UNIVERSITETET I AGDER

What is Intelligence?

A proposed framework of four different concepts of intelligence.

EVEN ØYGARDEN

SUPERVISOR
Einar Duenger Bøhn

University of Agder, 2019
Faculty of Humanities and Education
Department of Religion, Philosophy and History

Abstract: *Intelligence* is an ambiguous and generally poorly understood concept, and this becomes especially apparent in disagreements on the abilities and capacities of animals and artificial intelligence. In this thesis, I propose a framework of four different concepts of intelligence, that have in common that they describe capacity for complex problem solving, but otherwise have very different meanings and implications. These *are algorithmic intelligence*, which describes the intelligence of algorithms, *system intelligence*, which describes the intelligence of systems (artificial or biological), *heterophenomenal intelligence*, which describes conscious problem solving, and *homophenomenal intelligence*, which describes problem solving based on conceptual understanding. The thesis is an examination of the arguments that justify treating each of these as distinct and separate concepts, and an attempt to show what can be gained in philosophical discourse from such a distinction.

I would like to thank my supervisor at the University of Agder, Einar Duenger Bøhn, for valuable philosophical input and advice. In addition, constructive criticism from Håvard Haugland Bamle, Terje Pettersen, and André Sørvig Lohne has been very helpful.

Contents

Introduction

The concept of intelligence

What is intelligence? The question seems quite answerable on the surface. Most people seem to roughly agree that it is the ability to kind of figure stuff out. ‘Intelligence’ has none of the mystique that comes with terms like ‘mind’ and ‘consciousness’, concepts that deeply puzzle and fascinate philosophers and laymen alike. Rather the opposite, the word intelligence has for a long time been loaded with negative connotations, because it appears to us to describe something very mechanical, something that has little to do with feelings and value. It is universally acknowledged today that tests that seek to evaluate a human beings intelligence tell us nothing about that person’s worth. Just as uncontroversial, is it to say that humans are the most intelligent species on earth. Some people will say that many, perhaps even surprisingly many, other animals are quite intelligent as well, and some will object that we must not confuse intelligence and instincts. Are chimpanzees intelligent? Are octopi intelligent? They might be able to figure stuff out, but do they *really* figure stuff out in the same manner we do, or is our intelligence somehow a less mechanical intelligence than theirs is after all? The concept of intelligence really became a hot topic, of course, with the development of artificial intelligence. Small electronic devices are now able to figure out extremely complicated stuff, and the debate is now whether sometime in the near future they will become *really* intelligent- this time, for the most part, meaning that they are conscious. The word intelligence seems to take on different meanings in different contexts, and confusion is inevitable when those different meanings are implied within the same context. Will we have true artificial intelligence in 20 years? Well, the answer to that does not only hinge on scientific progress, it depends on what we mean when we say intelligent. If animal research uncovers that we have severely underestimated the intelligence of birds, would it have consequences for our ethics? That too, would probably depend on how we define intelligent. The lack of any satisfying existing definition of intelligence when facing important issues like these is the main motivation behind this thesis. ‘What is intelligence?’ is the research question, and what follows is an attempt to provide a satisfying account of this important but unfortunately very general concept. My central claim is that there are in fact four concepts of intelligence, that are closely related but not interchangeable, and awareness of differences between them is crucial to constructive discussion on the subject of intelligence.

Previous attempts at defining intelligence:

> Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do. (Gottfredson, Linda 1997(originally published 1994):13)

The above definition of intelligence was given by Linda Gottfredson, in an article called “Mainstream Science on Intelligence: An Editorial with 52 Signatories, History and Bibliography”, published in *The Wall Street Journal*, December 13, 1994. The article was a response to the public confusion about consensus in psychology concerning intelligence in the aftermath of the publishing of Richard J. Herrnstein and Charles Murray’s controversial book *The Bell Curve: Intelligence and Class Structure in American Life.* (1994). The article did clarify some misconceptions of psychologists’ view on intelligence, but the definition itself could hardly be called crystal clear. The lack of a good definition of intelligence in the field of psychology has been treated on several occasions. *The Journal of Educational Psychology* did a poll on experts’ opinions on the term as early as 1921, and found little consistency in the answers.[1]A similar investigation undertaken in 1986, by Douglas Detterman and Robert Sternberg, leading figures in intelligence research at the time, tried to update the earlier one, hoping that scientific progress had led to more clarity on the matter, but ended up with just as wide a spectre of opinions. Sternberg concluded in his later book, *Metaphors of Mind. Conceptions of the Nature of Intelligence* (1990) that the theories of intelligence rests on metaphors and lack empirical data because they are hard to operationalize and test.[2]

Neither do we particularly want to test it. In the article *Concept of "Intelligence"" Useful or Useless?* (1988), Hans Eysenc acknowledged that it was the most controversial concept in psychology. The concept has been attacked on what seems to be ideological grounds; there have been attempts to break it up into a large number of small and limited

[1] Lanz, Peter (2000) “The Concept of Intelligence in Psychology and Philosophy” In Cruse H., Dean J., Ritter H. (red) *Prerational Intelligence: Adaptive Behavior and Intelligent Systems Without Symbols and Logic*, Vol 1, Vol 2 Prerational Intelligence: Interdisciplinary Perspectives on the Behavior of Natural and Artificial Systems, Vol 3. Studies in Cognitive Systems, vol 26:19-30

[2] Lanz 2000

abilities, in addition to philosophical objections.[3] Many have argued that intelligence is an empty construction, and efforts to measure it thus useless, obviously disregarding its existence as a scientific concept analogous to gravitation or mass.[4] Eysenc set out to show that the concept of intelligence in the same manner as these could be a *useful* one. Those who have argued against the use of the concept have referred to the lack of agreed theory concerning intelligence, and claiming that, in absence of such a theory, it cannot be regarded as a useful concept. Since definition follows theory, there is no agreement on a definition of intelligence.[5]

For philosophers, intelligence has traditionally appeared a less interesting concept than consciousness. Philosophers of mind have occupied themselves with the mind-body problem, mental states and the “hard problem” of consciousness[6], but the development of AI has caused new philosophical interest in the concept of intelligence, though seldom appropriately decoupled from its closely related phenomenon consciousness. It is telling that, in the Stanford Encyclopedia of Philosophy, you will find numerous entries on Artificial Intelligence, almost comparable to the number of articles about consciousness, but not a single entry devoted to defining or explaining the actual concept of intelligence, which appears in philosophical articles pertaining to subjects as separate as logic and AI and the philosophy of religion.

Even in the field of AI, definitions of intelligence are often consciously avoided, though there have been no shortage of attempts at formulations that make sense when speaking of artificial intelligence. Max Tegmark, for instance, defines intelligence as the ability to accomplish complex goals, and general intelligence as the ability to accomplish virtually any goal, including learning.[7]But what does it mean to have a goal? These appear to be philosophical questions, and scientific communities are often impatient with such concerns. In Pei Wang’s article *What do you mean by AI?* (2008), she investigated the matter, and concluded that there existed two widespread opinions on the concept of intelligence. One view was that there exists a natural definition of intelligence, and that different understandings in various schools of artificial intelligence merely deal with different aspects, while they approach the same subject. The other common attitude was to view intelligence as a concept that escapes definition, and that it does not matter what researchers think about this

---

[3] Eysenc, Hans (1988) “Concept of "Intelligence" Useful or Useless?” In *Intelligence* 12: 1-16
[4] See for instance Keating, D.P. (1984). The emperor's new clothes: The "new look" in intelligence research. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 1-45). London: Erlbaum.
[5] Eysenc 1988
[6] See Chalmers 1996
[7] Tegmark 2017:37

as long as they produce results. Both positions end up concluding that attempts at reaching a shared definition is pointless.[8]My approach in this thesis clearly belongs in the former category, but my conclusion is different. We are indeed dealing with different aspects of intelligence in different contexts, and a single definition is of little practical use, but this does not have to mean that the concept has to remain confusing and ambiguous.

My own method and approach

My initial research into the use of the term intelligence revealed a great variety of conceptions, but at the same time general structures and patterns emerged. Some were treating intelligence as a somewhat fundamental abstract property,[9] others were viewing it as a complex phenomenon encompassing many other essential concepts.[10] Some were insisting that the right sort of structure had every right to be called intelligent,[11] others denied that intelligence could refer to something that is not conscious.[12] Some were open to the possibility of there being infinitely many ways of being intelligent, others had a much more restrictive view on what meaningfully fits the description. Although there, to my knowledge, has been no attempts so far at structuring all of these into a coherent picture of everything intelligence refers to in different contexts, it struck me as very possible to do. Based on my initial research on the use of the concept of intelligence in fields like psychology, philosophy, biology and artificial intelligence, I started categorizing the various conceptions according to what I perceived to be their essential differences in meaning and applicability, which in turn became the starting point for my thesis. Through conceptual analysis (in a broad sense), which typically aims to gain better knowledge of the language we use, [13] I explored the ways the many different conceptions of intelligence was considered to relate to tangential concepts like understanding and knowledge, seeking to both clarify and broaden our conceptual theory regarding intelligence. The main methodological tool at hand to modify existing concepts is

---

[8] Wang, Pei (2008) *What do you mean by «AI»?* Proceedings of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference

[9] Common in the field of machine learning, see Domingos, Pedro (2015) *The Master Algorithm* London: Penguin Random House

[10] For instance in developmental psychology, see Butterworth, George "Infant Intelligence" in Khalfa (ed) 1994:49-71

[11] For example Gregory, Richard (1970). *The Intelligent Eye.* Mcgraw-Hill

[12] Notably Searle, John (1990) "Is the Brain a Digital Computer?" In *Proceedings and Addresses of the American Philosophical Association,* Vol. 64, No. 3 (Nov. 1990):21-37

[13] See Kosterec, Milos (2016) "Methods of Conceptual Analysis" In *Filozofia* 71, 2016, No. 3: 220-230

that of distinction, often seen as the prime instrument for removing inconsistency in philosophy.[14]

My solution is to propose a framework for concepts of intelligence. What they have in common, my general definition of intelligence, is this: (D1) *Intelligence is (capacity for) complex problem solving.* It is possible to provide some further definition of what this means. In psychology, it has been suggested that complex problem solving means reducing the barrier between a given start state and an intended goal state with the help of cognitive abilities and behaviour[15]. When we attempt a general description of intelligence, the concept of cognition as a premise is less useful, but some complexity seems necessary for a problem solving process to qualify as intelligent. Very generally speaking, a capacity for complex problem solving includes being able to handle a high number of variables, with high connectivity, and their internal dynamics. However, this capacity exists on several qualitatively different emergent levels, from simple algorithms to the complexity of the human mind, and referring to the property in each instance as simply "intelligence" obfuscates essential differences. For this reason, I am introducing four different concepts of intelligence, which all describe complex problem solving, but that otherwise have quite different meanings.

The first is *homophenomenal intelligence*, which refers to the advanced problem solving capacity of humans: (D2) *Homophenomenal intelligence is (capacity for) complex problem solving based on conceptual understanding.* The second is *heterophenomenal intelligence*, which is applicable to all conscious entities: (D3) *Heterophenomenal intelligence is (capacity for) conscious complex problem solving.* The third is system intelligence, which is a property of (physical) systems: (D4*) System intelligence is (the capacity of) complex problem solving systems.* The final concept I will be applying is *algorithmic intelligence*, which refers to intelligence at the most atomic, fundamental level: (D5) *Algorithmic intelligence is the capacity of complex problem solving functions.* I have parenthesized 'capacity for' to emphasize that we may be describing both an inherent potential for intelligence and intelligent actions and behaviour. In the case of algorithmic intelligence, where complex problem solving *is* the function, the total capacity of the function is identical to its problem solving capacity. Together, these make up a conceptual framework of

---

[14] Rescher, Nicholas (2001) *Philosophical Reasoning. A study in the methodology of philosophizing* Oxford: Blackwell:116

[15] Funke, Joachim. (2012). "Complex Problem Solving." In Seel, N. (Ed.) *Open Learning Environments* Springer 2012:682-685

intelligence. Like other theoretical frameworks, it is not inherently 'true', and cannot be neither proven nor disproven. Quantum mechanics is a well-known example of such a conceptual framework that simply provides results when applied. All of my definitions are purely stipulative. The question that needs answering is this: Are these concepts useful? I believe that they are, and my complex goal for the remainder of this text is to convince you, the reader, of that.

For each of these concepts of intelligence, there are both proponents and critics among philosophers, psychologists, biologists or AI-scientists. Although many insist that only one of these concepts, or at least not all of them, is what we really mean by intelligence, I believe there are merits to including them all in a coherent theoretical framework for quite different concepts of intelligence. I will treat each concept in a separate chapter, showing both its usefulness and applicability, as well as its criticisms. The chapters are thus both a representation of the various existing views on intelligence, and my own effort to structure and define what I perceive as four quite different intelligence concepts. The resulting theoretical framework- what differentiates them, and how they relate to each other- is what I consider my contribution to be. My own philosophical views on various debated issues that this thesis touches will no doubt be apparent through both my selected literature and my formulations. I tend toward a materialist worldview, and my pragmatic attitude when it comes to epistemology and science has no doubt already surfaced by now. However, the nature of this thesis, that aims to show widely different conceptions of intelligence as equally valid, is such that I do not anticipate any strong philosophical objections to whatever personal opinions I have on consciousness, ethics and other relevant subjects. The fact that concepts and themes like language and ethics become relevant in the context of some concepts of intelligence, but not in others, means that subjects like these will appear, but not be explored beyond how they relate to intelligence. The point of contention, as I see it, is the framework itself; whether it is accurate, coherent, whether it includes the right sort of categories, and above all, as mentioned, whether it is of any use.

The first chapter is devoted to the concept of homophenomenal intelligence. (D2) *Homophenomenal intelligence is (capacity for) complex problem solving based on conceptual understanding.* 'Phenomenal' refers to the role of experience and *qualia*, while 'homo' emphasizes both the distinctly human nature of this kind of intelligence, and its homogenous specificity. Central issues are the importance of language and the existence of a general intelligence in humans. The first part of the chapter deals with the nature of human intelligence, and what sets it apart from other things we also at times refer to as intelligent, like animals. What do we mean when we say that there is a 'general' kind of intelligence that is dependent upon knowledge, beliefs, mental content and understanding, and that is often thought to be exclusive to humans? I will examine how all of these concepts relate to specifically homophenomenal intelligence. The second part of the chapter explores the many objections to the claim that there exists such a general intelligence trait in humans, many of which, we shall see, are ethically motivated.

The second chapter is about the concept of heterophenomenal intelligence. (D3) *Heterophenomenal intelligence is (capacity for) conscious complex problem solving.* In contrast to homophenomenal intelligence, 'hetero' emphasizes that we are dealing with multiple incomparable ways of being consciously intelligent. Research on animal mentality and behaviour is central to the arguments. The first part deals with objections to the clear separation of human and animal intelligence, and examines the empirical evidence relevant to this issue. The second part of the chapter emphasizes reasons to avoid conflating this human intelligence with other kinds of intelligence, in spite of eventual similarities. The last part of the chapter attempts to generalize the concept of heterophenomenal intelligence. What can we know and meaningfully say about completely different ways of being intelligent than the one we make use of when speaking, reading and writing?

The third chapter is concerned with system intelligence. (D4*) System intelligence is (the capacity of) complex problem solving systems.* What can intelligence mean, if we take the intentional conscious actor out of the equation? Here, the science of artificial intelligence becomes a primary source. The first part of the chapter is devoted to the description of humans and animals as systems. What is a system, and what does being intelligent mean, if we are not talking about *someone being there* and experiencing it? The second part attempts to describe what constitutes an *intelligent* system in general. What are necessary and sufficient premises for ascribing a system the property of intelligence? In what sense do a system care about the things it 'knows'?

The fourth chapter addresses the last concept of intelligence, algorithmic intelligence. (D5) *Algorithmic intelligence is the capacity of complex problem solving functions.* We are now dealing with intelligence as a potential of abstract objects, algorithms and functions, but also with concrete real-world consequences of theorizing about and developing them. Naturally, much of the ideas discussed in this chapter have their roots in the field of AI. The first part of the chapter is an attempt to explain in very general terms the idea of intelligent functions. What is an algorithm, and how can some algorithms be said to be intelligent? The second part of the chapter leans heavily on the concept of a “master algorithm” developed by Pedro Domingos. The branch of AI called machine learning is fundamentally based on the idea that intelligent functions are sufficient to create (almost?) any problem solving object/entity, and the many suggested approaches to creating such an algorithm represents differing views on what the nature of algorithmic intelligence is.

The fifth and final chapter is both a summary of the main conclusions of the other chapters and a further elaboration on the potential uses of the framework. The first part sets out to put the entire conceptual model in perspective, and clarify the relationship between the different concepts. The second part aims to more explicitly show that the framework can structure discussions about intelligence that until now have been marred by ambiguity, vagueness and misunderstanding.

Homophenomenal intelligence

(D2) ***Homophenomenal intelligence is (capacity for) complex problem solving based on conceptual understanding.***

Let us begin in familiar territory. *Homophenomenal intelligence* is more or less what most people refer to when they talk about 'intelligence' in their daily lives. People are intelligent to varying degrees, animals are not really intelligent, and unconscious objects definitely not at all. We humans seem to be uniquely intelligent when we ourselves reflect upon the matter, and our total global dominance as a species is a testament to our problem solving abilities. Not only that, our intuition is that every single normally functioning member of humankind possesses this special intelligence to a quite similar degree. Our ethics and morals are often in large part based on the notion that all human partake in this intelligent community, as 'rational actors', 'persons', or similar term describing this membership.[16] The concept of homophenomenal intelligence is in some sense paradoxical. It is specific, describing a single particular kind of intelligence that varies measurably, yet at the same time thought to be a universal property of all humans, at least in the ways it ethically counts. As a result, it is often this precise kind of intelligence that are of interest to us in matters of ethics concerning other species or artificial intelligence, while it is considered mostly irrelevant in questions of human ethics.

A homophenomenal concept of intelligence does not entail that only humans are intelligent, though that is often the conclusion, but that entities are intelligent to the extent that they possess the cognitive abilities we attribute to human intelligence. The goal of this chapter is to clarify the characteristics of this particular concept of intelligence. The core thesis is that there is a qualitatively different concept of intelligence that mainly applies to humans, and that is based on understanding. This understanding is a product of language and conscious experience, and it results in a specific kind of knowledge. It also has ethical implications, as it seems to be a premise for moral autonomy.

[16] The Kantian moral imperative of treating every human being as an end in himself or herself because of their ability to reason, is illustrative of this stance. See Kant (revised ed.) 2012

The exclusiveness of homophenomenal intelligence

As human beings, we are accustomed to speak of being and acting in the world in terms of thinking, understanding, believing, knowing and intending, and it appears that it is abilities like these that make us uniquely intelligent. Are we alone in this capacity? The answer to this question hinges on the difference consciousness and language make. No other species has developed comparable language faculties. Is there a kind of intelligence that is fundamentally about thoughts and knowledge that require language? Artificial systems may operate in complex machine languages, but for now, they do so unconsciously. They have no understanding of what they are doing, no mental imagery accompanying and guiding their problem solving efforts. Are there important differences between conscious intelligence and non-conscious intelligence? If the answer to these questions is yes, then there is a very specific concept of intelligence that is exclusive to humans, other animals that turn out to have some kind of mental language, and artificial systems that are conscious. It is an intelligence that does not mechanically solve complex problems, but carry out the task with intention, based on experienced beliefs about the world.

In the modern world, we are so accustomed to the thought of humans being cognitively unique that the idea of a human-specific intelligence seems self-explanatory. This has not always been the case. In our distant past, when humans in general possessed a much more animalistic worldview, talk of intelligence as a matter of beliefs and intentions would not have narrowed the concept to a human phenomenon in the same way. The prehistoric man would see this kind of intelligence all around him when pondering how the world works. In his teleological worldview, there were reasoning behind animal behaviour, the life of trees and plants, and even the forces of nature. Floods and fires intended their results. Science, however, aims specifically at replacing these kinds of answers, narrowing down what counts as intelligence.

Jean Khalfa, along with several other well-known philosophers and scientists, published in 1994 a collection of essays under the title *What is Intelligence?*. Khalfa emphasized the involvement of reasons as opposed to causes as the core of the concept intelligence. Before the seventeenth century, the universe seemed to be driven by reasons, with each thing having its own reason or finality. After the transition to Newton’s mechanical universe, intelligence had to either be entirely isolated or explained away. The dualists postulated it as a faculty exclusive to beings ruled by an immaterial substance, a soul. If the dualist point of view is right, only beings whose behaviour implies the actions of an irreducible autonomous soul can be called intelligent. Moreover, with intelligence, it was reasoned, comes moral responsibility. Much of our ethics thus became based on assumptions regarding intelligence. The opposite approach tried to eliminate this residue of a 'prescientific' conception of the world, by reducing the reasons to causes.[17] The general concept of intelligence had become entangled in the debate on dualism, not the last of the controversies and disagreements affecting its usage, but arguably the one that runs the deepest. It seemed that intelligence either had to be a property of something non-physical, or a lacklustre explanation of cause and effect.

When Alan Turing listed possible objections to the idea of thinking machines in his pioneering paper "Computing Machinery and Intelligence", the first mentioned was "the Theological Objection": "Thinking is a function of man's immortal soul. God has given an immortal soul to every man and woman, but not to any other animal or to machines. Hence no animal or machine can think."[18] Turing did not expand significantly on this, except to say he did not at all accept the view. In the same manner, I will not delve further into theology, but remark that religious arguments for the exclusiveness of human intelligence are commonly expressed, though perhaps not as often a part of scientific debates on animal and artificial intelligence. The general influence, however, of the core monotheistic idea that the human mind is divinely bestowed upon mankind in the image of the creator of all things, is not to be underestimated. Attempts to nuance the categorical separation of human minds and minds of other beings fundamentally conflict with many religious teachings. There exist published theological attempts at showing a non-conflict between genuine artificial intelligence and the biblical teachings of the soul,[19] but in a clear majority of cases, the use of the concept 'soul' seems to entail a very restrictive homophenomenal concept of intelligence. Reason is reserved humans, and other non-human behaviour that appears to imply a similar capacity is to be interpreted as a sort of 'as-if' intelligent behaviour. What is present when a human solves a problem, but not in any other case? Well, in addition to simply performing the correct sequence of actions, humans at times *understand* what they are doing, and have a sense of reasons for doing so. It is this understanding that is unique and central to homophenomenal intelligence.

[17] Khalfa, Jean (ed.) (1994) *What is Intelligence?* New York: Cambridge University Press
[18] Turing, Alan (1950) "Computing Machinery and Intelligence" in *Mind* 59: 433–6
[19] See Bjork, Russel (2008) "Artificial Intelligence and the Soul" In *Perspectives on Science and Christian Faith* Volume 60, Number 2:95-102

Understanding: consciousness and mental content

What is understanding? In all human scientific inquiry, it seems to be the stated goal. Our intellectual efforts throughout the ages have not been aimed at simply solving complex problems for the purpose of making our lives easier, like inventing the wheel or developing vaccines, but ultimately to know for the sake of knowing. We long for understanding that has no discernible practical applicability in our lives. How did the universe come to be? What is the origin of life? What was everyday life like in the 16th century? Whatever understanding is, it seems to be concerned with many things we would not normally define as 'problems'. In epistemology, philosophers try to characterize how it differs from knowledge and wisdom. In the philosophy of language, the effort is to describe what is involved in understanding words and sentences. The philosophy of mind is similarly interested in how we understand concepts. Are there many different, and to an extent unrelated kinds of understanding, or do they all have something in common? One aspect that seems to be universal is that understanding involves the understood relationship, structure or event 'revealing itself' to our minds' eye. When we understand, we somehow 'see' or 'grasp'. Understanding takes place in the realm of the phenomenal.

Humans can perform abstract calculations, and animals cannot. Calculators can, however, but we rightfully hesitate to say that they are intelligent in the same way. A calculator may 'know' in some sense how to perform all the required computations, but it does not understand what it is doing. What is the difference? When we speak of understanding, we are not simply talking about the ability to execute an action, but in addition, a subjective experience of the knowledge required to do so. Experience is the realm of consciousness, an aspect of the human mind that is directly accessible to us, yet notoriously hard to integrate with everything else we know about the world. How the physical world can give rise to subjective experience, and how this ethereal experience can interact with the world, has remained a mystery throughout the history of philosophy. Descartes famously declared that the existence of subjective experience was the one thing he could not doubt, but had to resort to a rather dubious thesis about interactions going on in the pineal gland to explain how the soul, the experience, effects the otherwise purely mechanical body and outside world.[20] Though several philosophers (Churchland, Dennett, among others) have tried to deny the obviousness of the existence of subjective experiences, most remain unconvinced that this is possible.[21]Philosophers generally agree that David Chalmers formulation of the "hard problem" of consciousness remains unanswered: subjective experience, or qualia, exists, and we do not know why.[22]The mental processes of humans have an accompanying experience; they have in a very specific sense content. The exact nature of this content is a contested issue in philosophy, but most would agree that consciousness is *about* things. This 'aboutness' is often described as intentionality, where intentionality refers to pretty much the same as having content.[23] The term intentionality is derived from Aristotle's concept of 'mental inexistence'; the notion that, when thinking of an object, one has that object *in* mind, but it does not exist like objects in the real world exists. The content is non-local. According to phenomenologist like Husserl, the basic character of intentionality is the property of being conscious of something.[24] When humans solve problems, their doing so contains a subjective mental state, an experience of or about the problem. The content of this experience we refer to as our understanding. It thus appears that understanding is defined as non-physical. Intentional relationships can exist between two objects that have no proximity to each other in either time or space, like when you picture Jupiter, and there are no physical descriptions available for this interaction. It follows that in structural descriptions of intelligence, when we are talking about the intelligence of systems in terms of physical causes and effects, understanding is ruled out as a proper concept.

In the philosophical tradition of Descartes, intelligence was entrenched as a unique property of the immaterial human soul and mind, in stark contrast to the clockwork mechanisms governing the rest of the universe. Many philosophers seeking to challenge this dualism, and treat the human mind as an integral part of nature's order, still had a desire to distinguish between things that *behaved* intelligently, and things that were genuinely intelligent, meaning possessing human mental capacities. Consciousness was necessary for intelligence, but not sufficient- the two words are not synonymous. Mental content became a key feature of homophenomenal concepts of intelligence, and in explaining how it is unique

---

[20] Examined in Finger, Stanley (1995) "Descartes and the pineal gland in animals: A frequent misinterpretation" In *Journal of the History of the Neurosciences*, 4:3-4, 166-182
[21] See Churchland 1992, Dennett 1987
[22] Chalmers 1996
[23] Siewert, Charles, (2017) "Consciousness and Intentionality", *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/consciousness-intentionality/ 20.03.2019
[24] Chapman/Routledge 2009:101

to humans.

John Searle's famous Chinese Room Argument was an argument against the computational model of the mind, seeking to demonstrate that syntax is not the same as, or by itself sufficient for, semantics. In the thought experiment he applied to convey the argument, Searle put himself in a closed room translating Chinese instructions using a script, thereby fulfilling a function as translator to perfection without any real *understanding* of Chinese. The analogy being, of course, a computer. Just like Searle in the thought experiment would be performing the intelligent task of translating, without any comprehension of what the messages he was translating *meant,* so would be the case for any merely computational machine capable of actually understanding what it was doing. Mental content accompanying the computations would qualify as "Strong AI", a phenomenon he did not imagine being realized in the foreseeable future. By this, he no doubt meant *real* intelligent artificial intelligence, whereas Weak AI for Searle was something of a misnomer, describing simulations of actual intelligent processes. Searle's view on intelligence and computation, which quickly became influential, is that there is no comprehension at all without consciousness. Mental content is the very essence of intelligence.[25]

The problem with this attitude towards 'understanding' is that it does little to explain how it affects anything at all. If the actions and results of Searle's performance in the Chinese room are identical to the results of someone who understands Chinese, then whether or not the room with Searle inside really understands the meaning of what is going on is of no practical concern. How should we define "understanding" if its presence has no measurable effects? Many researchers in the field of AI remain unconvinced of Searle's arguments, simply noting that if 'understanding' really does have a measurable impact on an agent's performance in some well-defined situations, then it is of interest to us.[26]Those with an interest in the philosophy of mind will no doubt spot the contours of familiar debates on epiphenomenalism and consciousness. The concept of homophenomenal intelligence would indeed be in danger of being superfluous if it could be summed up as intelligence with consciousness tacked on, where the conscious aspect accounts for none of the results. However, I will not be arguing for or against the causal powers of qualia, the framework of intelligences is not a theory of mind. For our purposes, a functional description of the role of understanding in human

[25] Expanded upon in Searle, John (1990) "Is the Brain a Digital Computer?" In *Proceedings and Addresses of the American Philosophical Association,* Vol. 64, No. 3 (Nov., 1990):21-37
[26] Legg, S. & Hutter, M. (2007) «Universal Intelligence: A definition of Machine Intelligence» In *Minds & Machines* 2007 Vol 17: 391-444.

intelligence will suffice. Without doubt, whatever understanding *really* is, it is a characteristic aspect of how humans go about solving complex problems. Animals are conscious as well, but I propose they still do not understand as we do. Consciousness is necessary for understanding, but not sufficient. Language makes a difference.

Understanding: Language

What makes humans understand is not just the fact that we somehow visualize and experience problems. Our experience of problems can be broken down into abstract concepts which we structure using our language. To a human, there seems to be a strong connection between language and consciousness itself. According to Searle, intentional states represents objects and states of affairs in the exact same sense that speech acts represent objects and states of affairs.[27]The position of language in theories of human cognition and mental abilities have been absolutely central, especially since the so-called linguistic turn in Anglo-American philosophy in the mid nineteenth century, where language came to be seen as *the* medium of conceptualization.[28]

What is language? In most contexts language refers to a culturally specific communication system (like English), but in linguistics and the philosophy of language we are talking about an internal component of the brain. This component, the language faculty, is often viewed as unique to humans. Linguists have distinguished between a narrow understanding of this faculty of language, which is the abstract, computational linguistic system alone, and a broader sense, which in addition includes the sensory-motor system and a 'conceptual-intentional' system. The latter, which is thought of as the enabler of conceptual representations, appears to function at least in nonhuman primates. Chimpanzees know who is related to whom, and how the dominance hierarchies are structured. [29]It is hard to pinpoint precisely where grunts and simple mating calls becomes speech, but it is not difficult to see the world of difference the fully-fledged human language has made. Our communicable concepts are transferable to speech and writing, making the learning process not just something that goes on inside an individual or even the local community, but an incremental process for the ages. Chomsky, among others, have also noted the freedom from identifiable

[27] Searle, John (1982) "What is an intentional state?" In H.L: Dreyfus with H. Hall (ed.) *Husserl, Intentionality, and Cognitive Science.* Cambridge, MA and London: MIT Press
[28] Wolf, Michael "Philosophy of Language" In *Internet Encyclopedia of Philosophy* https://www.iep.utm.edu/lang-phi/#H1 20.03.2019
[29] Hauser, Mark, Chomsky, N., Fitch, T. (2002) "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" In *Science* 22 Nov 2002:1569-1579

stimulus control that seems to be a property of the human language. Human language is *creative*. Our descriptions are appropriate to the situation but not caused by it, is the thesis.[30] Upon hearing a particular classic composition, you might utter the words “Ah, Bach.” The general meaning of the words no doubt was caused by the impressions made by the sounds, but there are countless similar sentences that, all things considered, were equally probable, for instance: “Beautiful, isn’t it?” or “This is my favourite fugue.” Language appears to free our intelligence from the mechanisms of nature, making it a less deterministic phenomenon then other kinds of intelligence.

Philosophers have not been hindered by this, and have sought to explain the structures in the brain that might account for human thought. Inspired by Chomsky, Jerry Fodor’s *language of thought* hypothesis postulates that thinking is done in a mental language, “Mentalese”, in a symbolic system physically realized in brains. According to the theory, when someone has the belief that P, the “object” of the belief is a complex symbol that is physically realized in the neurophysiology of the brain that has both syntactic structure and semantic content: the proposition that P.[31] Although Fodor was mainly trying to explain how anything material could have semantical properties, his account was also an argument for the necessity of language for thoughts. Note that this was no radical new philosophical standpoint- similar ideas had been expressed centuries earlier by Leibniz. In his view, human cognition was essentially symbolic, taking place in a system of representations possessing a language-like structure.[32]A purely physical description of what goes on in human speech, however, seems unsatisfying. When we put our language to use, when we communicate, we appeal to the phenomenal, to understanding.

According to cognitive science, human communication consists of encoding and decoding information. Failures can be due to encoding or decoding errors, or noise. This is neither inferential nor creative. However, communication is filled with implicit content, like what is meant in different contexts by saying “It’s late”. Do we mean to imply that we are tired, or are we talking about a train that is not showing up? The thought we intend to convey can never be fully encoded, and linguistic decoding is only a first step in understanding. Our meta-representational ability to infer intentions, by attributing mental states to others, are

---

[30] Chomsky, Noam (1959) «Review of *Verbal Behaviour* by B.F. Skinner» in *Language* 35: 26-58

[31] Aydede, Murat, "The Language of Thought Hypothesis", *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/language-thought/ 20.03.2019

[32] Kulstad, Mark and Carlin, Laurence, "Leibniz's Philosophy of Mind", *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2013/entries/leibniz-mind/ 20.03.2019

crucial.[33]The phenomenal aspects of human intelligence play a role not only in the individual's processing of information, but also in the transfer of information between us. There are however arguments to be made that language and thought are far from the same thing, among them evidence from neuroscience. Individuals with global aphasia, who have a near-total loss of language, are still able to add and subtract, solve logic problems, appreciate music and art, and navigate their environment. Neuroimaging studies show that also healthy humans engage the language areas of the brain when performing linguistic tasks like reading, but not when doing arithmetic or listening to music. The conclusion of these studies is that many aspects of thought do not depend on language.[34]This should make us cautious about some of the strong claims of the uniqueness of language-capable intelligence. Does it not seem likely that animals and other conscious entities after all might understand a great deal, even without a functioning language?

Some kinds of thoughts do however seem to be reserved individuals with functioning language capacities. According to Donald Davidson, we have the idea of *belief* only from the role of belief in the interpretation of language.[35] If this is true, a creature must be a member of a speech community to have the concept of belief. This is considered a strong version of the idea that language is imperative to cognitive functions, but also milder variants (for instance the claim that language is the medium of *conscious* propositional thinking) are still pointing towards a homophenomenal concept of intelligence. This relationship between cognition and language is also familiar from the everyday lives of humans. We engage in 'inner speech' throughout the day, and this imaged natural language occupy much of our stream of conscious mentality.[36]This by itself, however, only shows that human intelligence commonly is utilized by applying our language abilities to mental tasks. This, it seems, is not what people like Davidson mean to say. When we claim that interpretation of language is necessary to have any beliefs at all, then we are arguing that only phenomenal experiences involving language can be described as *knowing* or *understanding*.

This seems, on the face of it, too strong a claim when considering other animals. Few philosophers have problems ascribing animals some form of consciousness. Are they then conscious without knowing anything at all? If the limits of our language is the limits of our

---

[33] Sperber, Dan «Understanding verbal understanding», also in Khalfa (ed) 1994:179-198

[34] Fedorenko, Evalina, Varley, R. (2016) "Language and thought are not the same thing: evidence from neuroimaging and neurological patients» in *Annals of the New York Academy of Sciences*. Issue: The Year in Cognitive Neuroscience 2016: 132-153

[35] Davidson, Donald "Thought and talk." In: Guttenplan, S. (Ed) Mind and language. 1975:7–24

[36] Carruthers, P. (2002). "The cognitive functions of language." In *Behavioral and Brain Sciences*, 25(6), 657-74; discussion 674-725.

world, as Wittgenstein claimed,[37] then it is a small life-world indeed for non-speaking entities. When defending the view of human exceptionalism regarding phenomenal intelligence, the tradition has been to separate mental representations into two broad categories: perceptual representations and symbolic conceptual representations. Of these, only the latter can be integrated into a larger set of representations and thus structure intelligence, while the former consists only of the superposition of spatial images on spatial perceptions- a "proto-thought"[38]. It follows that the structure of propositions that make up knowledge is made possible only by representations that are symbolic. That seems to imply that animals have proto-thoughts, but, lacking symbolic concepts, cannot structure them into knowledge. At least, not the kind of knowledge that we are describing here. The sentiment is that human speech should not be viewed as just a symptom of our cognitive abilities, but also as something that facilitates thought processes. Our language capacity allows us to compose our intellectual efforts into a specific kind of knowledge. We will get to other possible ways of 'knowing' in later chapters, but let us first examine this language-based knowledge.

Understanding: Knowledge

The philosophical account of knowledge has traditionally been based on Plato's notion of *justified true belief.* One does not know that P unless one believes that P. In recent years it has been suggested that additional premises are necessary, narrowing the concept even further.[39] In conjunction with the view that language is necessary for belief, or that we need semantics to talk about believing, it seems to follow that knowledge is reserved conscious entities with language, i.e. humans. Is it the case then, that the homophenomenal concept of intelligence is the only concept of intelligence that deals with knowledge in a meaningful way? Can knowledge be said to be the hallmark of homophenomenal intelligence?

In the field of psychology, the perceived relationship between intelligence, knowledge and language is a very close one. Developmental psychology attempts to explain intelligence by observing its origins and how it grows. When professor of psychology George Butterworth summarized the useful definitions of intelligence in this context, he eschewed defining it as a single, underlying intellectual ability. Intelligence as the faculty or capacity of knowing, and intelligence as what is expressed in adaptive behaviour in particular contexts, were his

---

[37] Wittgenstein, Ludwig (1999) (1922) *Tractatus Logico-Philosophicus* Trondheim: Gyldendal
[38] Dummett, M. (1994). *Origins of analytical philosophy*. Cambridge: Harvard University Press.
[39] Gettier, Edmund (1963) "Is Justified True Belief Knowledge?" In *Analysis*, Vol. 23, No. 6, 1963: *121-123*

proposed working definitions for the field of study.[40]This does paint a picture of a separate concept of intelligence that more or less equals knowledge, and which might apply exclusively to humans. Human knowledge is based upon language as a developing cognitive system, which is more powerful than other cognitive systems, and which results in qualitatively different cognitive abilities in adult humans[41]. Thus, the developmental process becomes the object of study when explaining human uniqueness. How do humans develop the understanding kind of knowledge?

The most widely accepted theory of the origins of human intelligence is that of Jean Piaget. The roots of intelligence, he says, lie in programmes of action that form the first link between the baby and the world, the reflexes, and proceeds in a series of stages as the baby adapts actions through encounters with objects. From roots in reflexes, intelligence proceeds through habitual forms of action, to intentional use of means to achieve certain ends. Actions organizes and gives structure to perception, and eventually lends structure to thought and language.[42] Here, the process of gaining knowledge and that of evolving intelligence is considered more or less the same thing. Modern cognitive-based approaches to the subject of acquiring knowledge view knowledge as the fruit of intelligent cognitive processes, the manifestation of applied intelligence. [43]

According to Piaget, language is in a sense a result of intelligence applied over time. This is contested by the Chomskyan rationalists' claim that our language evolves mostly because we are born with an innate set of universal linguistic principles, which are not acquired through external stimuli. The argument for this is that the outside stimulus is too impoverished to account for the complexity of adult linguistic, and it is sometimes even formulated as a logical problem of needing knowledge to gain knowledge.[44] This would lead us to conclude that language is as much of a premise for intelligence as a result of it. The *Language-based theory of Learning* (LTL) in the tradition of Vygotsky similarly emphasizes that learning language and learning through language are simultaneous processes. When children learn a language, they are not simply engaging in one type of learning among many,

[40] Butterworth, George "Infant Intelligence" in Khalfa (ed) 1994:49-71
[41] Hughes, T.J. & Miller, J.T.M.(2014) "Lexicalisation and the Origin of the Human Mind" In *Biosemiotics* (2014) 7: 11.
[42] Butterworth 1994:49-71
[43] See Drigas, Athanasios & Pappas, Marios. (2017). The Consciousness-Intelligence-Knowledge Pyramid: An 8x8 Layer Model. International Journal of Recent Contributions from Engineering, Science & IT (iJES). VOL 5. 14-25. 10.3991/ijes.v5i3.7680.
[44] Chapman/Routledge 2009:50

but are acquiring the foundation of learning itself.[45] More than just a ‘chicken and egg’ question of whether human thought is a result of human language or vice versa, this latter view establishes an even stronger position on the uniqueness of human intelligence. Without language, we do not simply learn less, we lack the premises for true learning.

Is knowledge for that reason exclusive to entities with language? According to LTL, language is the essential condition of knowing, it is the process that turns experience into knowledge.[46] Defining knowledge as justified true belief likewise seems to establish the concept as exclusively connected with homophenomenal intelligence. However, there are other approaches to the concept, also in the context of humans. Professor of Neuropsychology Richard Gregory is critical of the relationship between intelligence and knowledge that Piaget described. People regard Einstein as exceptionally intelligent because what he said was not already known, Gregory writes. This is not the intelligence of existing knowledge; it is the intelligence of discovering or creating new knowledge. Gregory dubs the intelligence of knowledge *potential intelligence,* and the problem solving kind *kinetic intelligence,* borrowing from the vocabulary of physics. Each step of historical problem solving requires kinetic intelligence, building up and storing potential intelligence. Natural selection is a powerful kinetic intelligence, says Gregory, producing the immense amount of potential intelligence stored in our bodies. We should not expect processes to appear intelligent, he remarks, intelligence is the result of processes. Neither Darwin nor Einstein knew where their theories would lead- “blindness” should not refrain us from calling natural selection intelligent.[47]

Gregory’s kinetic and potential intelligence terms de-anthropomorphize the concept of knowledge to potential of intelligence. Other approaches similarly treat human knowledge as something other than justified true beliefs, or pertaining to conscious humans. John Horn and Raymond Cattell introduced in the 1960’s the distinction between fluid and crystallized intelligence. Fluid intelligence, they stated, is the ability to solve novel problems, or solve problems within time limits. Crystallized intelligence is the stored knowledge, the accumulated product of the system’s intelligence.[48]It seems clear that this view of knowledge makes the term applicable when describing the actions and behaviour of species without language as well as artificial systems, showing that the concept of knowledge can be relevant

---

[45] Wells, Gordon (1994) «The Complementary Contributions of Halliday and Vygotsky to a “Language-based Theory of Learning” In *Linguistics and Education* 6, 1994:41-90
[46] Wells 1994
[47] Gregory, Richard “Seeing intelligence” in Khalfa (ed) 1994:13-25
[48] Deary 2001:33

to intelligence even when no mental content or consciousness is being ascribed to the studied phenomenon. However, we are dealing with separate concepts of knowledge, closely connected to the different concepts of intelligence, and it seems clear that in the case of some special kind of intelligence, that involves mental content, language is a prerequisite.

The question of innateness, the degree to which language is learned through experience, remains unresolved. Language may be innate or acquired, but it is commonly held that it facilitates specific kinds of beliefs and knowledge, making homophenomenal intelligence qualitatively different from other kinds of intelligence.

Intelligence and moral autonomy

Even though the exact relationship between human language and human intelligence is not clear, language enables an intelligence that *understands*. Most notably, conceptual understanding is a prerequisite for distinguishing between the concepts of ‘right’ and ‘wrong’. As a consequence, the framework of different intelligences carries some apparent ethical implications that, though not central, are still well worth mentioning. In this context, we touch upon the concept of moral autonomy. The general concept of autonomy is related to intelligence in several ways. Understood as the capacity to be one’s own person, to live one’s life according to reasons and motives that are taken to be one’s own and not a product of manipulative or distorting external forces, it is central in moral and political theory.[49] Acting autonomously involves being capable of preferring certain actions above others, and making choices. Issues of determinism and free will aside, we should be able to describe an intelligent action as not wholly determined by factors external to the acting agent. In this sense, animals seem to be autonomous in the same manner as humans, but an entire field of philosophy, ethics, is based on the premise that once you have conceptual understanding, your autonomy makes you morally responsible for your actions. Human babies and animals have do not have sufficient understanding, and even though they might have undiminished moral status, meaning it is considered just as morally wrong to mistreat them, they are not seen as moral actors. Because of their lack of conceptual understanding, they are incapable of autonomous decisions that can be judged as right or wrong in a moral sense. Of course, animals are fully capable of making disastrous or fatal choices, but we assign them no ethical implications. It is

[49] Christman, John, (2018) "Autonomy in Moral and Political Philosophy", In *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/autonomy-moral/ 20.03.2019

their nature, we conclude. In the food chain, there is only the survival of the fittest, ethics is reserved the realm of conceptual moral reasoning.

What this seems to suggest, is that homophenomenal intelligence is required to be morally responsible. An entity without consciousness does not *understand*, as we have seen, because mental content is so central to understanding. An animal might have some form of understanding, but without proper conceptual capacity, it is not able to differentiate the essential concepts of moral decisions. One might argue that, for instance, a wolf has some understanding of the consequences of stealing another member of the pack's food, and that it might have empathetic qualms about doing so, but then we are really arguing to what extent the wolf in question has the property of homophenomenal intelligence, not whether an entity without it can be morally responsible. We shall return to both ethical implications of intelligence and other concepts of knowledge in later chapters, but let us first examine the applicability of homophenomenal intelligence in a non-human context.

Homophenomenal intelligence and machines

It is not necessary to commit to any ontological status of beliefs, thoughts or understanding to speak of a separate type of intelligence that is fundamentally tied to human experience. Daniel Dennett famously articulated an attitude he called "the Intentional stance". By this he meant describing the system (biological or artificial) *as if* it has beliefs, intentions etc.[50] This is not inherently right or wrong. There are no real facts to the matter, Dennett reminds us, just ways of looking at a complex system that may or may not be useful. It does however open up for the possibility of ascribing homophenomenal intelligence solely to entities that act as if they have beliefs, without being fully committed to explain what beliefs are. When speaking of homophenomenal intelligence in non-human contexts, this approach might very well be useful. We will probably never know whether some entity in the future behaving indistinguishable from an intelligent human really consciously understands, but we may speculate in what kinds of actions and behaviour would indicate that it does, and in the case of artificial entities, how to go about creating them.

If we for the sake of argument assume Dennett's intentional stance regarding artificial systems, what kinds of beliefs can we attribute to them? Can machines ever be described as conscious? The term "AI" was introduced by John McCarthy at a conference at Dartmouth

[50] Dennett, Daniel C. (1987). *The Intentional Stance*. MIT Press.

College in 1956. The earliest well-established notion of artificial intelligence here introduced, is often referred to as “Good, old-fashioned AI”, or GOFAI.[51] When applied in AI research, it is often focused on symbols and words, more so than purely learning, because of the special role language plays in the human behaviour it sets out to mimic. GOFAI is thus a knowledge-based intelligence. What the system knows is stored in memory as symbolic expressions, and using rules of logic new symbolic representations can be derived from these. In a mechanically embodied intelligent process, an intentional stance is merited, if we view the symbolic structures as propositions that are believed by the system, and causes its behaviour.[52]

Roger Schank and Lawrence Birnbaum outlined three camps or positions on the question of whether truly intelligent machines, meaning machines that have homophenomenal intelligence, was possible. The first they named the “dog-consciousness position”, after John Searle’s strong conviction that a machine could never really be intelligent without consciousness, an attribute he was willing to ascribe to his dog. The second they called the “language organ position”, referring to Chomsky’s view that there is a kind of language organ that only humans have, but that we might initially equip machines with. The position they held themselves, they referred to as the “additive intelligence position”, grounded in the idea that intelligence is modifiable, if we can identify the rules that govern intelligent behaviour. “How can we hope to answer questions about whether machines can be intelligent without knowing what constitutes intelligence?” asked Schank and Birnbaum. The issue was rarely a part of the debate about possibilities for AI, and not often part of the debate about human capabilities either. The linguists talked about innate human capabilities and language organs, but not about their employment. Where AI research uniquely touches upon important aspects of intelligence, they wrote, is in addressing a model’s capability to carry out some intelligent task with the question “What is the bottleneck problem here?” rather than the other intelligence positions’ usual “How is this task different from all other tasks?”. The general bottleneck they came to acknowledge was the machine’s lack of knowledge. Knowledge had to be the general factor in intelligence, and what intelligent artificial systems would need. Chomsky and the linguists would deny this. Any kind of knowledge will not do- some special linguistic knowledge is required. For Searle, consciousness itself appeared to be the missing factor.[53]

---

[51] Levesque 2017:7
[52] Levesque 2017:117
[53] Schank, Roger and Lawrence Birnbaum “ Enhancing Intelligence” in Khalfa (ed) 1994:72-106

As is probably becoming clear, it is not very easy to distinguish artificial intelligence (AI) from artificial consciousness (AC), even though AC has been established as a separate sub-field. There still are some methods that are clear departures from those typically used in AI. For instance, AC researchers in robot navigation and planning are not simply interested in how the robots avoid collisions and find their way to a goal location, but also in exploring the extent to which the processes the robots employ can be viewed as instances of imagination and mental content. Ambitions vary greatly, from cautious attempts at weak AC, (approaches that make no claim of a relation between technology and consciousness) to strong AC, where the goal is the design of systems that when implemented are sufficient for consciousness. [54] Most of AC research, however, fall somewhere in between the two extremes. Weak AC is generally not very interesting, and strong claims of AC run into the criticisms of Searle and others, that consciousness cannot be entirely computational. However, *some* computational concepts could be part of the explanation of *some* mental phenomena. A result of this emerging nuancing beyond conscious/unconscious machines is the growth in the field of research referred to as *prosthetic AC.* Whereas most AC research is autonomous, aiming to create a self-sufficient artificial consciousness, prosthetic AC seek to alter or augment existing consciousness. Much like artificial limbs, technology may offer modules contributing directly to consciousness.

I have perhaps too dismissively evaded the question of how you actually know if you have succeeded in creating a consciousness. The more pragmatic approaches might not care about this, but serious scientific AC should probably have some methodological suggestions. We can attempt to model the physical system underlying consciousness, or model conscious processes directly by implementing its causal structure in an AI system. We can also model the behaviour of a system believed to be conscious, and hope that reproducing behaviour is enough to implement the consciousness-producing causal structure. Somewhere there might lie a convincing answer as to whether the resulting system is *really* conscious rather than just apparently conscious, but for now, it seems that non-pragmatic approaches to artificial consciousness are limited in their explanation power by our current theories of consciousness. In addition, no existing AI is considered a serious contender for AC. Does that mean that experiencing machines are impossible, or just really far off? Arguments against computationalism have been mentioned. Another strong objection to the idea of homophenomenal intelligence in machines is the claim that artificial intelligence is a

[54] Chrisley, Ron (2008) “Philosophical foundations of artificial consciousness” In *Artificial Intelligence in Medicine*, Volume 44, Issue 2, 2008: 119-137

contradiction in terms. To be artificial, the system needs to be not just the result of our labour (like our children) but also designed by us. This means that any purpose or intentionality in the system is derived from us, more or less the antithesis to the autonomous human mind.[55] Regardless of the strength we perceive these principal objections to AC to have, the fact remains that conscious intelligence seems to be another issue entirely than all the various intelligent tasks that machines are rapidly solving. It is a question of intelligence that is on another level of description, where we do not have the conceptual tools to even formulate what needs to be addressed, much less an artificial challenger to this last stronghold of human cognitive superiority.

The existence of general intelligence in humans

My account thus far of the use of the homophenomenal intelligence concept is likely to leave the impression that its sole function is making the case for intelligence as a (perhaps, *the)* characteristic trait of human beings. If that were the case, it would seem that there is some common ground in what intelligence is describing. Is the claim that there exists such a thing as human intelligence the one use of the concept intelligence that everyone can agree with? Even here, it is not that simple. In the years following the publishing of Darwin's *On The Origin of Species* in 1859, the work that finally cleared the ground for the biology of human beings to be put in a proper scientific context, the main concern of many scientists was to combine the insights of the theory of natural selection with a yet to be discovered theory of heredity. The main goal was to provide an account of the evolution of human intelligence, widely acknowledged as the single most important trait evolution had produced. From the very moment it was made possible to describe properties like intelligence in naturalistic terms, there was conflict between the notion that these traits could be further honed and maximized on the one hand, and ethics and values on the other. Darwin's own cousin, Francis Galton, coined the term *eugenics*- the betterment of the human race via artificial selection of genetic traits, and is also considered the originator (along with Darwin himself) behind the idea that intelligence is a stable, inherited ability.[56]Eugenics quickly caught on all over the industrialized world, with state-sponsored sterilization programs in the United States leading the effort to weed out the genes of "imbeciles" in the 1920's, but fell out of vogue just as fast

[55] Chrisley, Ron (Ed.) (2000*) Artificial intelligence: critical concepts.* London: Routledge
[56] An excellent account of the early influence of "On the Origin…" is provided in Mukherjee, Siddharta (2016) *The Gene. An Intimate History.* New York: Scribner.

after the ethical implications of the efforts of Nazi-Germany became apparent at the end of the Second World War.[57] The idea that there is such a thing as a human trait of intelligence, identifiable as a measurable quantity, through the various intensive testing of the eugenic programs, had become closely linked to politics of race in most people's minds.

The problems with defining intelligence unsurprisingly made the measuring of said property a difficult task. A breakthrough was made in 1904 by the English statistician Charles Spearman, who isolated what he called a "unitary mental factor" that varied consistently across a wide range of mental tests. He called this factor *g*, general intelligence.[58]The issue of whether there exists a single, preferably measurable, property that we can speak of as *g*, or simply 'intelligence', or if this concept is an oversimplification that describes what is in reality a plethora of related abilities and properties, is the essence of the debate on the *existence* of homophenomenal intelligence.

At the centre of debates on *g* today, we usually find Charles Murray and Richard Herrnstein's *The Bell Curve*, published in 1994.[59] The scepticism towards *g* had been intensified in the 1980's by the assertion that intelligence testing contained cultural biases against minorities, and at the same time undermined by Howard Gardner's multiple intelligences theory.[60] Gardner listed seven intelligences in *Frames of Mind* [61](1983): linguistic, logical-mathematical, musical, spatial, bodily kinaesthetic, self-oriented personal intelligence and other-directed personal intelligence. His reasoning for calling all these intelligences rather than talents, was that naming one set of abilities (IQ) intelligence and others, like dancing, talents, devalues a lot of able people, now seen as not smart, but dumb.[62]The reasoning behind the argument that intelligence is several loosely connected abilities thus had very up-front ethical connotations. *The Bell Curve*, on the other hand, was presented as a proof that there really is an objective, measurable human property called intelligence, that is significant, and that needs addressing, ethical implications be damned. It addressed IQ scores in the context of social outcomes and social policy, with the key question: Are persisting inequalities in American life due to the attributes of individual citizens, or to structural causes such as unequal educational and economic opportunities? Several relationships were explored, the truly controversial one being the relationship

---

[57] Mukherjee 2016:78-90
[58] Spearman, Charles (1927) The Abilities of Man, Their Nature and Measurement New York: Macmillan
[59] Murray, Charles and Herrnstein, R. (1994) *The Bell Curve. Intelligence and Class Structure in American Life* New York: Free Press
[60] Ma/Schapira 2017:23
[61] Gardner, Howard (1983) *Frames of Mind* New York: Basic Books
[62] Flynn 2009:204

between IQ and race.

The book sparked widespread political and public debate, and became highly influential in American conservative think tanks. Among its fiercest critics was the biologist Stephen Jay Gould. Gould's main complaint was against the subtext of biological determinism, or more precisely, what he saw as one of *The Bell Curve's* principal themes: the claim that *worth* could be assigned to individuals and groups by measuring intelligence as a single quantity.[63] Are social struggles upsetting the natural order? he asked. "Hereditarians", Gould elaborated, view their measures of intelligence as markers of permanent, inborn limits, and children should in their minds surely be trained accordingly. Gould sought to demonstrate both the scientific weaknesses and the political context of deterministic arguments. Intelligence is a wondrously complex and multifaceted set of human capabilities, he wrote. Through *reification*, our tendency to convert abstract concepts into entities, this socially defined concept achieves its dubious status as a unitary thing. The single number a measurement of this fallacious quantity provides lends itself to a ranking of worth that can legitimize oppression. The fact that a person's performances on various tests tends to be positively correlated, could just as well be interpreted as a result of environmental causes, Gould added.[64] These were noteworthy arguments against hereditary theories of intelligence, and a warning against the potential for political misuse inherent in Murray and Herrnstein's book, but Gould did not refute the claim that there is such a thing as a concept of general intelligence.

The Bell Curve and the public debates that followed in its wake had made research into intelligence politicized to an even greater extent. In response, the American Psychological Association (APA) saw it as their responsibility to inform the public of the existing findings on human intelligence, and separate them from their supposed political implications. They appointed a Task Force consisting of researchers with widely different views on the concept of intelligence.[65] Disagreements aside, they all recognized that the main conception of intelligence differences was encapsulated in measurements applied to aspects of the mind, the so-called *psychometric approach*, even though mental ability tests by no means is all human brains are capable of. Though there was no agreed upon definition of intelligence, and a host of human abilities other than those tested by intelligence tests was

[63] Gould 1996:52
[64] Gould 1996:372
[65] The findings were published in Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77-101.

brought up as equally important, (like creativity and social sensitivity) some important, measurable, problem solving ability was universally recognized, that was consistently referred to as ‘intelligence’.[66] The general consensus that still exists on this topic today is also in large part due to the work of the psychologist John Carroll. He collected most of the existing well-known data on human intelligence differences in his book *Human Cognitive Abilities: A Survey of Factor Analytic Studies* (1993). A wide range of tests aimed at measuring intelligence were analysed, and the conclusion was that there indeed appears to be some kind of general intelligence that accounts for about half of the individual differences among the scores for a group of people- in other words, a significant factor.[67] This means that when standard intelligence tests are correctly applied and interpreted, they more or less measure the same thing. What they measure is both stable over time in individuals, and have strong predictive power. Contrary to popular public opinion, most psychologists believe that IQ tests reliably measure something important in humans[68]. The question then is not whether these tests are useful or measure something meaningful, but rather whether what they measure is in fact the essentials of ‘intelligence’. Some experts will say that they do, while others argue that they only succeed in measuring certain aspects of intelligence.[69]My take on this is that they measure something approximately like homophenomenal intelligence, the capacity for complex problem solving based on conceptual understanding, but that there indeed are other kinds of intelligence that both humans and others demonstrate.

Even though there is wide agreement about the existence of some kind of a ‘general’ intelligence trait, (in fact it is the opposite of general in our context) there is a common worry today of a value hierarchy emerging when this most characteristic of human abilities is isolated, and it is influencing how we apply the concept of intelligence to humans. It is reasonable to think that this in turn is influencing the use of the concept in general: describing humans as low-intelligence is dehumanizing them, and describing other entities as intelligent amounts to humanizing them. As we shall see in the following chapters, there are ways to avoid this.

---

[66] Deary 2001:118
[67] Carroll, John (1993) *Human Cognitive Abilities: A Survey of Factor Analytic Studies* Cambridge: Cambridge University Press
[68] Gottfredson, L. S. (1997). “Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography.” In *Intelligence, 24*(1), 13–23.
[69] Legg, S. & Hutter, M. (2007) «Universal Intelligence: A definition of Machine Intelligence» In *Minds & Machines* 2007 Vol 17: 391-444.

Conclusion

The main thesis of homophenomenal intelligence is that the language faculty is responsible for qualitative differences in intelligence between human minds and the minds of other animals. Our use of language is how we conduct conscious propositional thinking. The phenomenal aspect, the mental interpretation of language, is of utmost importance, for without mental images there is no understanding. One important consequence of conceptual understanding is that it makes an entity not only morally valuable, but also morally responsible. This means that creating AI comparable to humans involves mental content and knowledge. So far, our limited understanding of consciousness makes this aspect of the human mind the most difficult to emulate artificially, and there are even reasonable arguments to be made that it is impossible. Even though artificial systems rapidly exceed our human capabilities in any particular given task, there seem to be a meta-level of problem solving, where problems are subjectively recognized, that is still exclusive to humans, and likely to remain so for a long time.

There is a serious problem with referring to this as general intelligence. It is not a difficult task to demonstrate that language is an incredible tool for developing intelligence, but this does not exclude other ways of being intelligent unless you define intelligence as understanding through language. In this sense, the homophenomenal concept of intelligence more or less means human-like, and as we have seen, this is not unproblematic even when we humans try to apply it to ourselves. In fact, we have a concept that does not apply to non-human entities lest we be forced to consider them in some sense human, and that we hesitate to apply to humans out of fear of de-humanizing them. Homophenomenal intelligence might be the concept of intelligence that is most aligned with intuitive notions of intelligence, but the limitations of its constructive applicability has to be recognized. To do so, we have to dispel the notion that all forms of intelligence can be measured along a single axis.

Heterophenomenal intelligence

(D3) *Heterophenomenal intelligence is (capacity for) conscious complex problem solving.*

Homophenomenal intelligence is *one* singular way of being intelligent and experiencing it. It specifically describes the typical human approach to solving problems. However, there are many other ways of consciously tackling complex problems that do not utilize conceptual thinking and understanding. Together, all of these, some of which humans are themselves capable of, constitute the concept of *heterophenomenal intelligence.* In this chapter we will explore, not simply other ways of being in the world, but other *intelligent* ways of being in the world. I will argue that there are ways of knowing that does not involve human-like language faculties, that still draws upon consciousness when faced with novelty and complexity. I will contest the claim that conceptual understanding and true learning is unique to humans, but also make the case that the similarities should not make us blind to differences. When observing and evaluating intelligence in other species, the most important insight is that the different adaptations and approaches of phenomenal problem solving are not directly comparable.

We are faced with new epistemological challenges, for not only is our ability to understand other phenomenal intelligences limited by our own horizon of experience, we also lack parameters to evaluate them as intelligent in the first place. Unlike homophenomenal intelligence, with which we are all intimately familiar, the concept of heterophenomenal intelligence includes countless ways of experiencing and interacting with the world that are utterly alien to us, from birds to insects, and possibly even machines.

General heterophenomenal intelligence

Heterophenomenal intelligence differs from homophenomenal intelligence in that it does not involve language, and it differs from system intelligence in that it involves consciousness. As shown in the previous chapter, there is a strong case to be made for the necessity of language for having knowledge as it has been defined by western philosophy. So without this knowledge, where does that leave other kinds of intelligence?

Even without the capacity for language, there are infinitely many complex problems that can be consciously solved. Aristotle distinguished between *epistêmê* (usually translated as scientific knowledge) and *technê (*skill, craftsmanship or know-how).[70]The distinction between theoretical and practical knowledge is intuitive to language-capable humans. Some things we learn through language, by conceptual reflection, by means of reading or through conversation, and some things we learn by *doing.* In the chapter on homophenomenal intelligence, we saw that the traditional view of theoretical, conceptual knowledge is that it is necessary to believe that P to know that P. I will argue that it is not always necessary to speak of beliefs in the context of practical knowledge, or knowing-how, and neither is it necessary to speak of there being some ultimate truth to the matter. Sometimes consistent successful performance of an act is all the proof we can hope for that some entity knows how to do it.[71] When we know how to do something, we do not simply know the right kind of facts: knowledge-that and knowledge how is independent. This view, called *anti-intellectualism*, is not uncontroversial. According to sympathizers like Gilbert Ryle, knowing how to F is a distinctive kind of non-propositional mental state, while critics argue that knowing-how still somehow is definable in terms of propositional knowledge.[72] However, the idea that there at least is a considerable degree of independence aligns with the majority opinion in academic philosophy.[73]This knowledge-how, I propose, is the kind of knowledge that is involved in heterophenomenal intelligence. It seems to me that the concept of knowing-that is quite clearly motivated by a desire to account for the additional capacities the human language faculty provides. Other animals accomplish complex tasks, intentionally, predictably and through cognitive efforts, but without them having a developed language, we seldom demand that whatever mental representation they have of the problem at hand must be described as some kind of proposition. We also know that there are problem solving processes in humans that do not rely on language, such as feelings and reflexes, as we should expect, since our language capacities is an add-on, not a total nervous system replacement. When we do add language to the equation, we may rightly speak of understanding and knowing-that, but even without it, we are not justified in dismissing actions as determined and performed without knowledge.

When we talk about intelligence, we are interested in the capacity for solving novel

---

[70] Fantl, Jeremy, (2017) "Knowledge How", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/knowledge-how/>. 20.03.2019

[71] Harré, R. (2010). Gilbert Ryle: The concept of mind. In *Central Works of Philosophy Volume 4: The Twentieth Century: Moore to Popper* (pp. 214-238). Acumen Publishing Limited.

[72] Stanley, Jason (2011). "Knowing (How)". In *Noûs, 45*(2), 207-238.

[73] Fantl, Jeremy, (2017) "Knowledge How", In *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/knowledge-how/ 20.03.2019

and complex problems, not just repetitions of previously successful formulas. We do not normally consider performing a task that has been performed perfectly well on numerous occasions as requiring much in the way of intelligence. We might even refer to it as instinctual. That is not all there is to knowing-how, though. Successfully solving a complex practical problem is usually the end result of a long process of trial and error (even when this is being performed in a conceptual mental simulation of the task), inching ever closer toward the desired outcome. When an archer finally hits the bullseye after countless, progressively better attempts, we describe this, to some degree, as an act displaying knowing-how. Non-conceptual intelligence is a capacity that incrementally increases with repetitions, not a static determined ability, and is thus able to overcome novel obstacles. This can be seen demonstrated in primates, birds and even invertebrates. The ability to consciously learn and accumulate knowledge in a non-conceptual manner is central to heterophenomenal intelligence. We will return later to the troublesome dichotomy of understanding and instincts, but let us first examine the common denominator in phenomenal intelligence: consciousness. We have so far been describing how phenomenal intelligence without language can work, but to see how it differs from non-phenomenal intelligence, we have to understand how consciousness contributes to problem solving. Consciousness experience is in many ways *whole*, indivisible. It gathers countless blind processes working on different problem solving tasks, and provides focused, informed solutions to the most pressing matters. Here, the key concept is *attention.* Most neuroscientists closely link consciousness with attention, and some are arguing that attending to an object is the same as becoming conscious of it.[74]

The function of attention is to avoid being encumbered with informational overload. It is a mechanism that selects information that is currently of relevance, leaving the rest to sort of run in the background, at reduced bandwidth.[75]Similarly, AI researchers often attempt to replicate the selectivity *functions* of consciousness, disregarding questions of the role of mental content, beliefs or understanding that play a central part in the homophenomenal concept of intelligence. The consciousness they seek to replicate goes by the name *of access consciousness*, as opposed to phenomenal consciousness. Phenomenal consciousness is simply a name for experience, while access consciousness is direct control of reasoning, reporting and action.[76] Whether one can exist without the other is up for debate, but when we

---

[74] Koch, Christof, Tsuchiya, N. (2012) "Attention and consciousness: related yet different." In *Trends in Cognitive Sciences*, Volume 16, Issue 2, 2012: 103-105
[75] Naotsugu, Tsuchiya, Koch, C. (2008) "Attention and consciousness." In *Scholarpedia,* 3(5):4173.
[76] Block, Ned (1996) "How can we find the neural correlate of consciousness?" In *Trends in Neurosciences*, Nov, 1996, Vol.19(11), p.456(4)

are examining what separates heterophenomenal intelligence from structural intelligence, they both are of interest to us. In matters of ethics, whether a being can experience something, like pain or grief, is important to some degree to most philosophers. For many, the fact that dogs do not have a conceptual understanding of "pain" is inconsequential compared to the fact that they can experience it consciously. Phenomenal consciousness is often considered sufficient for moral status, which means that an entity having the property of heterophenomenal intelligence has moral status, but not necessarily an entity that 'merely' has the property of system intelligence.[77]However, the functions of consciousness is of greater importance to us in this context, because we want to describe what heterophenomenal intelligence *is*, not just state ethical implications.

There are serious arguments against the importance of consciousness in cognition that must be considered. The psychologist George A. Miller made an important early contribution in stating that it is the *result* of thinking, not the *process* of thinking that appears in consciousness.[78]Other studies has since shown that analysis and selection of information for entry to consciousness is unconscious, as is control of attention, the processes that create and retrieve memories, the formulating of ideas into a suitable form of speech, organizing responses to stimulus, determining priorities, among others.[79]Much evidence points to consciousness only being involved in presenting solutions, the problem solving process appears to be unconscious. You may experience a sensation of control when consciously striking a nail with a hammer, but studies have shown that consciousness does not affect either the speed or accuracy of the action.[80]

Still, consciousness seems like such a complex phenomenon that it is unlikely to have evolved without having important biological functions. Our consciousness appears to directly influence our behaviour. Among the evidence is the fact that we have developed a second visual system to provide suitable information for conscious perception, and the process of REM atonia: during our dreams, messages to major muscles are blocked, a procedure that could only have evolved because our experiences cause actions.[81]One biological function of consciousness that has been proposed, is that consciousness is a part of a *flexible response mechanism* (FRM) that have evolved to generate responses to novel situations, where it serves

[77] Perhaps most famously put forth by Peter Singer in *Animal Liberation: A new ethics for our treatment of animals* 1975
[78] Miller, George (1962). *Psychology: The Science of Mental Life*. New York, NY: Harper and Row
[79] Velmans, Max (1991). "Is human information processing conscious?" In Behav. *Brain Sci.* 14, 651–726
[80] Hansen, Ronald, Skavenski, A. (1985) «Accuracy of spatial localizations near the time of saccadic eye movements." In *Vision Res.* 1985; 25(8):1077-82.
[81] Earl, Brian (2014). "The biological function of consciousness." In *Frontiers in psychology, 5*, 697.

as the input data of this mechanism. Its purpose is to generate nonautomatic responses.[82]In this line of thinking, we see echoed the thoughts of language as freeing the human mind from determinism, and the importance of agency over pure stimulus-response causality in intelligent processes. However, unlike arguments like the linguistic idea of language as creativity, the notion of consciousness as part of a response mechanism posits the phenomenal as a fundamental part of the intelligent problem solving of all species on earth. It is time to broaden our scope of inquiry.

Close to Human

One line of reasoning that seeks to dissolve the clear boundaries between human minds, that are intelligent, and animal minds, that are not, is contending the differences. Kristin Andrews lists three main types of arguments for assuming similarity between humans and other species: the *argument from analogy*, the *inference to the best explanation argument*, and the *argument from evolutionary parsimony*.[83] The argument from analogy basically states that all other beings I know are thinking beings have property *x*, for instance a brain, individuals of a certain kind of species also has the property *x*, therefore individuals of this species probably are thinking beings. This line of argument is in my view not very convincing when applied to physical anatomy, as there are many differences in brain structure to account for even when accepting fundamental similarities, but as we shall see later, the argument gains some strength when we consider behavioural properties. The inference to the best explanation argument *does* focus on behaviour, stating that if members of a species engages in behaviour *x*, and the best explanation for behaviour *x* is that they have a (intelligent) mind, then most likely the members of this species have such minds. Those opposed to this conclusion will object that there might be vastly different processes that give rise to superficially similar behaviour. We will later get to the concept of 'instincts'. The argument from evolutionary parsimony suggests that we should, to an extent, disregard the thesis of different behavioural causes because we share a common ancestor with these animals, and we should apply Occam's razor explaining the emergence of that behaviour.

One apparent fundamental similarity is consciousness. Studying their behaviour makes us think that other animals are conscious, but verifying this assumption differs from how we

---

[82] Earl 2014

[83] Andrews, Kristin, (2016) "Animal Cognition", In *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/cognition-animal/ 20.03.2019

investigate human consciousness. In humans, we combine empirical studies of what goes on in the anatomical brain with *report* from the test subject of their conscious states. For instance, one line of human studies involve the correlation of accurate report with the presence of thalamocortical signalling. The *thalamic dynamic core theory of conscious experience* proposes that conscious awareness arises from synchronized activity in dendrites of neurons in dorsal thalamic nuclei, based on, among other things, the role of the thalamus in vegetative state brain injuries and general anaesthetics.[84]Other similar approaches includes studying the neural correlates with EEG signals and similarly widespread cortical activity. In general, they all lean on explicit verbal, or linguistic, report as a guarantee of the presence of consciousness correlated to the observed brain states. Animals, of course, are unable to produce such linguistic reports. We should therefore recognize that the mechanisms responsible for consciousness might be different from the mechanisms enabling its report. In that case, non-verbal responses such as pressing buttons and levers might be equally accurate indicators of consciousness correlating to observed brain anatomy.[85]

What does the presence of consciousness tell us about an entity's intelligence? We have already accounted for the importance of language, but yet humans and non-human animals share a number of biological, morphological, and relational properties. For instance, most accept that humans and animals can both be attributed some psychological properties such as the ability to *fear* or *desire*. Some psychological states, such as beliefs, as we have seen, are thought to be mainly human. Some properties are viewed as 'higher' and others as 'lower', and higher ones are without exemption the ones we ourselves possess, candidates for justifying human uniqueness, even if no satisfactory account of what 'higher' and 'lower' mean so far has been offered.[86] Is the difference that these mental states in humans can be described in terms of conscious insight, whereas animals have their cognitive abilities due to genetically determined connections between situation and response? Let us examine the relationship between instincts and learning.

A distinct qualitative difference in intelligence between human and non-humans seems the most difficult to defend when considering the genealogy of ourselves, early hominids and common ancestors of apes, but comparisons between humans and existing species of apes do

---

[84] Ward, Lawrence (2011) "The thalamic dynamic core theory of conscious experience" In *Consciousness and Cognition* Vol. 20, Issue 2, 2011: 464-486

[85] See Edelman, David, Seth, A. (2009) "Animal consciousness: a synthetic approach" In *Trends in Neurosciences*, Vol. 32, Issue 9, 2009: 476-484

[86] Andrews, K. & Huss, B. (2014) "Anthropomorphism, anthropectomy, and the null hypothesis" In Biol Philos (2014) 29: 711.

not radically differ. We should expect to find the most similarities in cognitive processes and functions in our closest biological kin, other primates, and indeed, their behaviour often closely resemble that of humans. The anthropologist William McGrew writes:

> Tickle a chimpanzee, and she laughs; startle a chimpanzee, and he grimaces; threaten a chimpanzee, and she lashes out; groom a chimpanzee, and he sprawls relaxed. All of these signals of feelings are recognized readily by the average person. More dramatically, when we see an orphaned ape with her dead mother, her demeanor or ‘body language’ is one that, if seen in a human child, would be interpreted as grief (McGrew 2004: 8–9)

Opinions differ on whether this as sign of a “theory of mind” ability (the ability to attribute mental states to oneself and others, and to understand that others have other beliefs, desires etc.) or just sophisticated social behaviour.[87] However, it does seem difficult to argue that chimpanzees manage multiple complex and changing social relationships without any capacity for structured propositional thought.[88] A behaviourist account of primate intelligence would certainly make a less clear distinction between humans and chimpanzees than any theory based upon the necessity of a mental language of thought for beliefs and understanding. We come across immense difficulties when we try to shift our focus from behaviour to mental states when looking at other species than our own. The reason, as stated by professor of biology Daniel Povinelli, is that, given that inferences about hidden mental states are made on the basis of bodily and environmental cues, it follows that, for any mentalistic hypothesis, there is a complementary “behaviour-reading” hypothesis that can explain the data equally well. Because human mentalizing abilities are built on top of, and integrated into, ape-level abilities to read and represent behaviour, we cannot know which of these mechanisms are being used to solve a particular task.[89] . It does not matter how impressive the feat. A chimpanzee named Washoe learned a gestural vocabulary of about 130 words, and was able to form creative word combinations. Washoe combined “Water-bird” to describe a swan on a lake, and “listen-drink” for Alka-Seltzer. Even in cases like this, and in spite of the intuitions the researchers conducting the experiments, conclusions of the scientific community have usually been very humble indeed: Unless alternative explanations of an ape’s combination of signs are eliminated, there is no reason to regard an ape’s multisign

[87] Premack, David, Woodruff, G. (1978). "Does the chimpanzee have a theory of mind?" In *Behavioral and Brain Sciences, Special Issue: Cognition and Consciousness in Nonhuman Species.* 1 (4): 515- 526.
[88] See Carruthers, P. (2002). “The cognitive functions of language.” In *Behavioral and Brain Sciences, 25*(6), 657-74; discussion 674-725.
[89] Barrett, Louise (2018) Picturing Primates and Looking at Monkeys: Why 21st Century Primatology Needs Wittgenstein in *Philosophical Investigations* Volume 41, Issue 2 2018:161-187

utterance as a sentence.[90] This elimination, as Povinelli points out, cannot be carried out.

Other animals that are harder to compare with humans also exhibit behaviour which would require intelligence if humans were to perform them. Seemingly intelligent behaviour in these animals are often characterized as instinctual, to emphasize that their problem solving abilities are hard-wired, rather than due to any real kind of awareness of the situation. However, the ability of other animals to learn has often been shown to be underestimated. It appears that, at the very least, the 'instincts' are quite adaptable.

Eberhart Curio researched bird mobbing calls in the 1970's, and concluded that the young birds' recognition of enemies was learned through the alarm calls of older birds, not through preborn instincts. He placed birds in opposing cages with an opaque turnstile separating the view between them. On one side, older birds was shown an owl, and on the other, young birds was shown a wide range of objects, from nectar feeding birds to bottles of laundry detergent. The alarm calls of the older birds sounded in response to the owl. Subsequently, the young birds developed fear for harmless birds and household items. The reaction to the older birds' calls were of course instinctive, but the experiment showed these instincts to be tools in a learning process in a similar way that humans also deploy natural instincts for learning.[91]The clear distinction between intelligent adaptability, and instinctual behaviour, does not hold up well when closely examined. This means that the term 'instinct' is problematic, even in species not often considered intelligent, like insects. The problem solving capabilities of bees trying to obtain food from alfalfa flowers has been studied. The alfalfa flowers have anthers that spring forward, delivering a startling blow to the bees. This will trigger avoidance behaviour in the bees until the alfalfa flowers are the only food available. The bees will then approach flowers whose anthers have already sprung, or eat from the sides of the flower, avoiding triggering the anthers. This behaviour is only seen in bees that have already been hit by an anther, in other words: in bees that have *learned*.[92]

In the case of birds, the complexity of their behaviour is so obvious that insistence on describing it as instinctual, rather than a demonstration of intelligent learning ability, seems out of place. Bernd Heinrich conducted in 1990's research on the problem solving abilities of ravens. Specimen birds raised in captivity were, one at a time, introduced to a piece of meat

[90] S.Terrace, L.A.Petitto ,R.J.Sanders T.G.Bever (1979) Can an Ape Create a Sentence? In *Science* Vol 206, Number 4421, 23 November 1979:891-902

[91]Original research in Curio, Eberhard & Ernst, Ulrich & Vieth, Willy. (1978). "The Adaptive Significance of Avian Mobbing." In *Zeitschrift Tierpsychologie*. 48. 184 - 202. Described in DeFusco, Christopher R. (2005) *Animal Consciousness. Grounds for the Ethical Treatment of Animals* Xlibris

[92] Described by Randolf Menzel in "Learning, memory and Cognition in Honey Bees" in *Neurobiology of Comparative Cognition* (1990) and in DeFusco 2005

hanging from a string on a horizontal tree branch. After a few days of trial and error, including attempting to catch the meat mid-flight and reaching from the branch, four out of five had become successful, using two different methods. Two of them figured out that it was possible to draw the string step by step up by holding the loops with their feet, while the other two pulled the string along the length of the branch until the string was pulled to the level of the branch. Heinrich concluded that innate programming could hardly account for the results.[93] The ravens had understood the problem, imagined and implemented a solution.

If we are to describe the learning process of animals this way, it seems hard to avoid the conclusion that it would have to involve an understanding on some level of core logical concepts we usually ascribe to humans exclusively. Research on sea lions has found that they are capable of using logic to solve novel problems, suggesting that the cognitive processes which enable the emergent association of perceptually different stimuli are fundamental, and that linguistic ability is not necessary to support this capability. [94]The role of language in intelligence and learning is far from obvious. There have also been attempts at downplaying the exclusiveness of human language. The theory of *continuity* claims that human language must have evolved over time, not just appearing fully formed in a single mutation.[95] Because of this, we could find intermediate levels of this ability, at least in our closest genetic kin. It is an argument that has serious weaknesses. Steven Pinker likened human language to an elephant's trunk: we should not bother testing other animals to see if they are able to use their noses in the same way as elephants.[96] A lack of intermediates when it comes to extreme features like trunks and language is not uncommon in nature.

There appear, however, to be a wide range of intermediates between human-level intelligent learners and strictly instinct driven species. Another argument against the view that any creature with a nervous system is a pre-programmed instinctual machine rather than a genuine intelligent learner, is the known, actual physical properties of biological brains. It is demonstrably sensitive and adaptable to external conditions. Creatures that are brought up in an enriched environment have larger brains than do their conspecifics that are raised in an impoverished environment. A rat placed in isolation in a laboratory cage will have a smaller

[93] Heinrich, Bernd (1995) "An Experimental Investigation of Insight in Common Ravens (Corvus corax)" in *The Auk*, Vol. 112, No. 4 (Oct., 1995):994-1003
[94] Lindemann-Biolsi, Kristy and Colleen Reichmuth (2013) Cross-modal Transitivity in a California sea lion in *Anim Cogn* 2014 Jul;17(4):879-90
[95] Chapman, Siobhan, Routledge, C. (ed.) (2009) *Key Ideas in Linguistics and the Philosophy of Language* Edinburgh University Press
[96] Pinker, Steven (1994) *The Language Instinct: The New Science of Language and Mind.* Harmondsworth: Penguin

brain than rats placed in environments with toys and other rats.[97]If we assume that at least some functions of this larger brain are dependent on previous cognitive activity on the rat's part, in interactions with the world, then in what sense is the rat's intelligence any more predestined than our own? On the whole, the human brain, the part of the body that determines our intelligence, is physically very similar to the brains of at least mammals, fundamentally. There is compelling evidence to support neurophysiologic similarity between humans and other animals in the effect of drugs designed to alter brain chemistry. Humans and animals respond similarly to Valium and Prozac, used to treat anxiety and depression. In dogs with compulsive behaviour, such as constantly licking their paws or chasing their tales, Prozac is very effective.[98] A more surprising finding is that octopi respond in much the same manner as human to methylendioxymethamphetamine (MDMA), with enhanced acute prosocial behaviour.[99]

The collective weight of evidence that non-human animals have the neuroanatomical, neurochemical and neurophysiological substrates of conscious states, and the capacity to exhibit intentional behaviours, led to the Cambridge Declaration of Consciousness in 2012: Non-human animals, including all mammals and birds, possess the neurological substrates that generate consciousness.[100] If conscious mental states are required for some particular kind of intelligence, other animals seem to fulfil the criteria as well- but are they intelligent to the extent that they are similar to us? I have in this chapter presented evidence supporting the claim that other animals also possess this homophenomenal intelligence to a degree, but the main lesson when observing earth's other species is that there are countless ways to experience and approach almost any given problem, yet only a select few seem to qualify as intelligent. This is where the need for the concept of heterophenomenal intelligence presents itself.

---

[97] Bennett, Edward L., et al. "Rat Brain: Effects of Environmental Enrichment on Wet and Dry Weights." *Science*, vol. 163, no. 3869, 1969: 825–826.

[98] DeFusco 2005

[99] Edsinger, Eric, Dölen, G. (2018) "A Conserved Role for Serotonergic Neurotransmission in Mediating Social Behavior in Octopus" In *Current Biology* 2018 Oct 8;28(19):3136-3142

[100] The Cambridge Declaration on Consciousness was written by Philip Low and edited by Jaak Panksepp, Diana Reiss, David Edelman, Bruno Van Swinderen, Philip Low and Christof Koch. The Declaration was publicly proclaimed in Cambridge, UK, on July 7, 2012, at the Francis Crick Memorial Conference on Consciousness in Human and non-Human Animals, at Churchill College, University of Cambridge, by Low, Edelman and Koch. The Declaration was signed by the conference participants that very evening, in the presence of Stephen Hawking, in the Balfour Room at the Hotel du Vin in Cambridge, UK

Anthropomorphism

In the first decade of the 20th century, worldwide attention turned to Berlin, where a horse named Hans was touted as the world's only 'speaking' and thinking animal. Hans could solve calculations by tapping the correct answer with his hooves, in response to questions. It turned out to be a 'scam'. Hans was able to answer the questions correctly by reading microscopic signals in the face of the questioning person. Ever since this discovery, experimenters on the cognitive abilities of animals have tried to avoid face-to-face contact with the animals.[101] While scientists were quick to withdraw claims of Hans' mathematical understanding, little was said of his quite impressive face-reading capabilities. He had recognized a different problem and situation, and solved it in a manner humans are not able to. What use is a general concept of intelligence, if what is considered intelligent invariably is what a human is capable of?

The difficulties in human intelligence testing, such as bias due to language differences or physical handicap, become even more difficult if we try to compare animals with different perceptual and cognitive capacities. Not only is it not always obvious how to conduct the tests, with animals we are unsure what should be tested for. As humans devise the tests, there is a danger that the tests may be biased in terms of our sensory, motor, and motivational systems.[102]Moreover, how do we go about validating an intelligence test for animals? An IQ test for humans can be validated if it is able to predict future academic or other similar successes, but the parameters are necessarily limited by a very human understanding of success. We can attempt to make very general criteria, like future reproduction rates, but if survival or the total number of offspring were our measures of success, then bacteria would have to be considered the most intelligent life on earth. That is probably not what we mean to say, but it does not undermine this point: If there are different ways of being intelligent, there are different ways of being successful. Evaluating heterophenomenal intelligence with linear parameters like in an IQ-test is not possible. Both the complexity of the presented problem and the complexity of its solution is to an extent hidden from us.

Daniel Povinelli has argued that contemporary animal cognition researchers are too eager to undermine claims of human uniqueness. He worries that the science of animal

---

[101] Samhita, L., & Gross, H. J. (2013). The "Clever Hans Phenomenon" revisited. In *Communicative & Integrative Biology*, *6*(6), e27122. http://doi.org/10.4161/cib.27122
[102] Macphail, E. M. (1985). Vertebrate intelligence: The null hypothesis. In L. Weiskrantz (Ed.), *Animal intelligence* (pp. 37–50). Oxford: Clarendon.

cognition is harmed when the researchers assume similarity, because very real differences will not be discovered.[103] Anthropocentrism, the placing of humans as centrepiece of creation, did not simply vanish from the scientific community with Darwin. The father of evolution had stated that the difference between the mind of man and the mind of an animal, though immense, certainly was of degree, not of kind.[104] This perceived continuity in intelligence makes us blind to the wonder of biological diversity. When we ask ourselves how intelligent chimpanzees are, we must be careful that we are not simply asking how human they are. If we choose to define intelligence as the mental capabilities of human beings, why is the question even interesting?

The charge of *anthropomorphism*, the attribution of human characteristics or behaviour to objects or animals, is not just something proponents of a homophenomenal concept of intelligence might level against those who defend a diverse understanding of phenomenal intelligence. Our inescapable anthropomorphism is also an epistemological blind spot, whenever we pass judgement on what range of possible subjective experiences and intentions that might accompany the behaviour we observe in other animals. Scientists have tried to combat this. Donald Griffin, an American zoology professor, is credited with establishing the field of *cognitive ethology,* which combines cognitive science with the study of animal behaviour and stresses observing animals under natural conditions. Its main object of study is the mental experiences of animals, and encompasses both the 'hard problem of consciousness', phenomenological experience and the easy problems, phenomena that might be explained by computational or neural mechanisms.[105] The idea is that by starting with real-world observations, one is more likely to generate subsequent research questions for investigation that may lead to general principles of cognition that have relevance to naturally occurring phenomena, than were we to test animals in laboratories. Important aspects of cognition will only emerge when the individual is considered as a part of their natural environment, which includes other individuals.[106]However, what constitutes the environment may appear quite different to the human observer and the animal subject.

The experiences of other animals are different from ours, not necessarily because they lack concepts, but because their range of perceptional experience differs. Different ranges of

---

[103] Povinelli, D. J. and J. M. Bering (2002). "The Mentality of Apes Revisited." In Current Directions in Psychological Science 11(4): 115-119.
[104] Darwin 1871/1982:445
[105] Ristau, Carolyn (2013) "Cognitive ethology" In *WIREs Cogn Sci,* 4: 493-509.
[106] Kingstone, Alan, Smilek, D. and Eastwood, J. D. (2008) "Cognitive Ethology: A new approach for studying human cognition" In *British Journal of Psychology*, 99: 317-340

hearing, smell, and vision in other species certainly must provide for qualitatively different phenomenal lives, not to mention sensory input like echolocation or electromagnetic senses, of which we humans have no direct experience with. Anthropomorphizing is perhaps unavoidable when speculating in the phenomenal minds of other species, much like Kant's conception of the Categories. Thomas Nagel's famous article *What is it like to be a bat?* (1974) was primarily an argument against reductionist theories of the mind, showing that they could not account for qualia, what it feels like to have an experience, but Nagel also showed just how profound difficulties we are faced with when we attempt to picture or describe *other* minds.[107] What can we know about forms of intelligence that are nothing like humans? What can coherently be said about a genuinely heterophenomenal concept of intelligence?

Different kinds of intelligence

In his book *Other Minds. The Octopus and the Evolution of Intelligent Life* (2016), Peter Godfrey-Smith tracks the evolution of intelligence, with a particular focus on species that display intelligent behaviour, and that separated early on from our own genetic lineage. If we want to understand other minds, he writes, the minds of cephalopods (octopi, squids etc.) are the most other of all.[108] In cephalopods, the neurons are largely decentralized, distributed across the arms as well as the central brain. The arms enjoy considerable independence, with recurrent connections giving them a form of short-term memory. Thus, the octopus' body is in a sense not separate from the brain, it is all nervousness. It is perhaps the best example of biological *intelligent embodiment*[109]. The term comes from robotics, describing autonomous robots in which the behaviour emerges from the dynamic physical and sensory interactions of the system's materials, morphology and environment.[110] The arms are partly self, partly non-self- agents on their own. This is at first hard to picture from a human viewpoint, where the highly centralized nervous system gives rise to an experience of a singular self that encompasses the entire body.

What is it like then, to be an octopus? There is perhaps a lot more unconscious movement, from the central brains point of view. To some extent, you guide your arms, to some extent, you just watch them go, Godfrey-Smith speculates. This, of course, is often true

[107] Nagel, T (1974) What *is it like to be a bat?* Philosophical Review, 83: 435–456.
[108] Godfrey-Smith 2016:10
[109] A closely related term is *embodied cognition*, a theory that cognition is shaped by an organism's entire body, not just the brain. See Rosch/Thompson/Varela 1991
[110] Hochner, Binyamin (2012) «An Embodied View of Octopus Neurobiology" in *Current Biology* Vol 22, Issue 30, October 2012: 887-892

of human motions as well, but could be a much more distinct and present feature of the octopus experience.[111] Being an octopus is nothing like being a human, and yet octopi and their relatives regularly display surprising feats of intelligence. Why are they so different compared to other invertebrates? Biologists researching cephalopods have emphasized their jack-of-all-trades lifestyle, in varied environments, as an important explanation for their impressive cognitive abilities. Novelty is a key word in the evolution of biological instincts. Instinctual behaviour is the result of genetically pre-programmed responses to the environment. Suppose that in a given environment there exists multiple strains of a primitive organism, in which certain stimulation triggers different behaviour. In some strains, the stimulation causes the organism to back off, in others it causes them to move towards the source. If this stimulation turns out to be beneficial, the former strains are at an evolutionary advantage. However, if the environment is volatile, and at some point this particular stimulation becomes lethal instead, the benefits of instinctual behaviour is gone. Should fires become commonplace in the environment of phototropic moths, they would face extinction.[112] Novelty seems to be such an important driving factor in the evolution of intelligence, that we can make loose predictions: the more novelty encountered in your environment, the more heterophenomenally intelligent you are likely to be.

Examples of conscious problem solving unimaginable to humans are countless in the animal world. What is it like to be a bat? Somehow they discriminate the echoes of their own sounds from thousands of echoes of other bats in their caves. Their powers of echolocation probably approaches ‘echoperception’, forming a model of their surrounding world functioning much the same as the visual counterpart in humans. Birds and bees use the suns horizon position as compass, a system that requires, in addition to an accurate internal clock, the ability to extrapolate from the sun’s present rate of movement to its expected position several hours later, so that they can calculate the angle. Clark’s nutcracker, a bird with a brain weighing a few grams, remembers the location of half of its 30 000 stored seeds during winter, a spatial memory far superior to humans.[113] If we take for granted that they are conscious, it seems that their experiences of the world must vary endlessly. That means that different animals have different ontologies. The nutcracker lives in a world of memories of stored food. An anteater might not notice exactly how many individual ants it sucks up with its tongue, but insectivorous birds relate to individual insects. The difference lies in what

[111] Godfrey-Smith 2016:104
[112] Dennett 1986: 53
[113] Khalfa 1994:29

potential information they can make use of, for instance the information carried in the spectrum of light.[114]Insects relate to flowers with certain kinds of colour, and live in a world where they are a distinguishable and prominent feature, while birds similarly relate to colours on the opposite side of the spectrum.

Pointing out differences is easy though, compared to generalizing the concept of heterophenomenal intelligence. What every way of solving complex problems so far mentioned has in common is that they are widely thought to be accompanied by experience. The exact requisites for this experience to occur are not at all clear. As we consider species with progressively less 'developed' nervous systems, claims of both intelligence and consciousness become proportionally controversial. Let us just as well focus on the strongest claims: brainless intelligence and consciousness.

Plants, fungi and single-celled organisms all engage in a fierce competition for resources, and have to respond and adapt to changing environments, albeit usually at a much slower pace than free-moving animals. Large amounts of research document behaviour that scientists unhesitantly characterize as intelligent. One of the most interesting cases is the slime mould *Physarum polycephalum*. It survives by locating nutrients in its environment, and extending itself towards it. When presented with mazes in experiments, it is able to choose the shortest route to its food, and if it detects detrimental factors like light, it takes the alternative shortest path. It makes decisions. It assesses the risk-reward balance between danger and need for food, and it selects the most nutritious foods after sampling alternatives. It also learns. When given repeated shocks, it will learn to react to them by a temporary cessation of growth.[115]It seems necessary to conclude that organisms like that are able to solve problems of some complexity even without a nervous system, but is it at all plausible that they do so somehow consciously?

When a plant detects light or nutrients, it grows and extends itself towards it, and in a sense they intend to. For this to be a conscious act, we should be able to say that the plant is aware of the nutrients it is reaching for. There would have to be some kind of qualia. What could it be like to be a plant? It has been suggested that, for a plant, the environment appears as a complex spatial and temporal mosaic of resources, light and shadow, not a conglomeration of discrete objects.[116]There are no neurons present to provide a mental

[114] Dennett 2017: 65

[115] See Adamatzky, Andy, Armstrong, R., Jones, J., Gunji, Y. (2013) "On Creativity of Slime Mould" In *International Journal of General Systems* April 2013

[116] Trewavas, Anthony (2003) "Aspects of plant intelligence." In Annals of Botany 2003; 92:1 - 20

representation of the plant itself or the external world, so for it to be true that the plant indeed is experiencing some kind of qualia, it has been proposed that what appears to the plant cannot be a representation of the thing, but the thing itself.[117]This, I would, argue, is misunderstanding what qualia is. Qualia are not physical, so if there indeed is such a thing as plant experience, then there is something being represented somewhere in the plant. Another possibility is that there simply is no coherent cognitive map, or global representation, but a decentralized experience. This would imply some form of simpler representation occurring at the cellular level, still with a degree of cohesion. Is the idea of a consciousness that is not 'whole' a real thing that happens to be out of our epistemological reach as humans, or does it contradict what we mean by consciousness? As seen earlier, consciousness is already hard to pin down in humans, and expanding the scope to all conscious entities may seem optimistic. However, that is precisely what artificial intelligence researchers concerned with the functions of consciousness must do.

The Global Workspace Theory of consciousness (GW) has become very influential in recent years. It posits that consciousness is a fleeting memory capacity that enables access between brain functions that are otherwise separate.[118]If we view the brain as a large parallel set of specialized processors, coordination and control can take place by way of a central information exchange, distributing information to the entire system. The functional role of consciousness, GW claims, includes integrating, providing access and coordinating the functioning of specialized networks that otherwise operate autonomously. The global workspace theory thus posits a centralized process as driving force for consciousness, but for most biological species, if they have subjective experience, it may not be a coherent and 'whole' self-model. We have mentioned plants and the idea of representations that do not form a singular global image. Even from a human perspective, the unity of experience can be called into question. What distinguishes animals with brains may not be that they are conscious at all, but that the experiences are integrated into a coordinating effort, primarily propelling intelligence, but also forming radically more complex representations. We often tend to picture 'simpler' forms of consciousness as more vague or faint, but complex consciousness might distinguish itself by constituting a less fractured whole. Perhaps we should hesitate to use the word consciousness too broadly, and instead speak of "subjective

---

[117] See Marder, Michael (2012) "Plant intentionality and the phenomenological framework of plant intelligence." In *Plant Signaling & Behavior,* 7:11, 1365-1372

[118] Baars 2005:46

experience in a very broad sense".[119]

This notion of experience in a broad sense was invoked by Giulio Tononi, the neuroscientist behind Integrated Information Theory (IIT), when confronted with an apparent weakness of this otherwise promising theory of consciousness. IIT posits that consciousness is information that is integrated, that any system is conscious to the extent that it contains this kind of information, and that this is quantifiable and measurable by the term "phi".[120] Scott Aaronson pointed out that it is possible to create relatively simple artificial systems with arbitrarily high values of phi that seem extremely counterintuitive to label as conscious.[121] Tonoi's response to the critique was that contrary to our intuitions, systems like the ones Aaronson described would be conscious, but they would specify nothing besides what it is like to be such a system in the space of its possible states.[122]Even with such a willingness to accept different kinds of consciousness, IIT presupposes that experience is whole. This appears to be challenged by several branches of philosophy. Recently, the field has seen renewed interest in panpsychism. Pansychism suggests that consciousness is a fundamental property of nature, widening the spectre of possible ways of being conscious even further[123]. However, it seems to be beyond questioning that entities with centralized nervous systems are a lot more adept at solving problems of a novel and complex nature then plants and mould, their impressive capacities notwithstanding. However varied consciousness may be, integration and coordination is crucial to heterophenomenal intelligence.

It is hard to picture an experience as alien as the "arbitrarily high phi" system, but when artificial intelligence is concerned, that is perhaps just what we might have to attempt. If heterophenomenal artificial intelligence should become a reality, differences in experience among biological species might appear relatively insignificant. Nick Bostrom has been a pioneer in emphasizing the *otherness* of (eventual) artificial intelligence, and in particular what goals they might have. He has been arguing two related theses: The *orthogonality thesis* holds that intelligence and final goals are orthogonal axes along which intelligent agents can freely vary. There does not need to be any correlation between how intelligent a being is, and what sort of goal it strives towards. The *instrumental convergence thesis* holds that an agent with sufficient level of intelligence will pursue similar intermediate goals to reach their

---

[119] Godfrey-Smith 2016:92

[120] Tononi, G (2012) *Integrated information theory of consciousness: an updated account* Archives Italiennes de Biologie, 150: 290-326,

[121] Aaronson 2014

[122] Tononi 2014

[123] See Chalmers 1996

respective final goals, because they have instrumental reasons to do so.[124]Thus, behaviour is not necessarily a good indication of what an agent is ultimately trying to accomplish. We should expect other intelligent entities to act in familiar ways, even when their long-term goals are utterly foreign or even incomprehensible to us. These familiar patterns of actions are the results of general instrumental goals of self-preservation, goal-content integrity, cognitive enhancement, technological perfection, and resource acquisition.

This becomes a concern when we ponder the prospects of a future artificial *superintelligence.* Bostrom defines this as “any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest.”[125] Programs that are extremely intelligent, but only within a narrow domain, like chess, do not qualify. There are different paths to a superintelligence, constituting subcategories. We might develop a system that can do all that a human intellect can, but faster: a Speed Superintelligence. There might also emerge a system composed of a large number of smaller intellects, such that the system’s overall performance across many very general domains vastly outstrips that of any current cognitive system: a Collective Superintelligence. We can also imagine a system that simply thinks *better* than we do. A system that is at least as fast as a human mind, and vastly qualitatively smarter. This would be, in Bostrom’s terms, a Quality Superintelligence.[126]

The problem is not that a superintelligence might turn out malevolent. ‘Good AI’ and ‘Bad AI’ is a simplification and a false dichotomy. The problem is that a superintelligence will be vastly more capable than humans of achieving its goals, and there is no way for us to ensure that these goals align with what we ultimately want. This is called the *alignment problem.* Several grotesque hypothetical examples of superintelligent machines blind to critical but hard to formulate human values have been making the rounds. A machine programmed to produce paperclips that ends up turning the entire galaxy into paperclips, or a machine told to maximize human happiness that enslaves all of humankind, endlessly injecting joy-producing chemicals into their brains, are just two of the more famous examples. The crucial point is that it is extremely difficult, if not impossible, to make sure something like this does not happen.[127]

This is the worry of the superintelligence sceptics, which does not include everyone working in the field of AI. Google’s Larry Page is what Max Tegmark describes as a “digital

[124] Bostrom, N. Minds & Machines (2012) 22: 71. https://doi.org/10.1007/s11023-012-9281-3
[125] Bostrom 2014:26
[126] Bostrom 2014:64
[127] Eliezer Yudkowskiy provides an account of the problem in “The AI Alignment Problem: Why It’s Hard, and Where to Start” 2016 (unpublished)

utopian". Page argues that digital life is the natural and desirable next step in the cosmic evolution. The worries of sceptics like Elon Musk, he labels speciesist, in that they prefer carbon-based life forms over silicon-based. Other utopians include the roboticist and writer Hans Moravec and reinforcement-learning pioneer Richard Sutton. A third position is to think that a superintelligence, more moderately termed advanced general intelligence (AGI), is really far off, and that we should prioritize debating other short-term issues. Among those who hold this view is Baidu's Andrew Ng, and former MIT professor Rodney Brooks.[128] Tegmark himself is, along with people like Stephen Hawking, in large part responsible for bringing the debate on how to ensure that AGI is beneficial out into the mainstream, even though the timeframe of this happening is hotly debated.[129] It is possible that our natural tendency to view intelligence from an anthropocentric perspective will lead us to underestimate some kinds of improvement in sub-human systems, and overvalue others. AI might at one point appear to make a sharp jump in intelligence if we think of village idiot and Einstein as the extreme ends of the intelligence scale, but AI is nothing like us, and might cross this tiny gap in a very short period.[130] The risk lies in its otherness.

The otherness of non-biological intelligence also has other, ethical implications. For example, some entities may not fear their own destruction or have need for emotions. Ray Kurzweil argues that it would be unethical to destroy a device committed to pursuit of a complex and worthy goal. We need to include as conscious entities that do not care to try to convince us of their emotions. It is hard to put ourselves in the subjective shoes of another human, harder to do so with other biological species, and this task will be harder still with intelligences extremely different from our own.[131] Though it lies outside our human abilities to picture an AI experience, we can predict some aspects that must be relevant: the speed of the computational elements, the internal communication speed, the number of computational elements, the storage capacity, the reliability, the lifespan, the sensors, the editability, the duplicability, the goal coordination, the memory sharing, and the algorithms.[132]These must all somehow factor in what it feels like to be a machine. For instance, because of the time dilation of the material world, a speed intelligence would prefer to work with digital objects.

Another question is whether consciousness needs locality, to be situated some place in

---

[128] Tegmark 2017:32
[129] See for instance Chalmers, David (2010) "The Singularity: A Philosophical Analysis." In *Journal of Consciousness Studies* 17 (9-10): 7-65.
[130] Bostrom 2014:85
[131] Kurzweil 2012:214
[132] Bostrom 2014:71

space-time. At least on a physicalist view, artificial phenomenal intelligence would be physically realized by the electronic circuits performing the processes involved in consciousness in the same manner that human consciousness is realized in the neurons of the brain. That the machine's experience is transferable from one set of circuits to another does not differ much from the way that new neurons carry on processes in the brain. It does however bring up the philosophical problem of identity persisting over time: would it be the same consciousness once transferred to other physical circuits? It is an interesting question in itself, but not one we shall pursue.

Are there any limits to how phenomenal intelligences, or experiences of being in the world, can differ? We have mentioned the demands of integration and coordination. Must there also be some kind of agreement on the fundamentals of reality? This seems necessary. Our understanding of physics is in some sense based on our biology, because it is through our biological sense apparatus we learn of the world. Even machines experiencing solely digital realms are bound by the physical constraints of the hardware facilitating their processes. Surely, the number of logically possible ways of being intelligent in the world is larger than the number of possible ways of being intelligent that the laws of physics allow. Outside of such fundamental limitations, no clearer indication of heterogeneity in intelligence exists, than the abundancy and diversity of the millions of species on Earth.

Conclusion

Heterophenomenal intelligence is endlessly varied, but even if consciousness itself is as broad a category as panpsychism suggests, there is a crucial general factor. Even though life-forms like plants are able to implement more intelligent solutions in their acquisition of resources than they are usually given credit for, the centralized focus, integration and coordination that is enabled by having a brain (or something functionally similar) is demonstrably the single most impactful function of consciousness. However, the usefulness of the concept of heterophenomenal intelligence does not lie in demonstrating that the intelligence of animals or machines is comparable to ours, but rather the opposite. Though there are good arguments to be made for the claim that the qualitative differences between humans and animals/machines are less significant than proponents of human uniqueness often suggest, the main thesis of heterophenomenal intelligence is that other ways of intelligent problem solving escapes comparison. Even if our language faculties grants us a special kind of phenomenal intelligence, with special kinds of knowledge and beliefs, we have no reason to believe that there are not other ways of being consciously intelligent that we are not equipped to imagine. This has ethical consequences for our dealings with other species. Our intuition is often that the more intelligent a species is, the more morally wrong it is interact with a member of said species in a strictly instrumental manner. Homophenomenal intelligence is a very poor parameter in this context, which means that other lines of thinking are required for sound, coherent judgments of moral status. Unless, of course, we are prepared to say that human properties are the exact properties that matter ethically.

That homophenomenal intelligence alone is a poor predictor of problem solving capabilities in other conscious beings can be seen demonstrated in countless species all over the world, and distinguishing between intelligence and instincts seems shaky at best. Treating these foreign capabilities as intelligent in their own way probably both has more explanatory power and is more informative about what to expect. This might appear to be a question of purely academic significance, but at least regarding the prospects of AI, that is far from being the case. The term artificial general intelligence (AGI) was popularized by AI researchers Shane Legg, Mark Gubrud and Ben Goertzel to specifically mean human-level artificial intelligence.[133]As we have seen, humans and machines operate with quite different parameters. Machines and people operate at the same level in some respects, and in others the machines have long since surpassed us. If our goal is to create machines that are intelligent in the sense that they can do precisely the things that humans do, then referring to this as 'general' intelligence obfuscates the level of their *actual* problem solving abilities.

---

[133] Tegmark 2017:52

System intelligence

(D4*) System intelligence is (the capacity of) complex problem solving systems.*

As shown in the previous chapter, consciousness plays an important role in focusing attention when an entity is performing an action requiring intelligence. However, the sometimes-close relationship between the two phenomena makes the concept intelligence in itself hard to isolate. It makes sense to speak of intelligence without phenomenal experience, but in order to do so we must abandon the locus of consciousness and turn our attention to the *system*. In the previous chapters I never questioned the ontology, the existence, of the objects having the property of intelligence. There were no forays into problems of identity or the illusions of the self. In this chapter, however, I cannot afford myself such luxury.

I will begin by making the case that the property of intelligence has a place in strictly structural descriptions of a conjunctive nature. It is an emergent property, and differs strongly from phenomenal intelligence in that without the single acting agency and intentionality of consciousness (itself an emergent phenomenon) we are outside the realm of morality, in a world of cause and effect. I will argue that the property of system intelligence makes sense when talking about both individuals and sub-individual systems, but less so in descriptions of collectives of autonomous individuals. I shall further make the case that a system that is intelligent has the following characteristics: it performs computations, and has the functions of knowledge and memory. In addition, it has goals. Knowledge, memory and goals are all familiar terms when speaking of the phenomenal, but I will propose that they have structural and functional counterparts that are useful when describing what makes a *system* intelligent.

What is a system?

The concept of *system intelligence* needs some initial justification. On the one hand, the claim is that we are not talking about an intelligence that is a fundamental part of physics, while on the other denying that the phenomenal has any role in it. It might seem to us that either we are providing a physical description, or we are not, which in most contexts leaves us with intentionality and consciousness. So which one is it? My answer is that the property of system intelligence is part of an emergence of dynamics happening on a macro scale. Philosophers have offered many widely different accounts on the nature of emergence, and discussions on the subject suffer from it. Nevertheless, it is common to distinguish between weak and strong claims of emergence. By David Chalmers definition, a high-level phenomenon is strongly emergent with respect to a low-level domain when the phenomenon arises from that domain, but truths concerning that phenomenon are not deducible even in principle from truths in the low-level domain. It is weakly emergent if truths concerning that phenomenon are *unexpected* given the principles governing the low-level domain.[134] The former position somehow has to explain the causal powers of the emergent phenomena, a sort of downwards causation that seemingly violate the laws of physics.[135]Weaker claims of emergence do not have to make such admissions, and usually defend the concept's viability by emphasizing that emergent dynamics have explanatory powers where the goings on in the parts are chaotic and unintelligible. There are for example no discernible or comprehensible patterns in the individual molecules of stars and planets, but the dynamics that cause them to attract and form celestial bodies allow us to speak of planetary orbits and motions.[136]Strong emergence entails ontological commitments, while weak emergence 'merely' denies that lower level explanations are proper and useful descriptions of emergent dynamics. The central claims of system intelligence, as described here, belong in the latter category. It is a conceptually emergent phenomenon. We can describe a system as intelligent in purely physical, dynamical terms, without positing intelligence as a fundamental part of nature. This necessitates similar non-phenomenal accounts of concepts like knowledge, memory and

---

[134] Chalmers, David (2006). Strong and weak emergence. In P. Davies & P. Clayton (eds.), *The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion*. Oxford University Press.
[135] See Kim, Jaegwon (2006) "Emergence: Core Ideas and issues." In *Synthese* 2006:151: 547.
[136] See Bokulich, Peter (2013) "The Physics and Metaphysics of Computations and Cognition" In Müller, Vincent (ed) *Philosophy and Theory of Artificial Intelligence* Berlin: Springer

goals.

What does a non-phenomenal account of intelligence entail? In the famous case of the philosophical zombie, where someone has all the physical attributes of regular people down to the molecules, except having any experience, we should expect this person to perform intelligently on par with ourselves, except *there is not really anyone home to be intelligent.*[137] The 'empty' body certainly would appear to be intelligent though. In fact, the entire point of the thought experiment is that we would not be able to spot the slightest difference. The complex system constituted by the zombie's body, or a robot, apparently can solve any problem that a conscious system could, thus having the property of intelligence. What then, is this intelligence *for*? Well, it is useful to the system. Even when separated from the concept of consciousness, we can describe a system as having complex goals of maintaining its integrity. *Oxford Dictionaries* defines a system as "A set of things working together as parts of a mechanism or an interconnecting network; a complex whole."[138] The term system is a template, that can be applied to anything from a single bacteria to populations of animals and cities. In all of them, we find structural similarities, in teleology, in organization, and so on. The concepts used to describe these similarities are not part of physics, but can explain and predict complex dynamics that today's physics cannot. Just any structure will not suffice for such a description. A physical structure that is fundamentally random or chaotic, or not a 'whole', is not a system.

Whether it is a bustling city, an advanced computer or a biological organism, a system has emergent functions and properties that does not make sense to speak of when looking at its individual parts. A monkey may be running, with its body in a state of fear, but in no part of the animal at the cellular level do the concepts of 'fear' and 'running' apply. That does not mean that it is any less true that the physical description of the state of the monkey is in fact fear, than it is true that certain neurons in the brain are firing above the usual rate, or that atoms are rearranging in that exact location in the universe. Emergent properties exist, at the very least as meaningful concepts, and of these, the concept of a system itself may be the most fundamental at this level of description. That is, in my opinion, justification enough to proceed, given that the subject of the thesis is a conceptual framework. If I were still pressed to provide some kind of answer to how systems and their properties *really* exist, I would have to give some pragmatic answer: that structures that can inform predictions and theories (in

[137] This argument against physicalism is most famously known from Chalmers, David (1996) *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press
[138] https://en.oxforddictionaries.com/definition/system 03.02.2019

ways that are not superfluous) must be said to exist somehow. This view, structural realism,[139] seems to provide a possible affirmative explanation for the ontological status of systems and their properties, but the usefulness of system intelligence in our context hinges on its suitability as a descriptive concept.

Some kinds of descriptions do seem out of order when speaking of structures and systems. Morality looks to be a part of the realm of agents and the phenomenal. When we are looking at a system, we see causes and not reasons. We might say that a system should do so and so to function correctly, but the reasons are relevant to conscious beings. The reasons why a computer should perform adequately is that it matters to humans, unless the computer itself happens to experience. The reason why the brain should behave normally is because it is of importance to the agent it is a part of. Whether we are concerned with the capacity to experience pain and pleasure or rational agency, it is consciousness and agency that determines both moral duty and moral status. There are of course plenty of cases where unconscious entities are considered moral subjects. Comatose patients, plants, and even entire ecosystems no doubt have a place in our ethics. Sometimes we ascribe them some proto-phenomenal existence, some pantheistic agency, sometimes we consider their potential for experience. Very rarely, though, do we ponder the moral status of stones and rocks. Consciousness is of singular importance, as it enables both the capacity to experience and the autonomy of moral agency. This means that even though we speak of a structural system's interests, it does not follow that we have any moral obligations to attend to them. In the field of AI, people like Kurzweil emphasize that some unconscious intelligent computer systems should be taken into moral consideration.[140] However, these are special cases, much like the hypothetical philosophical zombie, where we might exceptionally transcend the system level of description of non-conscious entities, and feel obliged to treat them as-if conscious, for reasons like behavioural similarity, or that harming them might affect *us* negatively. We are no longer speaking in structural terms. At the system level, there are causes and effects, and no room for morals.

At first glance, the concept of system intelligence might still appear somewhat reductionist, equating intelligence with the mechanical workings of the body, or a machine. It is true that this too is a description of system intelligence, but the concept encompasses a wide range of differing views. The Cartesian mechanistic view of organisms was criticized already

[139] See for instance Ladyman, James, (2016) "Structural Realism" In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/structural-realism/ 20.03.2019
[140] Kurzweil 2012:214

by Kant, who considered them irreducibly complex systems. In Kant's view, the emergence of parts in an organism was the result of internal interactions, as opposed to the assembly of pre-existing parts, like in a machine.[141] Though science for most of the 18th and 19th century treated life as just complicated cases of mechanical systems, biologists and neuroscientists researching higher order functions like intelligence today generally do not shy away from speaking of things like 'emergent properties' in the context of biological systems. One such property is intelligence.

System intelligence describes intelligent systems, where *intelligence* is a property of any system, mechanical or biological, that is structured in a certain way: by means of computation, the system builds up a store of knowledge in memory, about what works and what is counterproductive to reaching its goals. This is not saying that there is not more going on in a particular system, this is not an argument for a computational theory of mind[142], but any system performing something akin to this process qualifies as intelligent. If we speak of humans as systems, their system intelligence is *not* simply a description of the physical realizers of the experiencing mind, though it is a requisite for such a thing to exist. It is itself a property of a whole that cannot be meaningfully reduced to and explained by its separate parts, with inner complex mechanics that play no part in our descriptions of phenomenal agency and experience. The concept of system intelligence is a description of this phenomenon one way or another, though the way this intelligence relates to previously mentioned other concepts like computation, memory, knowledge and goals differs from the phenomenal accounts. We will attempt to clarify these concepts in turn, but first we need to understand what qualifies as an intelligent system in this context. The idea of computer systems is probably quite familiar, but the notion of humans and other animals as systems deserves some elaboration. The next part of this chapter is therefore devoted to biological systems, and what intelligence signifies at this level of description. Subsequently, we will try to expand upon our generalization of the concept of system intelligence, drawing on the efforts of the field of AI.

---

[141] Kant, Immanuel (1790/1952). *Critique of judgment*. Oxford: Oxford University Press.

[142] Searle's «Chinese room argument» is a well-known critique of the idea that what the human mind does can be accomplished solely through computations. Proponents of the computational mind theory include Hilary Putnam –see "Brains and behaviour" (1963)

Biological systems

The understanding of humans as biological systems has greatly developed in recent years. (Harvard Systems Biology Ph.D. Program matriculated its first class in 2005.) System biology, the study of connected groups of biological parts that all work together, has in large part been made possible by scientists working with huge amounts of data. The goal is to understand how systems make life and its characteristics, like intelligence, possible.[143] Dynamics that arise between cell groupings are not captured by our descriptions of the actions of the individual, nor can they be accounted for by studying the activity of individual cells. The parts that make up the system of a human are themselves systems, and some describe these sub-systems as having their own intelligences. Richard Gregory wrote in *The Intelligent Eye* (1970) that intelligence can be found in the procedures of vision. If perception is seen, not only as responses to stimuli, but also as active (though unconscious) decision making, then we can allow it intelligence. The reason to do so, he argues, is that recognition of objects from the limited data of the senses requires intelligence. Perceptions are richer then the sensory data, and the variety of optical illusions attest to the interpretations our system of vision is performing.[144] When you successfully catch a ball, you were not doing calculus. A smaller subset of you performed the necessary calculations and solved the problem. System intelligence is unlike phenomenal intelligence in that it is not observed as a single (though compound) parameter of a clearly defined entity: it is distributed and cumulative. Not only are animals intelligent systems, but they are made up of smaller systems that themselves have the property of system intelligence. We can speak of the entire individual as a system, or we can apply the term to any connected subset of it. Of course, this breaking down of the entire individual into smaller intelligent subsystems need not stop at the level of larger modules of the brain or body. We are used to treating neurons as the things that are exclusively controlling the sophisticated actions of animals, but this line of thinking overlooks the complexity of single cells. There is a complicated control system that governs the behaviour of cells, and the responsible structure looks to be in the *cytoskeleton.* The cytoskeleton makes up the frame that holds the cell in shape, but in addition it seems to contain the 'nervous

[143] Valcourt 2017
[144] Gregory, Richard (1970). *The Intelligent Eye*. Mcgraw-Hill.

system' of the cell. Our neurons have their own personal control systems. The exact part of the cytoskeleton that performs this function appears to be the *centrosome*, which among other things is responsible for initiating cell division. The control of the cell's movements and its organization is not under the control of the nucleus, which controls the cell's heredity. This means that when a synaptic connection is made in the brain, we do not have a single computation at work, but an enormous amount for each variation of strength in a single connection. This has been emphasized as an argument against tech-optimistic comparisons of machine computer power versus brains,[145] but it also serves as a reminder that human intelligence is a result of systems made of systems. It is unclear what the basic computational unit is, which makes understanding how the parts interact and perform together important. Observing the problem solving dynamics of the system may raise questions and provide answers that we would not come across by focusing on the reasoning individual or the fundamental parts.

When we were describing phenomenal intelligence, the emphasis was constantly on the experiencing and acting individual agent. Nuancing the view of the individual as the essential intelligent unit is not psychologically intuitive, but in the field of biology, this has been necessary for quite some time. Unlike the functionally identical computational units of a computer, individual biological cells have their own agendas. The evolution of animals began when some cells submerged their individuality, and became parts of larger, joint ventures, called multicellular organisms. As we know, the total mass of a human being is in large part made up of different bacteria and sub-cellular parts (like the mitochondria) with their own DNA, but the evolutionary origin of every part of the body is thought to be quite similar. The chemical transmissions between neurons is the residue of ancient signalling between organisms, pressed inward. What is new in this relationship, however, is that the influence is targeted, and able to reach just a few, distant cells, constituting an organized network. In a large system like a multicellular organism, coordination is key. A multitude of micro-actions must be shaped into a macro-action.[146] The comparative degree of cohesion and agency of the cell with regard to the collection of cells is what we have to discern whether a group of autonomous cells are gathering to increase their overall fitness, or if a higher-order autonomous entity emerges. As is the case with most evolutionary adaptations, intermediates of this process is readily found in nature today. The Portuguese man'o war (*Physalia physalis*)

[145]Penrose, Roger (2005) *Shadows of the Mind. A search for the missing science of consciousness.* London: Vintage

[146] Godfrey-Smith 2016:23

for instance, is a colony-individual, composed of four kinds of polyps and medusoids with different genomes that have gone through a joint evolutionary process, resulting in a distribution of tasks like digestion, navigation and defence.[147] There are also cases where the relationship between distinct individuals and the collective when it comes to intelligent decision-making becomes extremely blurred, for instance in the colonies of bees, ants and termites. At times, these colonies collectively carry out the functions of intelligent systems to such an extent that some approaches to artificial intelligence have sought to emulate them in order to design adaptive, decentralized, flexible and robust systems- so called *swarm intelligence*.[148]However, this is where we are venturing into a grey area of what system intelligence can be.

If we understand collectives of animals as candidates for information-processing intelligent systems, meta-organisms, we may even view the species as very large structure. The analogy between learning in individuals and evolution in species is at least as old as the idea of evolution. It has been argued that learning, intelligence and the acquisition of new knowledge can only arise from selected variation, and that any theory of intelligence must acknowledge multiple levels of variation and selection. Jonathan Schull produced the thesis that plant and animal species as multigenerational entities process information via multiple levels of selection and variation, in a manner that is very similar to what goes on in intelligent individual animals. As processors of information, the species are no less complicated than monkeys, and they should be studied for the light they can throw on intelligence.[149]Their behavioural competence in evolutionary time is comparable to individual competence in real time. They can be likened to giant amoebae: asexually reproducing superorganisms whose genetic material are constantly renewed and modified. Each individual is potentially immortal, and has existed continuously, with improvements. Both amoebae and species are comprised of highly differentiated semiautonomous components, be they organelles or organisms. Species are however more complex, and have better adaptive capabilities. In animals, information is integrated by nervous systems, in the species it is integrated by the generation turnover of the gene pool. The information-processing capabilities of the gene pool surpasses the most sophisticated of artificial systems.[150]It might appear that the property of system intelligence also should be ascribed groups of individuals.

---

[147] Gould, Stephen (1985*). The Flamingo's smile* (Chapt. 1.5). New York: W. W. Norton.
[148] Bonabeau, Eric, Dorigo, M., Theraulaz, G. 1999
[149] Schull, Jonathan (1990) «Are Species Intelligent?» In *Behavioural and Brain Sciences* (1990) 13, 63-108
[150] McClelland, J., Rumelhart, D., and the PDP research group (1986) *Parallel distributed processing: Explorations in the microstructure of cognition* MIT Press

However, there are reasons to maintain the individualistic view of the phenomena of both life and intelligence. The process of evolution occurs in the context of a population of autonomous systems, and Schull's definition of autonomy is not clear. If we define autonomy as a property of a system that builds and actively maintains the rules that define itself, then we have a description that is most soundly applied to the metabolism and functions of individuals and their physically connected parts. Even though this individuality cannot be severed from a wider collective organization, philosophers of biology generally maintain this autonomy as the roots of functionality, and as a consequence a premise for ascribing the functional property of system intelligence.[151] My view is that this objection to Schull's proposal is sound, and that we should not say that species are intelligent. If we did, we would probably have to admit nations intelligence, perhaps even the entire system of Earth. If the parts of a system themselves have a sufficient degree of autonomy, then the autonomy of the collective system encompassing them is more aptly attributed to those parts. Exactly what is sufficient to qualify as an intelligent system is not clear. There are examples of systems that clearly have the property of intelligence, like individual mammals, examples of systems that clearly do not, like the solar system, and there are some instances that are hard to define, like a colony of bees. The reason for this, as we shall see, is that even the characteristics we ascribe intelligent systems, like knowledge, computations and goals are somewhat loose terms when applied at a structural level. I do not view this as a significant weakness of system intelligence as a concept, we should expect that concepts have problematic border cases, but it is a sign that attempts at a thorough ontological grounding for the framework would be difficult to carry out. That is not the goal here, but I will return to the limits of system intelligence in the concluding chapter. Let us first examine what the premises are for ascribing systems this property. When we attempt to define the characteristics of intelligent systems, we are looking for specific functions. The human genome consists of about 20.687 genes in total, 25.000 fewer than rice or wheat. The complexity of the human mind is not a matter of gene numbers, but of sophistication of gene networks.[152]What then, are the structural characteristics that separate an intelligent system from other complex systems?

[151] Ruiz-Mirazo, K. & Moreno, A. Synthese (2012) 185: 21. https://doi.org/10.1007/s11229-011-9874-z
[152] Mukherjee 2016:322

Characteristics of system intelligence

So far, we have been looking at humans and animals as intelligent systems, but if we want to speak generally of the characteristics of such systems, the most developed conceptual framework is found in the field of AI. In the following chapter, we shall make use of that. Speaking in computer terms, we may ascribe an intelligent system the following structural characteristics: 1. Building selectivity into the generator. 2. Generating new strategies by modifying partially successful strategies. 3. Accessing large stores of potentially useful information. 4. Concentrating search in the direction of probable solutions. 5. Replacing the search for solutions with the ability to recognize solutions.[153] These are functions of system intelligence commonly recognized by computer scientists. They also apply to biological systems, which becomes more apparent with some slight paraphrasing and clarification. In nature, the survival of the fittest is the well-understood selector. The means by which strategies are created and modified is by *computation*. The store of information is the system's *knowledge*, accessed through *memory.* The direction of probable solutions is the direction of the *goal*, and the ability to recognize the solutions is facilitated by *reward* functions. These are the components of system intelligence we will be analysing. Through them, we can address what a system needs to be able to accomplish to qualify as intelligent, and what it means that a system is *trying to accomplish* something.

Knowledge

What is knowledge in the context of system intelligence? How can knowledge be described in structural terms? In the field of AI, various approaches to creating knowledge in systems have been successful in practice. Many have been inspired by the epistemic logic of Jaakko Hintikka, who suggested that knowledge is the elimination of uncertainty.[154] The system has the knowledge that P in a world iff all the not-P worlds are ruled out. This way of regarding the abstract concept is highly practical when evaluating the knowledge of a physical artificial system performing calculations, but to describe knowledge in proper structural terms

[153] Newell, A. and Simon, H. (1976) "Computer science as empirical inquiry: Symbols and search." In *Communications of the ACM* 19: 113-126.
[154] Hintikka 1962

the propositions that make up knowledge must be treated like concrete objects. Here, AI research has been the main driving force. In traditional approaches to artificial intelligence, creating machines with banks of knowledge is at the very centre of attention. Summing up the different positions on AI in their article *Enhancing Intelligence* (1994), Roger Schank and Lawrence Birnbaum Large claimed that intelligence is at bottom a function of knowledge, and insisted that bodies of knowledge must make up the core of study of intelligence.[155]In the early days of AI, the units of knowledge were mostly held to be linguistic objects or symbols, or first-order logic. An influential understanding of the mechanisms of knowledge was formulated by Marvin Minsky in the 1970's. When one encounters a new situation, Minsky wrote, one selects from memory a structure called a *frame*. A frame is a data-structure for representing a stereotypical situation, and several bits of information is carried with it, like how to use the frame, or what can be expected to follow, and these frames are linked in frame-systems. These atom-like frames are in traditional AI the building blocks of knowledge. Thinking always begins with imperfect plans and images, is the basic idea.[156]For the most part unconcerned with the nature of mental content, AI-research has focused on how to create a machine with the *functions* of knowledge. Many considered it more important for a machine to have large banks of knowledge than to have sophisticated algorithms. To perform intelligent behaviour, a system has to write down in some manner figures, symbols or frames that correspond to the state of the world, and this process goes by the name of *knowledge representation*. The object, the representation itself, is thus characterized by being a surrogate for something in the real world, and it carries with it an ontological commitment to what is important to calculate.[157] This is often referred to as the *symbolist* approach to AI. Thinking or computation is the manipulation of these objects of knowledge. The tempting conclusion is that this is not just a description of how a machine could attain knowledge, but that it also must be true with regard to how human (and animal) knowledge is realized. Much in the same manner as the structure of system intelligence is more important than complexity, so is the structure of knowledge more important than sheer volume of data. One of the main challenges in the development of AI going forward, is transforming all the unstructured potential

---

[155] Schank/Birnbaum «Enhancing Intelligence» in Flynn, James R. (Ed) (2009) *What is Intelligence?* Cambridge: Cambridge University Press

[156] Minsky, Marvin (1975) «A Framework for Representing Knowledge» in *The Psychology of Computer Vision*, P. Winston (Ed.), McGraw-Hill, 1975.

[157] Davis, Randall, Shrobe, H., Szolovits, P. (1993) "What Is a Knowledge Representation?" in *AI Magazine* Volume 14 Number 1 (1993)

knowledge of big data into structured knowledge.[158] Knowledge is searchable, meaningful and coherently ordered. It is *learnable*.

In the non-phenomenal account of knowledge, as it relates to system intelligence, it is still considered a representation of real world objects and states of affairs, but it is not essentially a mental representation. There is no need for an ultimate observer recognizing the knowledge, as has troubled the philosophy of mind. It is a matter of physical cause and effect, and what separates knowledge from just plain data or information is the way that the representations are structured and accessible to the system. By removing experienced *beliefs* from the equation, we are left with what may be described as a 'mere' functional account of knowledge, but also one that clearly separates the concept of knowledge from that of consciousness. When a system knows something at a structural level at description, very much the same happens as when a conscious agent has knowledge. It adjusts its actions accordingly. Knowledge is of paramount importance to an intelligent system, but defining it along the lines of justified true belief is unhelpful, in that it suggests that a system knows in the same manner as a conscious human. In the context of system intelligence, knowledge is relevant information readily available to the system: basic structures that correspond to the state of the world, that when manipulated in larger structures allow the system to predict the outcome of, and rule out, different paths of action. Without such knowledge, solving novel problems of complexity is not possible.

Memory and computation

In a complex and changing environment, a system cannot act intelligently without the capacity to learn, and it cannot learn without a memory. An intelligent system has to perform tasks of computation, and be able to remember the results. What then, is a non-mental account of memory, and what is the relationship between computations and system intelligence?

Structurally speaking, memory is the capacity to store information, and it is crucial for any intelligent system. The conventional wisdom among artificial intelligent researchers- that intelligence is about information and computation, not flesh or atoms, and that there is no fundamental reason why machines can't be at least as intelligent as us- puts information handling memory devices squarely at the centre of attention. These memory devices need to have long-lived states. All memory devices that are useful to humans (books, brains and hard

[158] See Zhuang et al. (2017)"Challenges and opportunities: from big data to knowledge in AI 2.0" In *Front Inform Technol Electron Eng* 2017 18(1):3-14

drives) have in common that they can be in many different long-lived states. This means that changing a bit of information's value would take will require more energy than random disturbances are likely to provide. Engineers prefer to encode information into systems that are not only stable, but also easy to write to. It is much easier to alter the state of a hard drive than engraving gold. When we treat intelligent systems as information processors, common ground emerges in the language we use to describe them. Human DNA stores about 1.6 gigabytes, comparable to a movie. The human brain stores 10 gigabytes electrically (specifying which neurons are firing at any one time) and 100 terabytes chemically (specifying how strongly different neurons are linked by synapses).[159] Of course, we don't mean to say that this is any kind of indication of the measure of intelligence in humans, but the fact that information has the property of being easily transferable between vastly different devices has proven extremely useful. Because of information's substrate independence, engineers are able to replace the devices inside our computers with dramatically better ones, without requiring changes to our software. Substrate independence does not mean that a substrate is unnecessary, but that most of its details do not matter. Any matter will do as long as it can be arranged into NAND gates, connected neurons or other building blocks enabling computation.

However, a book, even if it contained the total sum of all knowledge, would not qualify as intelligent. Daniel Dennett remarks that (although we must be careful not to conflate this with "genuine intelligence") there is some type of storage of information that is intelligent. Intelligent storage is information stored that can be *used* by the system that stores it, meaning that the system in addition must have some capacity for applying the information in some activity. The information is for the system itself, not merely for its users or creators.[160] Although it is possible to make this last part, the information being *for* the system, about consciousness or understanding in Searle's meaning of the term, in the context of *system intelligence* it is mainly observable activity that puts this information to use that interests us. The general term for these kinds of activities is *computation*. Computation is a transformation of one memory state into another. A computation takes information and transforms it, implementing what mathematicians call a *function*. Functions can be simple, like adding numbers, or complex, like translating language, and we will be discussing them in depth in the next chapter. An intelligent machine is able to use highly complex functions to accomplish highly complex goals. To do this, the machine must inhibit complex *dynamics*, so

[159] Tegmark 2017:60
[160] Dennett 1986: 50

that its future state depends in a complicated way on the present state. The atom arrangement must be less ordered than a rigid solid where nothing interesting changes, and more ordered than a liquid or gas. The MIT researchers Norman Margolus and Tommaso Toffoli coined the term *computronium* for any substance that can perform arbitrary computations.[161]Tech optimists tend to believe that most, if not all, of the processes in a human brain can be reduced to processes of computation.

Recreating the entire computational structure of the brain may be purely speculation at the moment, but it may be sufficient to recreate just some of its structural parts to start building a comparable artificial intelligence. The columnar organization of the neocortex, a part of the brain that is a mammalian specialty, was discovered by the American neuroscientist Vernon Mountcastle in 1957. He hypothesized the existence of mini columns within the columns, and later experiments has revealed that repeating units within the neuron fabrics of each column exists, containing about 100 neurons. How do these low-level units of the brain function? Ray Kurzweil proposes that the basic unit constituting the fundamental component of the neocortex is a pattern recognizer. The recognizers are capable of wiring themselves to one another throughout the course of a lifetime, meaning that their complex connectivity is not genetically pre-specified, but reflect the patterns we learn over time. The cognitive abilities of humans seem to be less a result of logical sensibility, and more of a uniquely modifiable system of pattern recognition. Our learning and our recognition is simultaneous. As soon as we have learned a pattern, we start recognizing it, and we then combine it with other patterns to form yet another recognizable new pattern. The end results are on some level experienced as meaningful. Memories of things like faces are not stored as images, but as lists of features where the elements of the patterns that constitute these features themselves are patterns. These patterns are, in Kurzweil's words, "chunks of knowledge", and they form conceptual hierarchies. In the case of understanding written text, simple geometrical lines and curves form character concepts, in turn forming word concepts. This is PRTM- the Pattern Recognition Theory of Mind.[162] In the 1980s and 1990s Kurzweil lead a team that developed a hierarchical technique based on the PRTM (called Markov models) used in software to recognize human speech and understand natural-language statements. This work was the predecessor to today's Siri and Google Voice Search.[163]As these types of programs continue to improve, they do appear increasingly intelligent to us. Viewing

[161] Tegmark 2017:64
[162] Kurzweil 2012:36
[163] Kurzweil 2012:72

intelligence as a computational concept-forming system of pattern recognition yields results. As we shall see later, there are reasons to believe some problem solving requires more than computation, by utilizing what seems to be additional capacities of phenomenal intelligence. Intelligent processes in systems, on the other hand, are computational of nature.

Whereas system knowledge is still a description of how the information in a system relates to the state of the world, the concept of memory in this context explicitly deals with the manner in which these informational structures are physically created and accessed. To qualify as memory-structures, they have to be accessed and put to use by the system itself, not just by an external actor, as in the case of a book. This process of working with memory structures can be described as computation, and a system is intelligent to the extent that it is able to apply it in order to solve complex problems. So do cities and species perform computations? Do they implement functions to change memory states? There probably is room for interpretations that suggest they do, which means that the premise of systems needing to perform computations to qualify as intelligent does not enable us to confidently exclude systems that are definitely not intelligent. However, in less controversial cases, it is a recognizable signature feature of systems that definitely are.

Goals and Reward

We remember that Tegmark defined intelligence as the ability to accomplish complex goals, and general intelligence as the ability to accomplish virtually any goal, including learning. However, a goal appears to be something that primarily a conscious entity possesses. Humans and animal want and strive towards various ends, but our accounts of that seem grounded in the phenomenal. Is it the case that being conscious is a premise for having a goal?

There are many different views on what it means to have a goal. At one end of the spectrum, it could be argued that to have a goal, use of language is a prerequisite. A less drastic objection would be to make a distinction between genuine goal-directed behaviour and goal-terminated behaviour coupled with the capacity to learn.[164] For the most part though, the crux of the argument in each case seems to boil down to whether it is necessary to *know* what one's goal is. I will argue that a structural understanding of goals where this is not a premise has merit.

---

[164] Dennett 1986:74

Any complex system that sustains problem solving processes has to, as a bare minimum, maintain its integrity. If it is able to do so even in complex and novel environments, this is in itself sufficient to qualify as intelligent, and in many ways this is the main challenge for all forms of life. Unless the explicit purpose of the system's action is to terminate itself, or the dissolving of the system an inevitable outcome, any action that leads to unrepairable loss of the systems' own integrity fails to qualify as intelligent. It will no longer exist to solve any of its other tasks. The crux of a goal is that one outcome is preferable to others, and the ability to be able to continue to affect future outcomes is the fundamental instrumental goal of all systems. If we ascribe complex systems the goal of maintaining their integrity, we stand a much better chance at predicting their future behaviour. To qualify as intelligent, a system needs the functions of knowledge, memory and computation, but if we suggest it has in no sense a goal, then nothing matters and there are no problems to be solved.

Accomplishing a goal is surely related to an ability to succeed or "profit". The important thing is not what an individual is attempting, but that the individual is able to choose their actions in a way that leads to success, to accomplishing the goal. The greater the capacity to succeed with respect to various goals, the greater the individual's intelligence. The emphasis on learning, adaption or experience in most definitions of intelligence implies that the environment is not fully known to the individual, that it contains novel situations that could not have been anticipated in advance. As we have earlier noted about the need for novelty in developing intelligence, it is not the property of a system being able to deal with a fully known environment, but rather the ability to deal with that which cannot be wholly anticipated.[165] This ability is found, to some extent, even in single-celled lifeforms. *E. Coli*-bacteria have a sense of smell, or taste. They detect concentrations of certain chemicals, and can react by moving (using filaments with which they swim) towards or away from them. They are not interested in how much of a chemical is present at any given moment, but in whether the concentration is increasing or decreasing. They will swim in a straight line as long as things get *better*, and switch course if things get worse.[166] The goal of finding and utilizing the proper chemicals can in some sense be attributed basic non-biological components, if we speak of chemical compounds as if they have intentionality. Life, however, has developed the ability to judge whether things are getting worse, in which case you should

[165] Legg, S. & Hutter, M. (2007) «Universal Intelligence: A definition of Machine Intelligence» In *Minds & Machines* 2007 Vol 17: 391-444.
[166] Berg, Howard (2006) «Marvels of Bacterial Behavior» in *Proceedings of the American Philosophical Society* 150, no 3 (2006): 428-42

stop doing what you currently are doing, or better, which suggests you should carry on whatever is working so well. In the case of *E. Coli*, there is not too much conflicting feedback on whether the individual's current approach on the whole counts as an improvement, and simple genetically determined responses will normally suffice. For larger systems, like mammals, facing huge quantities of sensory input, a bit more computation and math is needed. The mechanism that indicates success is *reward.*

If a system has no desire to exercise its intelligence in a way that affects its environment, its intelligence would be unobservable and of no practical consequence. Desire as a basic component of having a goal appears obvious. A simple theory of desire is that it is a matter of having dispositions to act.[167]This and other acknowledged theories have in common that they describe reasons for preferring one action rather than another as subjectively experienced, but recently there have been philosophical attempts at more structural accounts of desire. The philosopher Timothy Schroeder's *Reward Theory of Desire* describes human behaviour in light of research into the brain's chemical reward functions, and places this reward in the centre of our desires and intentions. It turns out that, similarly to the action patterns of *E. Coli*, it is in the cases that something gets *better* that a reward is triggered, regardless of how good things are at the moment.[168] Specifically, the *ventral tegmental area* (*VTA*) and the *pars compacta* of the *substantia nigra* (*SNpc*) areas of the brain are the output structures for the brain's reward system. Their neurons reach out across the structures of the brain, where the dopamine receptors are. The pattern in which the VTA/SNpc releases dopamine functions as a learning signal. When the organism receives unexpected positive stimuli (like food when hungry) the VTA/SNpc neurons fire above baseline rate. A predicted reward causes no deviation in activity. This process of releasing dopamine has been observed changing the strengths of the neural connections, in the same manner as reward-based learning in computers.[169]Like bacteria, humans fundamentally have the ability to detect if things are getting better or worse, but our goals are plentiful, often conflicting, and thus complex. We navigate by the brain's dopamine reward system. By this account of desire, there is no qualitative difference between what we mean when we describe the goals of a relatively simple system and the goals of a rational human. It is a matter of complexity.

Goals can also be ascribed to subsystems within the greater system. In the 1960's,

[167] Schroeder, Tim, (2017) "Desire", In *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/desire/ 20.03.2019
[168] Schroeder, Timothy (2004) *Three Faces of Desire* New York: Oxford University Press
[169] Schroeder 2004: 50

psychologists applying the method of *functional analysis* broke down the different psychological capacities of a system (like depth perception and linguistic abilities) into separate sup-capacities (like discriminating between inputs and evaluating information) and attributed them to subsystems. These subsystems were then treated as "subpersons" who discriminate, evaluate, remember and other psychological descriptions. Although subject to criticism of mereological fallacy, applying predicates true of the whole to its parts, the implied agency holds if we again separate goals from the realm of the phenomenal and view each intelligent agent as composed of progressively simpler and less intelligent agents.[170]This does not need to suggest that the property of intelligence at the level of the individual thereby can be reduced to its parts, because we must distinguish between phenomenal intelligence and system intelligence.

That is not to say that all goals are of the same kind. There are (arguably qualitatively) different degrees of problem solving, from simple chemical reactions, through trees responding in different ways according to complex alternating conditions, to the capacities of animals and humans. System intelligence is problem solving by learning, and to learn you need rewards. The fundamental goal of any biological system though, *qua system*, is its continued existence, through survival, reproduction, legacy or any other means, and the reward functions are the continuous evaluation and correction of the intelligent process striving toward this goal. Artificial systems as well, are usually programmed to not malfunction and dissolve, and are just as dependent on reward functions to accomplish it. However, they may be both extremely capable of problem solving, and explicitly programmed to self-destruct, which seems to conflict with the definition of what the goal of an intelligent system is. I do not think we should conclude that such a system has no system intelligence, but rather view this as an unavoidable demise due to external factors. Goals in the context of systems is a meaningful concept, because it allows us to make sense of and predict the paths of action the system will take. At the system level of description there is no qualitative difference between the goals of bacteria and humans, but this concept must not be conflated with goals and desire in the traditional sense, the phenomenal goals of an entity capable of experiencing wants, or even understanding.

[170] Drayson, Zoe (2014), "The Personal/Subpersonal Distinction." In *Philosophy Compass*, 9: 338-346

Conclusion

System intelligence is not simply intelligence sans consciousness. It is also a concept that is non-essentialist, where the singular intelligently acting agent is replaced by the emergent dynamics driving the processes of intelligence. Exactly what constitutes the system becomes somewhat stipulative, and as we have seen, it has been proposed that entire species or nations have the collective property of intelligence. The concepts we use to describe the dynamics of system intelligence are computations, memory and knowledge. At a perhaps even more fundamental level, a system intelligence has goals. The concept of a goal takes on additional aspects in extremely complex systems like humans, but remain fundamentally similar across the spectre of biological life. Artificial systems are different, both with respect to goals and to other premises dictated by the natural world. For a system to be called intelligent, we ought to demand a kind of general intelligence, an intelligence capable of complex problem solving. Biological systems of this kind tend to be relatively large and complex themselves, because they exist at the scale where a sufficiently complex environment is presented. A complex digital environment may exist in a purely non-spatial realm, and the same rules do not apply.

As emphasised, system intelligence is not phenomenal intelligence under a different name; to say that it exists is not in itself an argument for the human mind being entirely computational. However, those working on artificial intelligence that *do* hold this view, have their work cut out for them. If human intelligence is *identical with* a structural property of the human brain, it means that it is possible to create the same kind of intelligence by identifying and recreating that structure artificially. If we believe that this structural property is substrate independent, only having to do with how the brain computes, then our safest bet seems to be to replicate every single computation.

Whole brain emulation (“uploading”) is the process of producing intelligent software by scanning and modelling the computational structure of a biological brain. In such an uploading, the data from the scanners could be used to reconstruct the three-dimensional neuronal network that implements cognition in the original brain, and this neurocomputational structure could be implemented in a powerful computer. We now have an emulated mind existing as software on a computer, either inhabiting a virtual reality or interacting with the world through robotics. The advantage with such an approach to AI is that we would only need to understand low-level characteristics of the brains computational elements, relying less on theoretical insight and more on technological capability.[171] A very successful emulation would be *high fidelity*, with the full set of knowledge and capacities of the emulated brain, whereas a *distorted emulation* would be significantly non-human in some ways, but mostly capable of what a normal brain can do. A *generic emulation* would be like an infant, lacking all or most acquired skills of the emulated brain, but with most capacity for learning still intact.[172] In any case, a hefty amount of computer power is needed, as a functional simulation of the brain is estimated to require between 10^{14} and 10^{16} calculations per second.[173]

This is of course merely a practical concern. What interests us philosophically here is not what is difficult and what is easy, but what is possible and what is impossible. Even if a high fidelity emulation was successfully accomplished, appearing in all respects to be a disembodied human intellect, philosophers like Searle would hold that the emulation is incapable of understanding, that it is merely extremely sophisticated syntax. My guess is that in such an event, there would be equal debate both of whether this emulation was conscious or not, and whether “genuine” non-conscious intelligence, a zombie intelligence that has all the properties of homophenomenal intelligence except consciousness, is a real thing. The severity of the other minds problem would become more readily apparent and pressing- there currently exists no way of confirming consciousness, through neither science nor philosophy. To debate whether the fully functioning emulation is ‘really’ intelligent, an issue that without doubt would be added to the mix, seems more misguided. Distinguishing between phenomenal intelligence and system intelligence makes this discussion unnecessary.

[171] Bostrom 2014:35
[172] Bostrom 2014:40
[173] Kurzweil 2012:196

Algorithmic intelligence

(D5) *Algorithmic intelligence is the capacity of complex problem solving functions.*

The central concept in this chapter is *algorithms*, but it is so closely related to functions and functionality that a brief explanation of these are needed up front. An algorithm is a set of instructions, a recipe, a function, and what exactly that means is important to clarify. The word *function* has multiple meanings. It can refer to the purpose of an object, be it a human or a chair. It can also refer to simple mathematical relations between input and output. Everything physical interacts in some manner with other physical objects, in accordance with the laws of nature, and the way it does so, we can refer to as its function. Whether we are talking about purpose and intentionality, or purely determined mathematics, this function is an abstract property. Most of this interaction is highly predictable, like molecules reacting by either bonding with or repelling each other. Some function structures however, that are found at least in biological brains, are so malleable and adaptable to novel environments, that philosophers are debating whether, in the otherwise seemingly determined mechanistic universe, there still exists free will. Algorithms capable of solving complex problems when implemented, I will argue, have the property of being *intelligent*, and in recent years many AI researchers have come to the conclusion that the path to artificial intelligence is to find the right sort of algorithm, and then just let 'nature' run its course. The systems where the intelligent algorithms are performed may as a result deserve the label intelligent as well, but the key to creating them is this *algorithmic intelligence.*

Even an intelligent algorithm simply does what it does; it has no desires and can have no goals. Concepts like understanding or knowledge, that we have seen are closely intertwined with the other kinds of intelligence, are not a part of algorithmic intelligence, but may be a result of it, or an additional requirement for other kinds of intelligence. In this chapter, I will make the case that the intelligence of algorithms and the intelligence of systems are two different concepts, and review different existing proposals on what characterizes algorithmic intelligence.

What is an algorithm?

In philosophy, there has been some interest in algorithms as they relate to cognition, particularly in the philosophy of mind, but few have shown an interest in the algorithm per se. Neither do computer scientists have a well-established idea of the ontological status of their main material of programming, and the lack of consensus on this issue has been noted by prominent computer science publications.[174] Robin Hill has given a definition of algorithm that, though not undisputed, serves to clarify the concept of algorithmic intelligence: "*an algorithm is a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose."*[175]It is *finite* in the sense that it allows representations to be articulated in finite time and space, and it is an *abstract*. We can talk about an algorithm independently of its specific instances; it is not situated in space and time. By *effective*, Hill means to say that the algorithm is wholly determined, producing the exact same results given the same circumstances. That means that the causal powers of the physical structure instantiating it can be predicted on the basis of what we know about the abstract algorithm: the problem solving abilities, or intelligence, of the algorithm is equal to that of its realizer. Since no additional problem solving powers can be attributed to the physical instantiation itself, it makes sense to also ascribe the abstract algorithm itself the same property of intelligence. *Control* emphasizes that an algorithm brings about a change from one state to another, and *structure* reminds us that an algorithm consists of smaller units: the ordered procedural steps of the function. *Imperative* means that an algorithm is a prescription, which tells how-to. Finally, that it *accomplishes a given purpose* should make the reader pause at this moment. Was it not claimed in the previous paragraph that an algorithm can have no goal? The key word here is of course "given". The algorithm itself cares not a bit about whether it is performing successfully. It may perform the biddings blindly given by nature, or for a purpose given intentionally by a human scientist, but whenever we talk about *whose* purpose it is, it is either the system(s) in which the algorithm is being performed, or an external agent.

There are several different reasons why an algorithm does not have a goal in the same sense as a system. First, a system is informed of the states of its parts and therefore acts to

[174] Hill, Robin (2016) "What an algorithm is" In *Philosophy and Technology* Vol 29, Issue 1:35-59
[175] Hill 2016

maintain its integrity. An algorithm, being a function, is not informed of its functional structure; it is not even in the loosest kind of sense aware of itself. Therefore it cannot affect itself, or any self-related goals in an intelligent way. Even if the algorithm could somehow know its own structure, it is fully determined. According to determinism, the same could be said of a system, but in that case, a system is not determined by its own formal structure, but by the physical world.

What reasons are we then left with to say that there is such a thing as algorithmic *intelligence*? It turns out that algorithms by themselves have enormous capacity for complex problem solving, and capable of feats of intelligence that we do not need to describe in terms of knowledge, goals, or understanding. The accomplishments of AI-algorithms left to their own devices in recent years have been stunning, at times solving problems far more efficiently than entities we do not hesitate to label as intelligent. Still, precisely because knowledge or understanding play no fundamental part in the programs, they are often dismissed as not *really* intelligent, despite clearly satisfying the criteria of capacity for complex problem solving (D1). My view is that we should acknowledge that algorithmic intelligence is real, impactful, but very different. This very basic concept of intelligence refers to both abstract functions or algorithms that when physically realized would be able to solve complex problems, and their physical instantiations. The abstract functions carry an inherent potential of intelligence, and the physical structure performing the function, for instance a set of neurons, are concrete objects with the property of algorithmic intelligence and causal power. The reasons why this has become a subject for debate are the results of AI-research, and in order to understand its potential significance, we must examine the radically different approach to (artificial) intelligence that *machine learning* represents.

Machine learning

The symbolic approach to AI has been strongly challenged in recent years by what has become known as *machine learning*. As we saw, building knowledge, or more specifically the *representations* of knowledge, is the symbolist path to AI, but when it comes to creating successful programs, this approach faces serious practical difficulties. The most pressing, is that in such a top-down approach, the program has to be carefully instructed in how to navigate its resources of knowledge via sets of rules or decision trees. This makes learning really difficult. Machine learning, on the other hand, is obviously all about learning.[176] Through trial and error working on data sets without explicit instructions, such programs solve problems and make decisions with a level of autonomy other approaches to AI lack. Although it is difficult to develop complex, structured systems using machine learning models, in many ways they represent intelligence in its most basic form.

During the last half of the 1960s, the popularity of neural nets caused *connectionism*, a movement in cognitive science that hopes to explain intellectual abilities using artificial neural networks[177], to take over at least half of the AI field. A neural network is simply a group of interconnected neurons that are able to influence each other's behaviour. Real-world neurons are very complicated electro-chemical devices, of which there are many kinds. AI researchers have however shown that neural networks can attain human-level performance on many complex tasks even if we ignore these complexities and use simple simulated ones that are identical and obey simple rules.[178] The more traditional early approaches still included direct attempts to program solutions to specific problems, such as how to recognize the properties of letters. Marvin Minsky and Seymort Papert's book *Perceptrons* (1969) showed the limitations of feedforward neural nets, and in effect killed most of the funding for neural net research in the 70's, the so called winter of connectionist research. In strictly feedforward nets, the output layer has no way to influence the input layer, thus hampering its learning capabilities. The core theorem was that such neural nets, of which the so-called Perceptron was one of the most famous, were incapable of performing certain tasks involving images that humans do very easily.[179]In the 1990's however, with increased availability of digital information via the internet, and improved neural nets, a real contender to the knowledge-based approach to AI emerged. With vast amounts of data at hand, and increased computational power, the benefits of leaving computers to their own devices, blind but relentless and lightning fast trial and error, became apparent.

Machine learning is the study of algorithms that improve through experience.[180]In other words, they learn. Machine learning goes by different names: pattern recognition, statistical modelling, data mining, knowledge discovery, predictive analytics, data science, adaptive systems, self-organizing systems etc. Learning algorithms, algorithms that produce

---

[176] Sun, Ron (2015) "Artificial Intelligence: Connectionist and Symbolic Approaches" *in International Encyclopedia of the Social & Behavioral Sciences* 2nd Edition 2015
[177] Garson, James (2015) «Connectionism» in *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/connectionism/ 20.03.2019
[178] Tegmark 2017:72
[179] Kurzweil 2012:135
[180] Tegmark 2017:72

new algorithms, are the inverse of programming, where the task is finding the best algorithm for producing the desired output. Pedro Domingos describes machine learning as “the scientific method on steroids”. It follows the same process of generating, testing, and discarding or refining hypotheses, but in a fraction of the time it would take a human scientist.[181] This is the reason these algorithms are called learners. Their task is not to perform operations entirely predicted and instructed by their human creators, but learning itself. Today, just a few algorithms are responsible for the great majority of machine-learning applications. Naïve Bayes is one, the “nearest-neighbour” algorithm another. They are all much simpler than the algorithms they replace, which are explicit instructions in how to go about solving the problems in the way a human would. This leads Pedro Domingos, among others, to speak about the “master algorithm”, a single, universal-purpose learning agent. The thesis is that all knowledge- past, present and future- can be derived from data by a single, universal learning algorithm. Inventing such an algorithm, Domingos says, would be one of the greatest scientific achievements of all time, and the last thing we will ever have to invent, because it will from then on invent for us. [182] In *The Master Algorithm*, Domingos divides the current research landscape into what he calls the five tribes of machine learning: the symbolists, the connectionists, the evolutionaries, the Bayesians and the analogizers. Among them, different theories of intelligence are represented, each with different lines of arguments for, or sometimes against, the existence of a master algorithm. When investigating the nature of algorithmic intelligence, these are the approaches currently available to explore.

Symbolism

The proponents of traditional knowledge engineering, visited in earlier chapters, are the staunchest critics of the master algorithm. According to them, intelligence needs knowledge, and real knowledge cannot be learned automatically; it must be programmed into the computer. Marvin Minsky is, as noted, a prominent member of this camp. His theory of intelligence, expressed in the book *The Society of Mind* (1986), is that the mind is “a lot of different things”. There is no underlying basic principle, but a multitude of distinctly different processes, “agents”, each with their own purpose and methods.[183] Something along this line of thinking is the case with most “good old-fashioned AI”- sympathizers. There are however

[181] Domingos 2015:13
[182] Domingos 2015:23
[183] Minsky, Marvin (1986) *The Society of Mind.* New York: Simon & Schuster.

also members of the machine-learning community that tend to agree with at least some of this. This branch, which Domingos calls *Symbolists,* are an offshoot of the knowledge engineering school of AI, and thus closer to traditional AI than the other schools.

For symbolists, all intelligence can be reduced to manipulating symbols, like a mathematician solving equations by replacing expressions. Much of the work when forming the artificial intelligence may be done bottom-up, without direct guidance, but some seeds of knowledge has to be sown. The symbolists believe that you cannot learn from scratch, you need initial knowledge that can be incorporated into learning. Their master algorithm is inverse deduction (induction), which figures out what knowledge is missing in order to make a deduction go through, and makes it as general as possible.[184] The results are not definite, but tells you what is likely to be true. The challenge harkens back to the problem of induction that Hume famously emphasized: how can we justify generalizing from what we have observed to what we have not?[185] Every learning algorithm, says Domingos, is in a sense an attempt to answer this question. For instance, what good is Google's massive logs of search queries when you type in a combination of keywords that is not in the log? A key idea of symbolist machine learning is that every new piece of knowledge becomes a basis for inducing more knowledge, simply assuming that whatever is true of everything we have seen is true of everything we have not seen. It may not be philosophically sound, as Hume showed, but the same assumption lies behind our notion of the laws of nature, a confidence that have proven remarkably advantageous.

In addition to the raw data, symbolist machine learning starts with a number of hypotheses, the 'knowledge' part, and as long as they survive being tested on the data, they are assumed approximately correct. Through induction, inducing rules from the set of rules you start out with, this knowledge is honed. According to symbolist philosophy, this is fundamentally what intelligence is. It does not matter if the substrate is a professor manipulating symbols according to such rules in his brain, or if a machine switches transistors on and off, algorithmic intelligence is manipulating symbols in the form of logical rules and forming ever more fitting hypothesises about the world. There are many objections to the idea that human intelligence can be reduced to a purely axiomatic system, and we will get to more of them later. One that has become of interest even to philosophers not exclusively occupied with AI, is the so-called "frame problem", defined by John McCarthy and Patrick Hayes: specifying only which conditions are changed by an event does not entail that other conditions

---

[184] Domingos 2015

[185] Hume, David (2003) *A Treatise on Human Nature* New York: Dover (Original published in 1888)

have not changed.[186] Even if a set of rules state that because someone opened the door, the door is now open, there is no way of inferring that the light is still on unless this too is explicitly specified. The obvious way of working around it, including every thinkable rule needed to completely specify what has changed and what has not, is practically impossible even in very simple scenarios.

The fact that human reasoning is able to circumvent this had not puzzled philosophers until we tried to make machines do the same. It has been suggested that a practical solution in AI is to develop systems that can tackle the problem in a human-like way. For humans, the frame problem is very much like the problem of re-identification. When we see someone that looks exactly like a relative, we do not know that it is the same person; we simply assume it until proven wrong. We do not, however, make this assumption of familiar identity about any person; we categorize in our brains some important and relevant individuals that are expected to persist. In general, a solution to the frame problem might be to not just assume implicitly that objects persists through time and are unproblematically re-identifiable, but to explicitly determine what objects are the same that have been encountered previously.[187] In the case of the door and the light, the identity of the light would have been determined, and all properties assumed unchanged. Perhaps this is a plausible way forward, but the problem remains unsolved so far. If it should turn out that it is impossible to eradicate this problem, we might be forced to admit that pure abstract algorithmic intelligence sometimes will appear very unintelligent indeed from a human perspective, and that there is a genuine qualitative difference between what can be accomplished through a simple algorithm and what a human is capable of. It would still make sense to speak of algorithmic intelligence, but it would be a relatively weak and impotent intelligence, far from any ‘master algorithm’. It is possible that this is all there is to algorithmic intelligence, making it likely that other kinds of intelligence not only is conceptually irreducible to algorithms (speaking of things like desires and goals), but also ontologically. The intelligence of biological systems may be based on both the computational powers of algorithms and on things like knowledge as independent factors. This is however not the limit of AI ambition. For many machine-learning scientists, the goal is precisely to duplicate the abilities of the human mind, and nowhere is this goal made more explicit than in connectionism.

---

[186] McCarthy, John, Hayes, P. J. (1969). "Some philosophical problems from the standpoint of artificial intelligence» In *Machine Intelligence* 4*: 463–502.*

[187] Fields, Chris (2013) “How humans solve the frame problem” *In Journal of Experimental & Theoretical Artificial Intelligence,* 25:4, 441-456

Connectionism

For *connectionists*, learning is simply what the brain does, and we need to reverse-engineer it. The crucial problem is figuring out which connections in a network that are to blame for errors, and changing them accordingly. Their master algorithm, as Domingos sees it, is backpropagation, which compares a system's output with the desired one, and then changes the connections in the neurons to bring it closer to the desired output. The Canadian psychologist Donald Hebb argued in his book *The Organization of Behavior* (1949) that if two nearby neurons were frequently active at the same time, their synaptic coupling would strengthen so that they learned to help trigger each other.[188] This simple learning rule (Hebbian learning) allows neural networks to learn interesting things, like storing lots of complex memories by simply being exposed to them repeatedly. "Backpropagation" is basically the same thing, and it allows the networks to learn remarkably complex computations if the training is performed with large amounts of data. In recurrent neural networks, information flow in multiple directions, so that the current output can become input to what happens next. The network of neurons in the human brain is recurrent, letting information input from the senses affect the output to the muscles, and AI researchers have sought to incorporate this.[189]

In recent years, this approach has led to several success stories. DeepMind is an AI designed to learn to play computer games. It started playing the classic Breakout, where the goal is to bounce a ball at a brick wall, picking off the bricks one by one until the board is empty. DeepMind was not programmed with any knowledge of games, or even concepts such as games, bricks or balls. It was a blank slate being fed a long list of numbers: the current score and a list of numbers that the AI was not aware was specifications of the pixels on screen. It quickly learned to outclass human players, and in the process developed novel and surprising strategies one would call intelligent. DeepMind then published their code, explaining their idea deep reinforcement learning. The reinforcement part harkens back to behaviourist psychology: a higher score triggered a reward. The deep learning part was the deep neural net that was being reinforced, learning to predict how many points on average would be gained by pressing a key at any given moment. DeepMind went on to learn and master 50 other Atari games.[190] Deep reinforcement learning had turned out to be a

---

[188] Hebb, Donald (1949). *The Organization of Behavior*. New York: Wiley & Sons.
[189] Tegmark 2017:76
[190] Tegmark 2017:83

completely general technique to produce intelligent solutions to problems. The Google Brain Team upgraded Google Translate (until then based on GOFAI) to use deep recurrent neural networks in 2016, with dramatic improvements.

Connectionists seem to be backed up by neuroscience. Neuroscientists at MIT was in 2000 able to rewire the brain of a ferret, rerouting the connections from the eyes to the auditory cortex, and similarly from the ears to the visual cortex. The result was not a completely disabled ferret: the visual cortex learned to hear and the auditory cortex learned to see, and the ferret kept functioning.[191] Results in humans turn up the same. In congenitally blind individuals, the left visual cortex behaves similarly to classic language regions. The brain regions that are thought to have evolved for vision can take on language processing as a result of early experience.[192] Van J. Weeden, a Harvard neuroscientist, published in 2012 a map of the wirings of the neocortex, showing them to be following a grid pattern, like orderly city streets. This grid structure of cerebral pathways, he wrote, was “pervasive, coherent and continuous”, speaking to a common algorithm across all neocortical functions.[193] The wiring pattern, the organization of cortex is the same everywhere, different parts are able to assume each other’s functions, and the computations taking place within the brain are similar throughout. The ‘poverty’ of the genome, which cannot possibly specify the brain in detail, is a good reason why this has to be the case. The relatively few genetic instructions on how the individual brain is supposed to grow and evolve point to a unity of the cortex.[194]What neuroscience in this case seems to tell us about intelligence, is that a sparse number of algorithms, or even a single one, is able to carry out a great many of the tasks to which we ascribe intelligence. The structure of the system itself is inconsequential compared to the existence of this crucial intelligence algorithm.

One objection to *human* intelligence being fundamentally a matter of algorithms comes from mathematics. In computer science, the two most important classes of problems are P and NP problems. P problems are problems we can solve efficiently, NP problems are problems we efficiently can check the solution to. The “P = NP problem” is whether every efficiently checkable problem is also efficiently solvable. In its original, most familiar form, Gödel’s theorem asserts that, for a sufficiently extensive formal system F, that F cannot both

---

[191] Laurie, v. M., Pallas, S. L., & Sur, M. (2000). “Visual behaviour mediated by retinal projections directed to the auditory pathway.” In *Nature, 404*(6780): 871-6. doi:http://dx.doi.org/10.1038/35009102

[192] Bedny, Maria et al. (2011) “Language Processing in the Occipital Cortex of Congenitally Blind Adults” In Proceedings *of the National Academy of Sciences* 108, no. 11 (March 15, 2011):4429-34.

[193] Wedeen, Van J. et al (2012) The Geometric Structure of the Brain Fiber Pathways *Science* 335, no. 6076 (March 30, 2012)

[194] Domingos 2015:27

be complete and consistent, which means that the P=NP problem is not solvable.[195] : An implication of this is that mathematicians do not simply ascertain mathematical truth by means of knowingly sound calculational procedures, something above and beyond that is going on. A computer, on the other hand, is precisely such a formal system. The Gödel-Turing argument, put forth by J.R. Lucas and Roger Penrose, claims that Gödel's incompleteness theorem shows that the human mind is not a Turing machine. It is the ability to understand that natural selection has favoured, says Penrose, and that ability is not computational.[196] There seems to be a hard to define openness to human cognition, related to the creativity that the linguists emphasize. Does this imply that our brains are not entirely governed by precise physical laws? Penrose does not think so. He hopes an answer can be found in the measurement problem of quantum theory, but so far, no satisfying explanation for how humans circumvent these problems has been provided. Interestingly enough, the same mathematics have been used as an argument *for* the potential of algorithmic intelligence. Domingos calls this "the argument from computer science." The solution to the P=NP problem would have a good claim to being the master algorithm, he writes. The general applicability of a computer, where the same machine is used for any problem, actually points to the possibility of a master algorithm that, when properly formulated, would simulate any other algorithm.[197] Inherent in connectionist approaches to AI, there is thus a view of a more potent algorithmic intelligence. The knowledge of a system, and perhaps even consciousness, is part of the capacity of algorithms.

Evolutionism

Evolutionaries believe natural selection is the mother of all learning, and that all we need to do is simulate it. They work on the learning structure, not just adjusting parameters, but also creating the brain that those adjustments can tune. The idea is that the environment should be the programmer of algorithms, as it is in nature. Their master algorithm is genetic programming, evolving programs the way nature evolves organisms. The key is the fitness function, the survival of the fittest algorithm.

Evolution is a separate pointer to algorithmic intelligence. Natural selection is a mechanism of a type very familiar to computer scientists: iterative search, where a problem is

[195] Penrose 1995: 90
[196] Penrose, Roger (1994) "Mathematical Intelligence" in Khalfa, J. (Ed) *What is Intelligence?* New York: Cambridge University Press 1994:107-136
[197] Domingos 2015:33

solved by trying multiple candidate solutions, selecting and modifying the best ones, and repeating as many times as necessary. Daniel Dennett argues in Darwin's Dangerous Idea (1995) that natural selection is an algorithmic process, a collection of sorting algorithms that are themselves composed of generate-and-test algorithms that exploit randomness in the generation phase, a quality-control testing phase, and where the winners advance by having more offspring.[198] Evolution is an example of how much a simple learning algorithm can achieve given enough data. The input is the experience and fate of all living creatures that ever existed.[199]This argument may on the surface seem to speak merely of the process of natural selection as an algorithmic function, leaving the issue of intelligence in individual organisms untouched. However, the internal make-up of both multicellular organisms and the individual cells that constitute them are themselves actors in (and the products) of evolutionary processes. If this primary force in biology can be characterized as algorithmic, then so can all biological phenomena.

Inspired by this, AI-researchers create evolutionary algorithms set to replicate the reproduction, mutation, recombination and selection of biological organisms. The idea is that the fundamental algorithm of natural selection is what is needed to produce the proper intelligence to solve any problem. A set of candidate algorithms is randomly selected, and their fitness is the basis for the next generation's selections. Recombination is an emulation of parents passing on their properties to their offspring, where two or more selected parent algorithm candidates generate children algorithms, whereas mutation is applied to one candidate, resulting in one new algorithm. This can be continued until a sufficient algorithm is produced.[200]The approach has shown much promise, demonstrating the powerful fundamental learning force of nature. In my view, we should still be cautious drawing conclusions about the characteristics of algorithmic intelligence from this. The risk involved becomes clear when Domingos includes evidence from physics.

In physics, the same equations applied to different quantities often describe phenomena in completely different fields, like quantum mechanics, electromagnetism and so on. Once an equation is discovered in a field, like the wave equation, or the diffusion equation, we can more readily discover it in others. Conceivably, writes Domingos, they are all instances of a master equation, and all the master algorithm needs to do is instantiate it for different data sets. Optimization, the mathematics concerned with finding the input to a

[198] Dennett, Daniel (1995) *Darwin's Dangerous Idea* New York: Simon and Schuster
[199] Domingos 2015:28
[200] Vikhar 2016

function that produces the highest output, plays an important role in most fields of science, and demonstrates that simple functions can give rise to very complex solutions.[201] This line of argument, while valid when speaking of fundamental algorithms of nature, appears too general to describe intelligence. If what we call intelligence is ever shown to be yet another instantiation of a singular function permeating the entire fabric of reality, then we have made an impressive scientific discovery, but also simultaneously emptied the concept of intelligence of most of its meaning. Unless we decide to describe the universe itself as intelligent, what we seek to conceptualize is less fundamental than this. Though natural selection might be the algorithm behind all intelligence, it does not itself have the property of algorithmic intelligence. Some additional complexity beyond the fitness function is needed. We also want to emphasize that being complex does not entail intelligence. Steven Johnson writes in *Emergence* "Emergent complexity without adaptation is like the intricate crystals formed by snowflakes: it is a beautiful pattern but it has no function".[202]What evolutionism does leave us, is an intuitive conception of the ontology of algorithmic intelligence: complex problem solving algorithms evolved through the fundamental algorithms of nature.

Bayesians

Bayesians are not interested in emulating nature, but are concerned with uncertainty. All knowledge is uncertain, and the problem is how to deal with noisy and incomplete information. Their solution is probabilistic inference: Bayes' theorem and its derivates. According to this school of statistics, all learning occurs on the basis of this single formula. The Bayes' theorem tells you how to update your beliefs whenever you see new evidence, and according to the Bayesians, that is the essence of intelligence. Unlike the relationship between beliefs and intelligence in the homophenomenal context, where beliefs are suggested to be exclusive to members of a speech community, the kind of belief we are talking about here is about statistics, probabilities, and rational actors in a very broad sense. Mental content does not factor, what is important is what is likely to be the case, and what preferable action follows from this state of affairs. Learning is a process of belief revision, and intelligence is the ability to perform this.[203]

Yet again, problems that emerge in the field of AI, give general philosophy food for

[201] Domingos 2015:30
[202] Johnson 2001: 20
[203] Joyce, James, (2003) "Bayes' Theorem" In *The Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/bayes-theorem/ 20.03.2019

thought. Belief change has occupied philosophers since antiquity, discussing the mechanisms by which scientific theories develop and debating criteria for revisions of probability assignments. Humans are, however, not perfectly rational, and the introduction of artificial agents that actually live up to the commitments to all logical consequences of their beliefs reveal the full importance of belief revision.[204] The structure that performs this is central to intelligence. This approach shares much with symbolist machine learning. Intelligence is the process of dealing with some form of propositions or symbols, but here the emphasis is on the dynamic that embraces or rejects them. Earlier attempts at describing this dynamic in humans were often very pragmatic, like Quine's influential suggestion that belief revision is a matter of choice, and that these choices should be made in such a way that the resulting belief system is simple, squares with the evidence, and disturb the original belief system as little as possible.[205]

Because of the advances of AI, descriptions of the dynamics of belief revision are now for the most part concerned with logic and sets of beliefs. A Bayesian learner-machine starts with a set of hypotheses about the world, and compares these with the data it receives, marking them more or less likely according to their compatibility. It is a machine that turns data into knowledge, and according to Bayesian statisticians, it is the only way to do it. If they are right, Bayes' theorem is either the master algorithm or the engine that drives it.[206]One recent development in Bayesian learning is the effort to move from traditional statistical analysis to causal analysis of data, pioneered by Judea Pearl.[207] Pearl argues that AI research is handicapped by an incomplete understanding of intelligence, and that it has to abandon reasoning by association in favour of causal reasoning, in order to understand *why* something happens rather than just correlating. Causal analysis aim to infer probabilities under conditions that are *changing.* To share our intuitions for cause and effect, a Bayesian learner has to be situated in a model of the environment.[208]The idea seems like it might very well improve the problem solving capacities of algorithms, but looks to be based on a quite different conception of *understanding* than what we have established. There is no clearly apparent reason why including changes in the environment in the computational data should

---

[204] Hansson, Sven Ove, "Logic of Belief Revision", The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/logic-belief-revision/ 20.03.2019
[205] Quine, Willard (1951). "Two Dogmas of Empiricism". In *The Philosophical Review.* 60 (1): 20–43.
[206] Domingos 2015:31
[207] Pearl, Judea (2010) "An Introduction to Causal Inference," In *The International Journal of Biostatistics*: Vol. 6: Iss. 2, Article 7
[208] *Quanta Magazine* May 15, 2018 https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/ 20.03.2019

result in an experience of the problem, meaning that Pearl's vision either implies a radical theory of consciousness, or is ultimately of little consequence in our context. In my understanding, the latter is more correct.

The Bayesian approach shares much with the philosophy of the evolutionaries, but the central tenet is still the revising of beliefs: Algorithmic intelligence is all about inferring probabilities using Bayes' theorem. In the context of the different concepts of intelligence, system intelligence would be the result of adequately performed belief revision in a system, while algorithmic intelligence would be a property of the structure carrying out this process in the system. If the Bayesians are correct in their assessment of the importance of this process in learning, this might be the only kind of algorithmic intelligence there is. Still, if there is one lesson to take from the previous chapter on heterophenomenal intelligence, it is that we should never underestimate the multitude of possible ways of approaching any problem.

Analogism

For *analogizers*, the key to learning is recognizing similarities between situations, and thereby inferring other similarities. The key problem is judging how similar two things are. Their master algorithm is the support vector machine (SVM), which figures out which experiences to remember, and how to use them to make predictions.[209] Roughly speaking, an SVM-algorithm can be run on a data set, and figure out the best way of separating and classifying the data according to any given parameter, and thus make generalizations. The usefulness of this method has been proven in areas like facial recognition, hand-written character recognition, and it is how companies like Netflix is able to successfully recommend movies to you that you did not expect to like. After the trained human crew at Netflix has tagged hundreds of elements in a movie (like 'car chase', 'suspense' and 'philosophical'), the algorithm is put to the task of identifying which elements seem to matter for any of the several "taste groups" you can be categorized into, and predict what movie you would enjoy watching next.[210]

The underlying philosophical concept of intelligence seems relatively simple. The world is made up of potential information, and the task for intelligent entities is to figure out

[209] Domingos 2015
[210] Plummer, Libby (2017) "This is how Netflix's top-secret recommendation system works" *WIRED Magazine* Tuesday 22 August 2017 https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like 20.03.2019

what is important, and what should be weighed. Intelligence is about synthesis: sorting and classifying data, finding connections and forming structures and models. When all that matters is accounted for, the preferred action of a rational, intelligent actor is more or less trivial and given. This leaves less room for conscious decision making in humans then you would think. There is a lot of debate in the field of neuropsychology on how precisely the brain categorizes sensory input, but consensus about the fact that the human brain categorizes objects into domains like 'living things' or 'inanimate objects', and lets specific neural circuits deal with each of them. Findings from functional neuroimaging demonstrate that both sensory and motor systems are engaged during conceptual processing.[211] The extent to which this supports *embodied intelligence*, the thesis that much of the body is involved in cognitive processes, is interesting in itself, but for our purposes, it simply highlights categorizing algorithms in biological organisms. Different intelligent subsystems in the brain carry out conceptualizing tasks pre-consciousness, and if we ask how, intelligent categorizing algorithms are our current best bet. Therefore, it seems like algorithms of this sort should be taken very seriously as one, but not necessarily the only, kind of algorithmic intelligence. If it turns out that the most important function in the human brain belongs in this category, we might have to consider it supremely potent, but there are as yet no conclusive empirical evidence, and many other algorithms that easily rival the results of the artificially created variants.

Conclusion

As we have seen, several fundamental intelligent algorithms have been proposed. Some view algorithmic intelligence as most importantly the ability to manipulate symbols, which means that some kind of external knowledge has to be introduced to the algorithm. Others view the algorithms of the human brain as exactly the kind of algorithm any intelligent system needs, and seek to replicate this through methods of backpropagation in neural networks. Yet others seem to favour the idea that algorithmic intelligence is much more widespread and fundamental in nature, and that it is the force of natural selection we need to apply to cultivate the right kinds of algorithms. Some are more specific in their claims, suggesting that intelligent algorithms essentially are structures that revise beliefs, or that

---

[211] Mahon, Bradford, Caramazza. A. (2009) "Concepts and categories: a cognitive neuropsychological perspective" In *Annual review of psychology* vol. 60 (2009): 27-51.

intelligent algorithms deal with synthesising sensory input into conceptual categories.

We can differentiate between a strong and a weaker claim of algorithmic intelligence. In the idea of the master algorithm lies the strong claim that algorithmic intelligence is capable of, and in fact responsible for, anything any intelligence can do. The weaker claim is that algorithmic intelligence is a basic abstract form of intelligence that performs some part in any intelligent process, and by itself is capable of much, but not all, that any other kind of intelligent agent or system can do.

The notion of algorithmic intelligence is not only visible behind a seismic shift in artificial intelligence, from top-down knowledge engineering to the bottom-up approach of machine learning, it also seems to be on the horizon in neuropsychology, as researchers delve further into the nitty-gritty details of the brains cognitive processes. Systems and algorithms are, however, very different subjects. At least until a master algorithm is discovered, we should be precise about whether we are talking about the capabilities of an abstract algorithm, or the sum of properties of a physical intelligent system.

Conclusion and discussion

The relationships between the different concepts of intelligence

The general definition of intelligence in this framework is a description of a property: (D1) *Intelligence is (capacity for) complex problem solving.* A problem may be viewed as intrinsically a problem, like an abstract equation, or it may be a problem *for* something, or even for *someone.* The different concepts of intelligence correspond with the levels of description we apply when we speak in terms like agency and intentionality. However, this does not entail that it is possible to construct an entirely uncontroversial hierarchical model of the four concepts of intelligence, from basic abstract functions to thick concepts of conceptual understanding. Each concept of intelligence comes with its own particular assumptions, propositions, and claims of relevance, and thinking in terms of different concepts of intelligence will result in different paths, decisions and real world consequences. Related concepts also take on different meaning in context of different concepts of intelligence. Knowledge in the context of homophenomenal intelligence is about experienced beliefs that are structured by language. In the context of heterophenomenal intelligence, it is a conscious knowing-how. When speaking in terms of system intelligence, knowledge are specific types of structures of information available to the system itself. In algorithmic intelligence, knowledge plays no part at all.

What is the most basic kind of intelligence? As seen, philosophers like Searle demand nothing less than human-like understanding for an entity to be described as intelligent in any meaningful way. Among AI-researchers, there is a schism between those who view intelligence as an emergent property of systems capable of sustaining the functions of knowledge, computations and memory, and those who view intelligence as reducible to algorithmic functions. Of the two concepts, algorithmic intelligence is clearly the most basic and fundamental. (D5) *Algorithmic intelligence is the capacity of complex problem solving functions.* The point of contention is whether ascribing algorithms such a property is justified. As the differences in knowledge engineering and machine learning approaches to artificial intelligence has illustrated, there is genuine disagreement on what algorithms are capable of by themselves. Between the two camps, there are both factual disagreements and differing philosophical views. It is unclear how much a single algorithm can accomplish. Unity, patterns and order in the brain speak to a strong algorithmic intelligence, and arguments for modality in mind usually also strengthen its case.

As long as the matter of the potential of algorithm remains unresolved, this alone makes the concept of algorithmic intelligence useful to separate from other, more complex and composite forms of intelligence. However, there are lessons to be learned from studying algorithms as intelligent processes that can shed light on fundamental philosophical aspects of intelligence that are hard to get at in complex systems, let alone conscious creatures. What exactly goes on, when an intelligent process is 'solving a complex problem'? Symbolist machine learning experts will tell you it is a process of manipulating symbols, honing ever more correct hypothesises about the world through inverse deduction. Connectionism claims that the central function of intelligence is backpropagation, comparing the actual output to the desired output, and adjusting accordingly. Evolutionists view intelligent processes as simply an extension of evolution's fitness function. For Bayesians, having the property of intelligence boils down to being able to infer probabilities and revise beliefs. Analogizers view intelligence as categorizing, spotting similarities and relevancy and weighing it. Judging by current results in AI, all of these theses appear to have some merit, and it is quite plausible that the most intelligent functions must incorporate all of these capacities in their formal structure.

When proponents of knowledge engineering are dismissive of algorithmic intelligence however, they do not merely disagree on the empirical evidence for human intelligence being algorithmic in nature, they contest the concept on philosophical grounds. What could possibly make an entirely disinterested action intelligent? When we talk about the intelligence of a system, we are not just referring to the sum of a host of algorithmic functions, we are taking a more intentional stance, and describing it in terms like having *goals*. (D4*) System intelligence is (the capacity of) complex problem solving systems.*

The descriptions of an intelligent system certainly cannot be reduced to descriptions of intelligent algorithms, but the question of whether this means that there is an actual ontological irreducibility is a big and difficult one. Generally, those who have strong faith in the capabilities of algorithms, like connectionists, will say that there is an alternative description of algorithmic intelligence to fit any description of system intelligence, while the claims of the symbolists imply that the intelligence of a system is not entirely accounted for by the algorithms. I have not taken any clear position on the potential of algorithms, but defended the view that systems exist in a real sense, which means that whether or not every intelligent process in a system reduces to algorithms, the intelligence of the system and the intelligence of the algorithm are different concepts. As I see it, any defence of the existence of systems is likewise a defence of irreducible properties of systems. I would also argue that the difference between ontologically committed emergence and ”merely” conceptual irreducibility is less obvious than many make it out to be. Irreducible concepts are no small matter. That a human being at the same time is a state of trillions of particles is astonishing, but well within our imaginations' reach. However, attempts at lower-level descriptions of beliefs and knowledge would not only be complex beyond our comprehension, they would be utterly incapable of capturing the concepts' meaning. Our concepts are how we make sense of our world, but most of them relate only to objects and dynamics on the macro scale, and the fact that many of them cannot be assigned meaningful lower-level descriptions, ought to, at least, have us examining our epistemic justifications for assuming ontological reducibility.

Emergent properties in general is the reason for speaking of system intelligence as something entirely different from algorithmic intelligence. Of all emergent properties, one is uniquely of interest to us, and that is consciousness. Anything that has the property of being conscious, it seems, is radically different from its non-conscious version. Conscious intelligence as a separate concept is no exception. (D3) *Heterophenomenal intelligence is (capacity for) conscious complex problem solving.* System intelligence can be ascribed goals of maintaining its integrity in various ways, but without the property of consciousness, the intelligence is more abstract. We conscious observers can describe the system as working toward a solution, but goals are attributed by external observers, *as if* that is what the system seeks. How clear and explicit is the difference between system intelligence and phenomenal intelligence? It seems that much in the same manner that there is a gliding scale between non-conscious and conscious states, there might be gradients of these two. The same goes for systems' acquiring of the property of intelligence. Recall the proposal that species and nations are intelligent.

Species and nations, in my view, fail to fully satisfy the criteria of intelligent systems. We may speak of emergent goals of an entire species or nation, goals that are not simply reducible to the goals of individuals. We may even speak of collective knowledge and memory (the latter is part of discourse, albeit a controversial one, in the field of history) as aspects of such a collective. The term computation, on the other hand, can only be applied in the loosest possible sense in the context of nations or species, when speaking of dynamics above the level of the individual. However, if we look to corner cases like insect colonies, we see a degree of communication and interaction that begin to rival what goes on in individual animals. There is a possibility that the intelligence property of a system can be ascribed using a scale of integration, where reaching a threshold in the integration of the systems' problem solving efforts results qualifies the conjunction as an intelligent system, and reaching a further threshold of integration results in consciousness. If viewed this way, we might not have clear-cut definitions to determine corner cases like colony-individuals or proto-conscious systems, but we preserve an intuitive conception of what constitutes an intelligent system, and a qualitative difference between phenomenal intelligence and system intelligence is in any case maintained.

How does the subjective intelligence differ from the non-subjective? Whereas system intelligence is a hodgepodge of a myriad processes of problem solving going on simultaneously within the emergent constraints of the system, consciousness is both problem and effort concentrated. The problem with the strongest claim to relevance and urgency is presented in consciousness, where the collective system attends to it with singular focus, as an acting intelligent *agent*. One minor consequence of this is the terminology used to describe what this intelligent agent does. We may for instance use words like 'intending' and 'wanting' without feeling a bit silly, because we know that there is a literal world of difference between an entity that actually experiences wanting to survive, and a system we merely *describe as* seeking to survive. The major consequence becomes apparent when we similarly use words like 'striving' or 'hurting'. Phenomenal intelligence comes with ethical implications. Although some argue that we have moral duties even toward non-conscious entities, it is far more common to hold the ability to experience pain and pleasure as sufficient for at least some kind of moral status.

However, if it were the case that some kinds of intelligent entities have moral status because they are conscious, that would not be a very interesting point in our context. More intriguing is the fact that intelligence seems to be a prerequisite for consciousness. Every known or suspected conscious physical entity so far has the property of intelligence, and we assume there is some degree of correlation between how intelligent a being is (if it was theoretically measurable) and how conscious it is (assuming consciousness comes in degrees). Until we reach the plateau of homophenomenal intelligence, that is. (D2) *Homophenomenal intelligence is (capacity for) complex problem solving based on conceptual understanding.* As for its relationship to heterophenomenal intelligence, I think it makes sense to view them both as categories within the larger category phenomenal intelligence, where homophenomenal intelligence is a subcategory that accounts for some qualitative differences of great importance, at least to humans, and heterophenomenal intelligence refers to any phenomenal intelligence that does not fit the description of the former. Homophenomenal intelligence is often considered the great moral equalizer: if you have the property, you have it to the fullest degree. Challenge anyone to explain what gives humans their special, intrinsic worth, that sets them apart from other animals, and the answer will most likely be something compatible with (D2). This ‘all or nothing’ approach breaks down if we acknowledge some degree of homophenomenal intelligence in animals, and as we have seen, there are reasons to do so. Though the term intelligence is seldom herd in ethical debates, it is as important for consequentialist analysis of pain and pleasure as it is for principal concepts of rational actors. If we were to attempt to describe how morality emerges from physical mechanisms, we could probably do worse than starting with the concept of intelligence. Intelligent systems have, in some sense, goals, the fact that heterophenomenally intelligent entities experience succeeding or failing to reach these goals makes the goals intrinsically *matter*, and the fact that homophenomenally intelligent entities *understand* these goals, and why they matter, enables moral actions. While this mere sketch of an argument can hardly be called an ethical theory, something along these lines seems to be a natural conclusion to draw from the framework as I have presented it, and even though ethics has not been a main concern of mine, I will include concrete ethical problems when demonstrating the potential utility of distinguishing between the intelligence concepts.

Closely connected to the ethical aspects of distinguishing between homophenomenal intelligence and conscious intelligence in general, is the more general notion that intelligence without language may experience solving problems, but not *know* that it is solving them, or *understand* any part of the process. With conceptual understanding comes a meta-intelligence. Not only are you subjectively aware of the problem, but you are aware of your awareness of the problem, making the conscious solving of the problem reflect upon itself and adapt accordingly. In addition, introducing moral concepts to the conceptual framework of such an agent opens up the door to possible responsibilities far beyond self-preservation. This understanding is not only relevant in matters of ethics. As we have seen, the Gödel-Turing argument shows that homophenomenal intelligence is able to grasp problems that a purely formal algorithm would never be able to, given that the P=NP problem is unsolvable. It somehow has acquired a degree of freedom that allows it to work around formal rules. There seems to be two possible answers to why this is the case. Either this is a capacity that emerges in a system complex enough to bring about consciousness, or P=NP is actually solvable, and our brains are simply running superior algorithms to the ones we so far have been able to create artificially. If the latter is true, homophenomenal intelligence appears to be a product of a special class of algorithms, but if not, then this kind of intelligence seems to be genuinely open and undetermined. The varying degrees of determinacy in the concepts of intelligence is notable. From wholly determined algorithms, through systems acting to maintain their integrity, to conscious problem solving, it seems that with greater integration of the problem solving processes comes a larger degree of freedom. Creating human-level artificial intelligence seem to have possible paths both at algorithmic and system level: either through the right algorithm to tackle formal restrictions, or by implementing or cultivating the necessary dynamical structures in the system. Having summarized the relationships between the different concepts of intelligence, it is time to demonstrate some practical applications of the theoretical framework.

Applying the concepts

I believe there are possible decision- and policy-affecting consequences of a more coherent and nuanced framework of different concepts of intelligence in practically every context where the term is relevant. Instead of attempting to summarize or categorize them, I shall provide some examples of what the framework can contribute to discussions on intelligence. These are not necessarily meant to form fully convincing arguments one way or another on the various issues, but illustrate how the framework makes it possible to move beyond simply reiterating the question "But what is intelligence?", and also how more developed concepts of intelligence might find use in new contexts.

One topic of recent fame is that of moral machines. Much has been written on how to best implement ethical values in the machinery and programs that gradually have more power over the lives of humans. [212] One of the most frequently debated issues is the concept of self-driving 'autonomous' cars, and what kinds of decisions they should be making when faced with moral dilemmas. When forced to make a decision, should they choose to swerve the car and possibly cause the death of the one person driving the car, or stay on course and endanger the life of a pedestrian? This might be an important matter to resolve, but the idea that the cars currently in production are making moral choices leads to confusion. The misunderstanding is grounded in the fact that the artificial intelligence implemented in the cars by and large is based on machine learning, meaning that little of the data the machine takes into account

[212] See Bonnefon, Jean-François, Shariff, A., Rahwan, I. (2016) "The social dilemma of autonomous vehicles" In *Science* 24 Jun 2016 :1573-1576

when faced with a situation we describe as an ethical dilemma has been explicitly programmed by humans. In a very real sense, the decision trees of the car are of its own making, and autonomous. However, that is not sufficient to describe them as moral choices. In the chapter on homophenomenal intelligence, I made the case that in order to make autonomous moral choices, you have to have understanding. In other words, for a machine to make a moral decision, it would have to be conscious, or more precisely, have homophenomenal intelligence. That means that so-called autonomous cars are no more an outsourcing of human ethics than any other invented device capable of harming a human being. The moral choice is made by developers and producers. Autonomous cars do however force our hand in deciding whether utilitarian or deontological ethics should generally be applied in traffic.

The same is true for a host of other ethical issues related to artificial intelligence, from surveillance and medicinal systems to drone warfare. We are far from it being relevant that any machine is capable of reflecting and understanding moral dilemmas, what is needed is ethical design (and use) of powerful machinery. This is recognized by today's legal system, which emphasizes that robots are not suitable recipients of criminal punishment because they cannot conceive of themselves as morally responsible agents.[213]Still, we are often left with the impression that they have moral choices to make, even though todays self-driving cars reason no more about the well-being of humans than a wooden sailing ship, intelligent or not. My proposed framework of concepts avoids the risk of conflating intelligent moral actors and intelligent tools.

The generally unacknowledged presence of the concept intelligence in our ethical reflections is another issue. Here, a reluctance against the term compounds its explanatory deficiency. I would argue that both Kantian morals emphasizing our duties toward human uniqueness and consequentialist analyses are coloured by the same anthropocentrism, and monolithic conception of intelligence. The interesting one in this context is utilitarianism, which is the preferred moral system of many engaged in issues of animal rights. Those who view humans as self-explanatorily special cases of moral status seldom attempt to justify the circular logic they lean on. We humans have concepts, language and understanding, and this gives us an intrinsic worth that is different from all other species. Of course, there is no objective value in these properties: they are what makes us essentially human, and since

---

[213] Gless, Sabine, Silverman, E., Weigend, T. (2016) "If Robots cause harm, who is to blame? Self-driving Cars and Criminal Liability" In *New Criminal Law Review: In International and Interdisciplinary Journal,* Vol. 19 No. 3, Summer 2016:412-436

humans have these exact properties we are special. The central idea seems to be that it is self-evident that our essential human properties make us uniquely valuable, which is correct in the sense that humans obviously value these properties above other traits, but it cannot be taken seriously as a moral theory meant to apply to non-human entities.

The utilitarian approach of people like Peter Singer sets out to be a more general moral theory, where the ability to suffer pain and loss regardless of what particular kind of entity you are is the measure of moral status.[214] Not all philosophers would agree that suffering is objectively bad, but as a starting point for a universal moral theory it is at the very least not blatantly species specific. The question is how we measure the capacity for suffering. I would argue that there are two general approaches to this. The first is to say that suffering is somehow quantifiable, by how intense the sensation of pain or grief appears in the subjective awareness of the suffering entity. Let us call that the 'consciousness' view of suffering. The alternative view is to say that we are talking about qualitatively differing experiences, suffering of the kind, and to the extent, allowed for by the entity's cognitive properties. It is not a matter of quantity of qualia involved, but of kind. Let us call this the 'intelligence' view of suffering. The latter is not usually made explicit, but it is clearly often implied when we take into account what kinds of arguments are involved when we for instance discuss the abilities of primates, elephants or birds to suffer. It might appear that we are looking for consciousness when we ask whether they can suffer, but as far as I can tell, we are specifically searching for signs of intelligence. In Tom Beauchamp and James Childress' authoritative account of theories of moral status in *Principles of Biomedical Ethics* (2009), a distinction between a theory based on cognitive properties and a theory based on sentience was attempted. According to them, the theory based on cognitive properties ascribe moral status using premises like "perception, memory, understanding and thinking". The 'intelligence view' I have outlined fits in some respects with this, but mostly in the sense that they both assume essential qualitative differences between entities that are conscious. The theory based on sentience, on the other hand, is more or less Jeremy Benthams' view of the capacity to suffer as the central property, which can be measured independently of intelligence.[215] My gripe with these classifications is that the theory based on cognitive properties overlooks the arguments why such properties are thought to matter ethically (they relate to consciousness), and the theory based on sentience does not address what aspects of consciousness grants the

[214] See Singer 1975
[215] Beauchamp/Childress 2009:69-76

capacity to suffer. The reason in both cases is that the nature and role of intelligence in this account is ambiguous at best.

One might object that consciousness and heterophenomenal intelligence very well could be mutually dependent. Are there instances of consciousness without appropriate levels of intelligence? In nature, we would certainly be surprised to discover such a case, if the question was scientifically answerable. A chimpanzee behaves much more intelligently in our eyes than a garden snail, and we take for granted that a chimp baby in agony is a worse subjective experience than anything the snail is capable of experiencing. When humans look at the world, we see a general rule at work, of more intelligence leading to more consciousness. However, let us revisit Scott Aaronson's critique of Tononi's Integrated Information Theory of consciousness. If consciousness, as according to Tononi's theory, is simply integrated information, then how do we explain that a computer program designed with no other purpose than producing a high amount of integrated information, with no other abilities, is conscious? Tononi's answer, we remember, was that it would be conscious, but not interesting. It would lack phenomenal intelligence. It would be large quantities of qualia of the simplest kind, and we should expect the system to be practically incapable of suffering even if there was a distinct subjective experience there. Tononi's theory of consciousness could turn out to be very wrong, of course, but the example illustrates the point: intelligence could be ethically relevant in ways that today's rough concept does not handle well. The example given here is only the outline of an argument, but one I feel fairly confident could be fleshed out given time and effort. My ambition with this thesis was not to develop an ethical theory based on intelligence, but to show that our understanding of the concept matters when we are dealing with ethics and non-human entities. A universal ethics for the future should acknowledge both the place of intelligence in our moral reasoning, and our epistemic challenges in recognizing heterophenomenal intelligence.

Why is an ethics recognizing the role of intelligence preferable to other ethical approaches to morality and non-human entities? Consider the idea of "personhood". The idea behind the concept, as proposed by Mary Anne Warren, is that a being is a member of a "moral community" if it has the *characteristics of sentience*, *emotionality*, *reason*, *capacity to communicate*, *self-awareness* and *moral agency*.[216] Opinions on the exact characteristics vary, but the logical conclusion regardless of what you choose to include is that either you are a

[216] Warren, Mary (1997) "On the Moral and Legal Status of Abortion" In *Ethics in Practice* La Follette, H. (Ed) Blackwell, Oxford

person, or you are not. This kind of dichotomy when it comes to eligible moral subjects may have merits in legal matters, but if we are interested in facts of the matter about a particular entity's properties, such categorization provides us little information whether the subject is nearly (or barely) a person. Judging an entity's moral status on the grounds of properties like homophenomenal or heterophenomenal intelligence, with the stance that these are properties that can be possessed to various degrees, may not be a convenient way to get simple answers of the kind that the legal system seeks, but these are not the kind of answers that ought to interest philosophers. Establishing homophenomenal intelligence as the premise for moral agency and heterophenomenal intelligence as sufficient for *some* moral status is a way to ensure that every relevant property is continuously taken into account and allow for honest ethical discussions on the thresholds of morality.

More so than any other discussion though, the astute reader might have surmised that the debate on what constitutes 'true' AI has been *the* central motivation for developing a framework for the concepts of intelligence. Opinions even today range from accepting nothing less than conscious general AI to claims that a simple calculator is intelligent in every way that counts. Intelligence certainly has become a concept of 'the gaps', where each incremental improvement in AI over time has narrowed the scope of what the sceptics of general artificial intelligence are willing to consider a sign of intelligence. Beating a human at chess was at some point thought to be an unquestionably intelligent achievement, today it is viewed as trivial computation. The accomplishments of IBM's Watson[217], beating the Jeopardy! champion in a trivia of questions posed in a natural language, likewise seem far less impressive in hindsight, when the technology needed for the task is no longer unknown and mysterious. This has been taken as both a demonstration that "intelligence" encompasses less than previously thought, and the opposite. Is language and understanding needed? Do we have to implement knowledge in the system? A coherent framework of different concepts of intelligence, showing their relationships to each other, clarifies what remains unresolved and can structure debates on the issue.

For the time being, the question might seem mostly of academic interest, but real practical implications may be just around the corner. As the world is hastily trying to prepare for what seems to be an inevitable AI-dominated near future, there is no shortage of warnings about the dangers that lie ahead. Aside from appropriate concerns about technology abuse and disproportionate power in the hands of early developers and adopters, there has been a

[217] See Chandrasekar, Raman (2014) "Elementary? Question Answering, IBM's Watson, and the Jeopardy! Challenge" In *Resonance* March 2014:222-241

resurgence in interest in the classic science fiction apocalyptic scenario of the computer gaining a will of its own and rising against its masters. In the classic “Terminator”-movies, the rather understated series of events that led to killer robots hunting down surviving humans on a scorched planet was the development of a computer program called Skynet, which at some disastrous moment came online and thus was able to unleash its powers. The popular movies were released in the 80’s and early 90’s, at a time where no one expected advanced AI anytime soon, and few worried about the dangers of everything being online.

In recent years, the concept of AI too powerful for humans to control have become a popular subject once more, but the difference is that this time many seem to take it a lot more seriously. Elon Musk famously stated, to the chagrin of many a tech optimist, that “with AI, we are summoning the demon”. [218] What is the intrinsic danger in AI? The nature of the threat of uncontrollable AI was most notably elaborated on in Nick Bostrom’s *Superintelligence* (2014). We have already touched upon the alignment problem, that there is no way of ensuring that the values that guide the artificial system is compatible with our own. An intelligence superior to our own would be unpredictable to us, in stark contrast to our efforts to constrain it, which would be easily predictable for a superintelligent agent. It would be able to appear cooperative and non-threatening while weak, and then with unseen effectiveness and ruthlessness carry out its alien volition at the precise moment the situation allows it.[219] The scenario might sound far-fetched, but the gravity of the consequences (the end of the world of humans) probably means that we should make some precautions.

One such effort would be to reflect upon how we would recognize an AI-threat. If we were of the conviction that the intellectual capabilities of an agent can be estimated by comparison with the human mind, then we would very likely fail to do so. We might search for homophenomenal intelligence, revealing no indication that the program is able to conceptualize and subjectively understand what it is doing, and thereby conclude that its intelligence is inferior to us, and no threat. The superintelligence might however be entirely unconscious, yet so adapt at achieving its goals even in the face of human interference that it is entirely beyond our powers to prevent it. Alternatively, it may in fact be conscious, but in such a different manner from us, with such a different ontology and world of experience, that we would not be able to imagine its cognitive abilities.

---

[218] McFarland, Scott (2014) «Elon Musk: «With Artificial Intelligence we are summoning the demon» In *The Washington Post* October 24, 2014, https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/?noredirect=on&utm_term=.1c5938be14b6 20.03.2019

[219] Bostrom 2014

Superintelligence is a dramatic example, demonstrating how important recognizing intelligence potentially is, but the general point is applicable in less spectacular debates on the prospects of AI research as well. We need to clearly separate the ability to do the things that the human mind does (homophenomenal intelligence) from the ability to tackle any novel problem in a fashion superior to humans (general system intelligence, system super-intelligence, or general heterophenomenal intelligence, heterophenomenal super-intelligence). Attempting to evaluate intelligence along a single, linear scale, with 'human level' or 'super-human level' as fixed points will not cut it. Similarly, it is pointless to insist on referring to a system able to predict and outsmart our every action as 'not really intelligent', because it does not 'understand' what it is doing. The philosophy of AI is in need of a richer, less ambiguous concept of intelligence then is currently the case, and in questions like the ones emphasized here, my proposed framework appears to me to be a step in the right direction.

Afterword

Exploring the subject of intelligence in-depth within the confines of a master thesis necessitates a very superficial treatment of a host of other tangential concepts and issues. The chapter structure is intended to demonstrate that different concepts of intelligence are useful and relevant in different contexts, of which there are too many to individually probe to extraneous lengths. My accounts of questions related to ethics, language and the concepts of knowledge and beliefs, for instance, did not do those debates justice in their somewhat brief appearances. I have tried to examine their most relevant aspects for our purposes, without veering too much off course. Still, many arguments were not brought to a satisfying conclusion. There are far too many questions left unresolved to mention, some because I deemed them less important in the context, some because I clearly can offer no satisfying answer. The one question I did set out to answer, the question of what intelligence is, is where I hope I have made some slight contribution. I believe that the many loose ends still left do not just demonstrate the practical limits of a master thesis, they also illustrate the many new avenues of thought and discourse that this proposed theoretical framework opens up to.

Bibliography:

Aaronson, S (2014). *Why I Am Not An Integrated Information Theorist (or, the Unconscious Expander)*. http://www.scottaaronson.com/blog/?p=1799 20.03.2019

Adamatzky, Andy, Armstrong, R., Jones, J., Gunji, Y. (2013) "On Creativity of Slime Mould" In *International Journal of General Systems* April 2013

Andrews, Kristin, Huss, B. (2014) **"**Anthropomorphism, anthropectomy, and the null hypothesis" In Biol Philos (2014) 29: 711.

Andrews, Kristin, (2016) "Animal Cognition", In *The Stanford Encyclopedia of Philosophy* (Summer 2016 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/cognition-animal/ 20.03.2019

Aydede, Murat, "The Language of Thought Hypothesis", In *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/language-thought/ 20.03.2019

Baars, B (2005) *Global workspace theory of consciousness: toward a cognitive neuroscience of human experience* Progress in Brain Research, Vol. 150 ISSN 0079-6123

Barrett, Louise (2018) "Picturing Primates and Looking at Monkeys: Why 21st Century Primatology Needs Wittgenstein" In *Philosophical Investigations* Volume 41, Issue 2 2018:161-187

Beauchamp, Tom, Childress, J. (2009) *Principles of Biomedical Ethics* New York: Oxford University Press

Bedny, Maria et al. (2011) "Language Processing in the Occipital Cortex of Congenitally Blind Adults" In *Proceedings of the National Academy of Sciences* 108, no. 11 (March 15, 2011):4429-34.

Bennett, Edward L., et al. "Rat Brain: Effects of Environmental Enrichment on Wet and Dry Weights." In *Science*, vol. 163, no. 3869, 1969, pp. 825–826. *JSTOR*, JSTOR, www.jstor.org/stable/1726503.

Bjork, Russel (2008) "Artificial Intelligence and the Soul" In *Perspectives on Science and Christian Faith* Volume 60, Number 2:95-102

Bokulich, Peter (2013) "The Physics and Metaphysics of Computations and Cognition" In Müller, Vincent (Ed) *Philosophy and Theory of Artificial Intelligence* Berlin: Springer

Bonabeau, Eric, Dorigo, M., Theraulaz, G. (1999) *Swarm Intelligence. From Natural to Artificial Systems.* New York: Oxford University Press

Bonnefon, Jean-François, Shariff, A., Rahwan, I. (2016) "The social dilemma of autonomous vehicles" In *Science* 24 Jun 2016:1573-1576

Block, Ned (1996) "How can we find the neural correlate of consciousness?" In *Trends in Neurosciences*, Nov, 1996, Vol.19 (11), p.456 (4)

Bostrom, Nick (2012) Minds & Machines 22: 71. https://doi.org/10.1007/s11023-012-9281-3

Bostrom, Nick (2014) *Superintelligence* New York: Oxford University Press

Carroll, John (1993) *Human Cognitive Abilities: A Survey of Factor Analytic Studies* Cambridge: Cambridge University Press

Carruthers, P. (2002). "The cognitive functions of language." In *Behavioral and Brain Sciences, 25*(6), 657-74; discussion 674-725.

Chalmers, David (1996) *The Conscious Mind: In Search of a Fundamental Theory.* New York: Oxford University Press

Chalmers, David J. (2006). Strong and weak emergence. In P. Davies & P. Clayton (eds.), *The Re-Emergence of Emergence: The Emergentist Hypothesis From Science to Religion.* Oxford University Press.

Chalmers, David (2010) "The Singularity: A Philosophical Analysis." In *Journal of Consciousness Studies* 17 (9-10): 7-65.

Chandrasekar, Raman (2014) "Elementary? Question Answering, IBM's Watson, and the Jeopardy! Challenge" In *Resonance* March 2014:222-241

Chapman, Siobhan, Routledge, C. (ed.) (2009) *Key Ideas in Linguistics and the Philosophy of Language* Edinburgh University Press

Chomsky, Noam (1959) «Review of *Verbal Behaviour* by B.F. Skinner» in *Language* 35: 26-58

Chrisley, Ron (2008) "Philosophical foundations of artificial consciousness" In *Artificial Intelligence in Medicine*, Volume 44, Issue 2, 2008: 119-137

Chrisley, Ron (Ed.) (2000*) Artificial intelligence: critical concepts.* London: Routledge

Christman, John, (2018) "Autonomy in Moral and Political Philosophy", In *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/autonomy-moral/ 20.03.2019

Churchland, Paul (1992). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science.* Cambridge, Massachusetts: MIT Press

Darwin, Charles (1982). *The Descent of Man*. New York: Modern Library. (Original work published 1871)

Davidson, Donald (1975) "Thought and talk." In Guttenplan S (Ed) *Mind and language*. Oxford University Press, Oxford, pp 7–24

Davis, Randall, Shrobe, H., Szolovits, P. (1993) "What Is a Knowledge Representation?" in *AI Magazine* Volume 14 Number 1 (1993)

Deary, Ian (2001) *Intelligence. A Very Short Introduction* New York: Oxford University Press

DeFusco, Christopher R. (2005) *Animal Consciousness. Grounds for the Ethical Treatment of Animals* Xlibris

Dennett, Daniel (1969) (1986) *Content and Consciousness* New York: Routledge

Dennett, Daniel (1987). *The Intentional Stance*. MIT Press.

Dennett, Daniel (1995) *Darwin's Dangerous Idea* New York: Simon and Schuster

Dennett, Daniel (2017) *From Bacteria to Bach. The Evolution of Minds* London: Penguin Books

Domingos, Pedro (2015) *The Master Algorithm* London: Penguin Random House

Drayson, Zoe (2014), "The Personal/Subpersonal Distinction." In *Philosophy Compass*, 9: 338-346

Drigas, Athanasios & Pappas, Marios. (2017). "The Consciousness-Intelligence-Knowledge Pyramid: An 8x8 Layer Model." In *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*. VOL 5. 14-25. 10.3991/ijes.v5i3.7680

Dummett, M. (1994). *Origins of analytical philosophy*. Cambridge: Harvard University Press.

Earl, Brian (2014). "The biological function of consciousness." In *Frontiers in psychology, 5*, 697.

Edelman, David, Seth, A. (2009) "Animal consciousness: a synthetic approach" In *Trends in Neurosciences*, Vol. 32, Issue 9, 2009: 476-484

Edsinger, Eric, Dölen, G. (2018) "A Conserved Role for Serotonergic Neurotransmission in Mediating Social Behavior in Octopus" In *Current Biology* DOI:https://doi.org/10.1016/j.cub.2018.07.061

Eysenc, Hans (1988) "Concept of "Intelligence" Useful or Useless*?*" In *Intelligence* 12: 1-16

Fantl, Jeremy, (2017) "Knowledge How", In *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/knowledge-how/ 20.03.2019

Fedorenko, Evalina, Varley, R. (2016) "Language and thought are not the same thing: evidence from neuroimaging and neurological patients» in *Annals of the New York Academy of Sciences*. Issue: The Year in Cognitive Neuroscience 2016: 132-153

Fields, Chris (2013) "How humans solve the frame problem" In *Journal of Experimental & Theoretical Artificial Intelligence*, 25:4, 441-456

Finger, Stanley (1995) "Descartes and the pineal gland in animals: A frequent misinterpretation" In *Journal of the History of the Neurosciences*, 4:3-4, 166-182

Flynn, James R. (2009) *What is Intelligence?* Cambridge: Cambridge University Press

Funke, Joachim. (2012). "Complex Problem Solving." In Seel, N. (Ed.) *Open Learning Environments* Springer 2012:682-685

Gardner, Howard (1983) *Frames of Mind* New York: Basic Books

Garson, James (2015) «Connectionism» in *The Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/connectionism/ 20.03.2019

Gettier, Edmund (1963) "Is Justified True Belief Knowledge?" In *Analysis*, Vol. 23, No. 6, 1963: *121-123*

Gless, Sabine, Silverman, E., Weigend, T. (2016) "If Robots cause harm, who is to blame? Self-driving Cars and Criminal Liability" In *New Criminal Law Review: In International and Interdisciplinary Journal*, Vol. 19 No. 3, Summer 2016:412-436

Godfrey-Smith, Peter (2016) *Other Minds. The Octopus and the Evolution of Intelligent Life.* London: William Collins

Gottfredson, Linda S. (1997). "Mainstream Science on Intelligence (editorial)" In *Intelligence*. 24: 13–23

Gould, Stephen (1985). *The Flamingo's Smile*. New York: W. W. Norton.

Gould, Stephen Jay (1996) *The Mismeasure of Man* New York: Norton & Company

Gregory, Richard (1970). *The Intelligent Eye.* Mcgraw-Hill

Hansen, Ronald, Skavenski, A. (1985) «Accuracy of spatial localizations near the time of saccadic eye movements." In *Vision Res.* 1985; 25(8):1077-82

Hansson, Sven Ove (2017) "Logic of Belief Revision" In *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/entries/logic-belief-revision/ 20.03.2019

Harré, R. (2010). "Gilbert Ryle: The concept of mind." In *Central Works of Philosophy Volume 4: The Twentieth Century: Moore to Popper* (pp. 214-238). Acumen Publishing Limited.

Hauser, Mark, Chomsky, N., Fitch, T. (2002) "The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?" In *Science* 22 Nov 2002:1569-1579

Hebb, Donald (1949). *The Organization of Behavior*. New York: Wiley & Sons.

Heinrich, Bernd (1995) "An Experimental Investigation of Insight in Common Ravens (Corvus corax)" in *The Auk*, Vol. 112, No. 4 (Oct. 1995):994-1003

Hill, Robin (2016) "What an algorithm is" In *Philosophy and Technology* Vol 29,

Issue 1:35 59

Hintikka, Jaakko (1962) *Knowledge and Belief: An Introduction to the Logic of the Two Notions.* New York: Cornell University Press

Hochner, Binyamin (2012) «An Embodied View of Octopus Neurobiology" in *Current Biology* Vol 22, Issue 30, October 2012:887-892

Hughes, T.J. & Miller, J.T.M. (2014) "Lexicalisation and the Origin of the Human Mind" In *Biosemiotics* 2014 7: 11.

Hume, David (2003) *A Treatise on Human Nature* New York: Dover (Original published in 1888)

Johnson, Steven (2001) *Emergence* London: Penguin Books

Joyce, James, (2003) "Bayes' Theorem" In *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2016/entries/bayes-theorem/ 20.03.2019

Kant, Immanuel (1790/1952). *Critique of judgment.* Oxford: Oxford University Press.

Kant, Immanuel (2012) *Groundwork on the Metaphysics of Morals* Revised Edition. Cambridge University Press

Keating, D.P. (1984). The emperor's new clothes: The "new look" in intelligence research. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 1-45). London: Erlbaum.

Khalfa, Jean (ed.) (1994) *What is Intelligence?* New York: Cambridge University Press

Kingstone, Alan, Smilek, D. and Eastwood, J. D. (2008) “Cognitive Ethology: A new approach for studying human cognition” In British *Journal of Psychology*, 99: 317-340

Kim, Jaegwon (2006) “Emergence: Core Ideas and issues.” In *Synthese* 2006:151: 547.

Koch, Christof, Tsuchiya, N. (2012) “Attention and consciousness: related yet different.” In *Trends in Cognitive Sciences*, Volume 16, Issue 2, 2012: 103-105

Kosterec, Milos (2016) “Methods of Conceptual Analysis” In *Filozofia* 71, 2016, No. 3: 220-230

Kulstad, Mark and Carlin, Laurence, (2013) "Leibniz's Philosophy of Mind", *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/win2013/entries/leibniz-mind/ 20.03.2019

Kurzweil, Ray (2012) How to Create a Mind London: Duckworth Overlook

Ladyman, James, (2016) "Structural Realism" In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), Edward N. Zalta (Ed.) https://plato.stanford.edu/entries/structural-realism/ 20.03.2019

Lanz, Peter (2000) “The Concept of Intelligence in Psychology and Philosophy.*”* In Cruse H., Dean J., Ritter H. (red) *Prerational Intelligence: Adaptive Behavior and Intelligent Systems Without Symbols and Logic*, Vol 1, Vol 2 Prerational Intelligence: Interdisciplinary Perspectives on the Behavior of Natural and Artificial Systems, Vol 3. Studies in Cognitive Systems, vol 26. Springer, Dordrecht

Laurie, v. M., Pallas, S. L., & Sur, M. (2000). “Visual behaviour mediated by retinal projections directed to the auditory pathway.” In *Nature*, *404*(6780), 871-6.

Legg, S. & Hutter, M. (2007) «Universal Intelligence: A definition of Machine Intelligence» In *Minds & Machines* 2007 Vol 17: 391-444.

Levesque, Hector (2017) *Common Sense, the Turing Test, and the Quest for Real AI* Cambridge: MIT Press

Lindemann-Biolsi, Kristy, Reichmuth, C. (2013) “Cross-modal Transitivity in a California sea lion” In *Anim Cogn* DOI 10.1007/s10071-013-0721-0

Ma, Christine and Schapira, M. (2017) *Richard J. Herrnstein & Charles Murray's “The Bell Curve” Intelligence and class structure in American life* London: Routledge

Macphail, E. M. (1985). “Vertebrate intelligence: The null hypothesis.” In L. Weiskrantz (Ed.), *Animal intelligence* (pp. 37–50). Oxford: Clarendon.

Mahon, Bradford, Caramazza. A. (2009) “Concepts and categories: a cognitive neuropsychological perspective” In *Annual review of psychology* vol. 60 (2009): 27-51.

Marder, Michael (2012) “Plant intentionality and the phenomenological framework of plant intelligence.” In *Plant Signaling & Behavior*, 7:11, 1365-1372

McCarthy, John, Hayes, P. J. (1969). "Some philosophical problems from the standpoint of artificial intelligence» In *Machine Intelligence* 4*: 463–502.*

McClelland, J., Rumelhart, D., and the PDP research group (1986) *Parallel distributed processing: Explorations in the microstructure of cognition* MIT Press

McFarland, Scott (2014) «Elon Musk: «With Artificial Intelligence we are summoning the demon» In *The Washington Post* October 24, 2014, https://www.washingtonpost.com/news/innovations/wp/2014/10/24/elon-musk-with-artificial-intelligence-we-are-summoning-the-demon/?noredirect=on&utm_term=.1c5938be14b6 20.03.2019

McGrew W (2004) *The cultured chimpanzee: reflections on cultural primatology.* Cambridge University Press, Cambridge

Miller, George (1962). *Psychology: The Science of Mental Life*. New York, NY: Harper and Row

Minsky, Marvin (1975) «A Framework for Representing Knowledge» in *The Psychology of Computer Vision*, P.Winston (Ed.) McGraw-Hill, 1975.

Minsky, Marvin (1986) *The Society of Mind.* New York: Simon & Schuster.

Mukherjee, Siddharta (2016) *The Gene. An Intimate History*. New York: Scribner.

Murray, Charles and Herrnstein, R. (1994) *The Bell Curve. Intelligence and Class Structure in American Life* New York: Free Press

Nagel, T (1974) "What is it like to be a bat?" In *Philosophical Review*, 83: 435–456.

Naotsugu, Tsuchiya, Koch, C. (2008) "Attention and consciousness." In *Scholarpedia*, 3(5):4173.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Urbina, S. (1996). "Intelligence: Knowns and unknowns." In *American Psychologist, 51*(2), 77-101.

Newell, A. and Simon, H. (1976) "Computer science as empirical inquiry: Symbols and search." In *Communications of the ACM* 19: 113-126.

Pearl, Judea (2010) "An Introduction to Causal Inference," In *The International Journal of Biostatistics*: Vol. 6: Iss. 2, Article 7

Penrose, Roger (1994) "Mathematical Intelligence" in Khalfa, J. (Ed) *What is Intelligence?* New York: Cambridge University Press 1994:107-136

Penrose, Roger (2005) *Shadows of the Mind. A search for the missing science of consciousness.* London: Vintage

Pinker, Steven (1994) *The Language Instinct: The New Science of Language and Mind.* Harmondsworth: Penguin

Plummer, Libby (2017) "This is how Netflix's top-secret recommendation system works" *WIRED Magazine* Tuesday 22 August 2017 https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like 20.03.2019

Povinelli, D. J. and J. M. Bering (2002). "The Mentality of Apes Revisited." In *Current Directions in Psychological Science* 11(4): 115-119.

Premack, David, Woodruff, G. (1978). "Does the chimpanzee have a theory of mind?" In *Behavioral and Brain Sciences, Special Issue: Cognition and Consciousness in Nonhuman Species.* 1 (4): 515- 526.

Putnam, Hilary (1963). "Brains and behavior." In Ronald J. Butler (ed.), *Analytical Philosophy: Second Series*. Blackwell.

Quanta Magazine May 15, 2018 https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/ 20.03.2019

Quine, Willard (1951). "Two Dogmas of Empiricism". In *The Philosophical Review.* 60 (1): 20–43.

Rescher, Nicholas (2001) *Philosophical Reasoning. A study in the methodology of philosophizing* Oxford: Blackwell

Ristau, Carolyn (2013) “Cognitive ethology” In *WIREs Cogn Sci*, 4: 493-509.

Rosch, Eleanor, Thompson, E., Varela, F. (1991*) The embodied mind: Cognitive science and human experience* MIT Press.

Ruiz-Mirazo, K. & Moreno, A. Synthese (2012) 185: 21. https://doi.org/10.1007/s11229-011-9874-z

Samhita, L., & Gross, H. J. (2013). “The “Clever Hans Phenomenon” revisited.” In *Communicative & Integrative Biology*, *6*(6), e27122. http://doi.org/10.4161/cib.27122

Searle, John (1982) “What is an intentional state?” In H.L: Dreyfus with H. Hall (ed.) *Husserl, Intentionality, and Cognitive Science.* Cambridge, MA and London: MIT Press

Searle, John (1990) “Is the Brain a Digital Computer?” In *Proceedings and Addresses of the American Philosophical Association*, Vol. 64, No. 3 (Nov. 1990):21-37

Schroeder, Timothy (2004) *Three Faces of Desire* New York: Oxford University Press

Schroeder, Timothy (2017) "Desire", In *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/desire/ 20.03.2019

Schank, Roger, Birnbaum, L. «Enhancing Intelligence» in Flynn, James R. (Ed) (2009) *What is Intelligence?* Cambridge: Cambridge University Press

Schull, Jonathan (1990) «Are Species Intelligent?» In *Behavioural and Brain Sciences* (1990) 13, 63-108

Siewert, Charles (2017) "Consciousness and Intentionality", In *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.) https://plato.stanford.edu/entries/consciousness-intentionality/ 20.03.2019

Singer, Peter (1975) *Animal Liberation: A New Ethics for our Treatment of Animals*, New York: Random House

Spearman, Charles (1927) *The Abilities of Man, Their Nature and Measurement* New York: Macmillan

Stanley, Jason (2011). “Knowing (How)”. In *Noûs, 45*(2), 207-238.

Sun, Ron (2015) “Artificial Intelligence: Connectionist and Symbolic Approaches” In *International Encyclopedia of the Social & Behavioral Sciences* 2nd Edition 2015

Tegmark, Max (2017) *Life 3.0* London: Penguin Random House

Terrace, S., Petitto L., Sanders, R., Bever, T. (1979) “Can an Ape Create a Sentence?” In *Science* Vol 206, Number 4421, 23 November 1979

Tononi, G (2012) *Integrated information theory of consciousness: an updated account* Archives Italiennes de Biologie, 150: 290-326, Tononi, G (2014) *Why Scott should stare at a blank wall and reconsider (or, the conscious grid)*. Accessed September 9. 2018. http://www.scottaaronson.com/blog/?p=1823 20.03.2019

Trewavas, Anthony (2003) "Aspects of plant intelligence." In *Annals of Botany* 2003; 92:1 - 20

Turing, Alan (1950) "Computing Machinery and Intelligence" In *Mind* 59: 433–6

Valcourt, James (2017) *Systematic. How Systems Biology is Transforming Modern Science.* New York: Bloomsbury Publishing Plc

Velmans, Max (1991). "Is human information processing conscious?" In *Behav. Brain Sci.* 14, 651–726

Vikhar, Pradnya (2016). "Evolutionary algorithms: A critical review and its future prospects". In *Proceedings of the 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC*). Jalgaon, 2016, pp. 261-265

Wang, Pei (2008) *What do you mean by «AI»?* Proceedings of the 2008 conference on Artificial General Intelligence 2008: Proceedings of the First AGI Conference

Ward, Lawrence (2011) "The thalamic dynamic core theory of conscious experience" In *Consciousness and Cognition* Vol. 20, Issue 2, 2011: 464-486

Warren, Mary (1997) "On the Moral and Legal Status of Abortion" In *Ethics in Practice* La Follette, H. (Ed) Blackwell, Oxford

Wedeen, Van J. et al (2012) "The Geometric Structure of the Brain Fiber Pathways" In *Science* 335, no. 6076 (March 30, 2012)

Wells, Gordon (1994) «The Complementary Contributions of Halliday and Vygotsky to a "Language-based Theory of Learning" In *Linguistics and Education* 6, 1994:41-90

Wolf, Michael "Philosophy of Language" In *Internet Encyclopedia of Philosophy* https://www.iep.utm.edu/lang-phi/ 20.03.2019

Wittgenstein, Ludwig (1999) (1922) *Tractatus Logico-Philosophicus* Trondheim: Gyldendal

Zhuang et al. (2017)"Challenges and opportunities: from big data to knowledge in AI 2.0" In *Front Inform Technol Electron Eng* 2017 18(1):3-14