7 LIMITATIONS

One of Popper's admirable recommendations is that one ought to expose potential weaknesses of one's theories (Popper, 1983). As per this thesis, there are several factors which ought to be taken into consideration while reading it and of which the reader should be made aware of. The first point concerns the Theoretical Framework. Indeed, the backbone of the argument which derives its structure from Popper's Three Worlds cannot be said to uniformly apply to every case of the AGI discourse or research. While the framework has proven to be arguably sound for some limited cases in Science and Humanities, the panorama can supposably vary for interdisciplinary cases. Some mathematicians are also trained philosophers and vice-versa. Further research could venture in examining such cases.

Moreover, the distinction which Popper meant to draw between the Worlds appears to be an ontological one. In his 'Objective Knowledge' he presents the idea of three different ontological worlds (1972). Furthermore, in his 'Knowledge without a knowing Subject' (1972) and in his 'Three Worlds' (1979) he repeatedly stresses the existence of World 3 independently on a subject perceiving it and it justifies its existence by means of the causal impact it has on other Worlds. Given these considerations, it ought to be stressed that this Thesis utilizes Popper's distinction to try to group different *readings* or different *standpoints* on the matter of AGI. However, it does not claim any ontological difference between the three Worlds.

8 CONCLUSION

The present Thesis has traversed the topic of AGI by first providing a brief account of its history, different approaches and projects. A more in-depth prospect on the matter has been presented by the Literature Review. The Theoretical Framework has served as a toolbox to analyse the AGI discourse from famous academics and scholars. At the very incipit the question was, '*How can Popper's Three-worlds paradigm be applied to frame the opinions of scientists and philosophers on Artificial General Intelligence (AGI) and what does it reveal about the way the topic of AGI is approached?'*. By the end of this Thesis, it can be argued that World 2 and World 3 can be utilized in framing and grouping the opinions of the two disciplines on the topic of AGI. More broadly, this can be framed in terms of approaching the topic by means of *thought processes* (World 2) and *thought contents* (World 3). This analysis can hypothesize discrepancies between the two 'worlds' of Philosophy and Science, when they tend to more strongly approach AGI from just one of the stances. Even though each stance provides a 'safe place' for each field, on one hand it is difficult to rely on World 3's objective knowledge and theories without taking into consideration the *mental processes* which output this knowledge. On the other, it is difficult to claim that 'understanding' automatically arises from World 2's mental processes by overlooking World 3.

In conclusion, what could we learn or advance from this analysis? Overall, the possible suggestions are innumerable but I believe that the incentivizing of interdisciplinarity can favour the opening of worldviews, communication between and within fields and finally, place the AGI discourse in Popper's World 3 where, either as a theory or as a mere human idea, it can be subject to critique. I contend that an interdisciplinary approach ought to be more cherished as it promises more realistically nuanced outcomes than trying to figure out and picture every possible future AGI scenario from each discipline. Furthermore, it can integrate the different stances from each field, transforming an obstacle into an asset. Researchers, professors but also students ought to be acquainted through their path of study with what other fields have to say and with their now still 'alien' worldviews. The ways to push interdisciplinarity on the agenda are innumerable, from curricula in schools and universities to open conferences, journals and more accessible popular events. As it is, AGI is an interdisciplinary matter in itself and it has the potential to lure people towards its topic from several angles. This would also avoid the spreading of *fear* towards the future of AI and AGI, a fear which many a times derives from miscommunication and misunderstanding. We should better concentrate together on what *is* possible rather than on what *might* happen.

REFERENCES

Basic Questions. (2017). *Www-formal.stanford.edu*. Retrieved 16 June 2017, from http://www-formal.stanford.edu/jmc/whatisai/node1.html

Hobbes, T. (1958). *Leviathan* (5th ed., p. 45). Library of Liberal Arts.

Brey, P. (2006). Evaluating the social and cultural implications of the internet. *ACM SIGCAS Computers and Society*, *36*(3), 41-48.

Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*, *4*(3), 185-211.

Chisnall, D (2011). *Influential Programming Languages, Part 4: Lisp*.

Clark, A., & Chalmers, D. (1998). The extended mind. *analysis*, 7-19.

Goertzel, B. (2007). *Artificial general intelligence* (Vol. 2). C. Pennachin (Ed.). New York: Springer.

Dennett, D. C. (2008). *Kinds of minds: Toward an understanding of consciousness*. Basic Books.

Dreyfus, H. L. (1972). What computers can't do. MIT press.

Dreyfus, H., & Haugeland, J. (1974). The computer as a mistaken model of the mind. In *Philosophy of Psychology* (pp. 247-258). Palgrave Macmillan UK.

Dreyfus, H. L., Dreyfus, S. E., & Zadeh, L. A. (1987). Mind over machine: The power of human intuition and expertise in the era of the computer. *IEEE Expert*, *2*(2), 110-111.

Dreyfus, H. L. (1992). *What computers still can't do: a critique of artificial reason*. MIT press.

Frowen, S. F. (Ed.). (2016). *Hayek: economist and social philosopher: a critical retrospect*. Springer.

Goertzel, B. (2007). *Artificial general intelligence* (Vol. 2). C. Pennachin (Ed.). New York: Springer.

Hayek, F. A. (1945). The use of knowledge in society. *The American economic review*, 519-530.

Hecht-Nielsen, R. (1988). Theory of the backpropagation neural network. *Neural Networks*, *1*(Supplement-1), 445-448.

High, R. (2012). The era of cognitive systems: An inside look at ibm watson and how it works. *IBM Corporation, Redbooks*.

Hobbes, T. (1958). *Leviathan* (5th ed., p. 45). Library of Liberal Arts.

Horsman, C., Stepney, S., Wagner, R. C., & Kendon, V. (2014). When does a physical system compute?. In *Proc. R. Soc. A* (Vol. 470, No. 2169, p. 20140182). The Royal Society.

Horwitz, S. (1998). Review of Frowen, Stephen F., ed., Hayek: Economist and Social Philosopher: A Critical Retrospect.

Horvitz, E. J., Breese, J. S., & Henrion, M. (1988). Decision theory in expert systems and artificial intelligence. *International journal of approximate reasoning*, *2*(3), 247-302.

Lloyd, S. (2004) *Programming the universe.* New York, NY: Alfred A. Knopf.

Miller, G. A., Galanter, E., & Pribram, K. H. (1986). *Plans and the structure of behavior*. Adams Bannister Cox.

Penrose R. (1989) *The emperor's new mind.* Oxford, UK: Oxford University Press.

Popper, K. R. (1972). Objective knowledge: An evolutionary approach.

Popper, K. R. (1978). Three worlds. The Tanner Lecture on Human Values. The University of Michigan. *Ann Arbor*.

Popper, K. (1983). *The logic of scientific discovery*. Routledge.

Rosenbloom, P. S., Laird, J. E., Newell, A., & McCarl, R. (1991). A preliminary analysis of the Soar architecture as a basis for general intelligence. *Artificial Intelligence*, *47*(1-3), 289-325.

Rosenbloom, P. S., Laird, J., & Newell, A. (Eds.). (1993). The SOAR papers: Research on integrated intelligence.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences*, *3*(03), 417-424.

Searle, J. R. (1990). Is the brain's mind a computer program. *Scientific American*, *262*(1), 26-31.

Searle, J. R. (1992). *The rediscovery of the mind*. MIT press.

Shanahan, M. (2012). Satori before singularity. *Journal of Consciousness Studies*, *19*(7-8), 87-102.

Shanahan, M. (2015). *The technological singularity*. MIT Press.

Shanahan, M. (2016). *Beyond humans, what other kinds of minds might be out there? – Murray Shanahan | Aeon Essays*. (2016). *Aeon*. Retrieved 16 June 2017, from https://aeon.co/essays/beyond-humans-what-other-kinds-of-minds-might-be-out-there

Simon, H. A., & Newell, A. (1958). Heuristic problem solving: The next advance in operations research. *Operations research*, *6*(1), 1-10.

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London mathematical society*, *2*(1), 230-265.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433-460.

Bush, V. (1945). As we may think. *The atlantic monthly*, 176(1), 101-108.

Sloman, A. (1985). A Suggestion About Popper's Three Worlds in the Light of Artificial Intelligence. *ETC: A Review of General Semantics*, 310-316.