Intelligence and simulation of minds

First Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@chalmers

Second Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@chalmers

Third Author
Affiliation / Address line 1
Affiliation / Address line 2
Affiliation / Address line 3
email@chalmers

Abstract

In this essay, we discuss the question of whether machines can have minds. We start off by introducing some of the arguments and ideas that have been proposed on the subject during the last 70 years. We then try to define what a mind is, showing some examples from the animal kingdom. We move on to describe some different methods of testing if something has a mind and discuss what the tests actually show, before considering a few examples of how simple minds can arise. We finally conclude that although human brains and modern computers are fundamentally different, the possibility of a machine mind seems plausible. However, a potential machine mind will most likely be very different from a human mind and thus make it impossible for machines and humans to fully understand each other.

1 Introduction

It all started with the question "Can machines think?". We as human civilisation have been creating tools to solve problems since the beginning of our time. As we grew better at creating tools and collecting them together, we were able to build more and more complex machines upon each other. Then came the computer, a machine first designed as a computational tool, which we later gave the ability to improve itself. Inevitably, we started to feel the uncertainty of what this can lead to.

Alan Turing avoided answering this question by introducing an imitation game experiment (Turing, 1950). Can machines behave indistinguishably as human from another human observer's perspective? This experiment itself raised another discussion among computer scientists and philosophers. As researchers in various areas improved from the time of Turing, instead of asking if machines think, we would like to explore the question of whether machines can have minds.

1.1 Strong versus weak AI

Before we dive into the discussion regarding machines with their minds, a clear framework should be illustrated in order to achieve the most fruitful discussion. We often hear "strong AI" and "weak AI" in the domain of artificial intelligence research, but these phrases can actually have two different meanings. Since the terms overlap, they are sometimes accidentally confused. We would like to clear the distinctions here.

Strong and weak AI used in computational theory of mind

John Searle is often regarded as the first person to use the term "strong AI" in this definition in his "Minds, brains, and programs" (Searle, 1980). According to John Searle, "strong AI" claims that programs (or programmed computers) that correctly maps the right input to the right output, should be regarded as possessing a mind just like humans possess minds. On the other hand, "weak AI" (discussed more in Section 1.2) claims this mapping is merely an imitation of mind and so cannot be accounted as a real mind. Searle's famous Chinese room thought experiment, which is explained in more detail in Section 2.1, serves as a refutation to strong AI claims. In this thought experiment, Searle introduced the concept of "intentionality" which is not possible to model or include in any program. A man in this Chinese room can perform a task perfectly by following the given instructions and rules while not knowing any Chinese. No intentionality at all is involved in the tasks carried out.

Strong and weak AI used to describe the ability to solve general problems within a single agent

Sometimes also referred to as full AI, strong AI in this branch can be defined as an agent which can demonstrate the ability to perform any intellectual task that humans are capable of (Kurzweil, 2005). That is, an agent can solve tasks that are not predetermined on a level matching a human. If there exists an artificial intelligence agent which can pass the Turing test, one can classify it as strong AI or full AI because it demonstrates the ability to perform tasks in terms of natural language dialogue equivalent (or indistinguishable from a third person's observation) to human level. However, the general consensus is that general intelligence must accomplish the following:

1. reasoning

2. learning

3. planning

4. natural language communication

5. Knowledge representation (KR)

Although discussion in either one of these branch often leads to some common arguments from another branch, it is better to clarify at beginning.

1.2   What is a mind?

Does simulated minds count as real mind?

As we mentioned earlier, Searle refuted the claim that simulation of mind, itself should be counted as a real mind. Till today this argument still remains a topic of debate. So why do people waffle about whether a simulation of mind can be attributed as a real mind or not? Another famous question of philosophy, the mind-body problem, could be linked to this simulation problem. The 17th-century philosopher René Descartes is commonly acknowledged to be the first person who formalised this problem (Descartes, 1641): "How could the immaterial mind (or soul) influence the material body?". Descartes also discussed where the boundary between the material world and the immaterial world is, if one regard information flow as input and mind as processor of information. Like most philosophical problems, the answer to the mind-body problem seems to depend largely on one's personal belief.

Intelligence in animals

In order to better understand the concept of minds, we take a look at intelligence in some known organisms. Many animals are known to exhibit intelligent behaviour in both individuals and communities. For example, many species of fish, bird and some mammals are able to periodically travel great distances to the location of their birth once a year. Other species such as dogs, dolphins and ants are well known for their social behaviour.

Mackintosh noted that crows had the ability to carry information from experiment to experiment while pigeons were notoriously bad at this, seemingly starting over from scratch (Mackintosh, 1988). He refuted the then common idea that intelligence in animals could be ordered using the evolutionary tree, i.e. the "phylogenetic hierarchy". Crows are also known to craft and use tools to gather food and defend against predators (Emery and Clayton, 2004). In an experiment, a crow could reason about which tool to use as well as making tools from new materials (metal wire) when no twigs were available. Furthermore, crows use strategy in both caching their own food for future use and pilfering food caches of other birds. Emery and Clayton also compared the brains of crows and chimpanzees. The ratio of brain size and the rest of the body is about the same and both species have an enlarged brain center for cognitive thought: neocortex in apes and nidopallium in corvids (crows). However the structure of the nidopallium is very different from the neocortex of the apes. We thus draw the conclusion that complex cognition can be brought about by different (brain) architectures resulting in similar levels of intelligence.

Which animals have minds?

Now that we have discussed some intelligent behaviour of various animals, we can try to distinguish which of them have minds and which do not. We can argue that humans have minds since we are capable of conscious thought and abstract reasoning. On the other hand bacteria may not possess minds as far as we can observe since they seem not be able to learn and remember. However, one can argue that bacteria learn as communities by mutation and that individuals reason about which action to take from their chemical perception of their surroundings. So even for a simple organism such as bacteria, to decide if they have minds is not trivial. Here we assume bacteria are mindless since they seem to depend more on random mutations than conscious actions.

It has been noted in the past that ordering animals' intelligence in a single dimension lacks proper foundation (Mackintosh, 1988) so we will not make this assumption here. However many dimensions are needed to present the intelligence of animals, there must be some classifying function between humans and bacteria that categorizes which have minds and which do not. The problem is that there seems to be little consensus in how to draw such a distinction.

One can argue that humans are unique in being able to evolve as a species in a much faster rate than evolution. Cultural changes, advances in medicine and technology can all be considered a conscious evolution of the species. While stray dogs may not be able to construct metro stations of their own, they can learn how to use them to navigate their neighbourhoods, not so differently from their fellow human passengers (McGrane, 2013). Crows have been recorded to drop nuts on pedestrian crossings and eating the crushed content when the traffic lights switches in favour of the pedestrians (Attenborough, 2007). These innovative and adapting behaviours can be considered proof that dogs and crows have minds.

A more general view would be that some set of aspects of intelligence such as spatial intelligence, long-term memory, learning abstract concepts, language and social behaviour would be sufficient or necessary for possessing a mind. Exactly which concepts are required and to what extent is a matter of debate.

2 Discussion

2.1 How to test if an agent possesses a mind?

Previously we noted that depending on the definition of mind, several animals could be thought of as having a mind or not. So instead of attempting to resolve this ambiguity we will attempt to answer the question of whether a machine can possess a human mind, that is to think like a human.

Architecture of the human mind

Goertzel collected a lot of research from psychology and neurology to form a conceptual model of the human mind from the different theories (Goertzel et al., 2012). The resulting diagram, illustrating high level cognitive-architecture, can be found in figure 1. Each box within figure can be further expanded into detail diagrams, in depth discussion can be found in Goertzel's referred work. Connections between boxes indicate dependencies during performing intellectual tasks. For example, to greet an old from you knew long ago, you would need the connection between long-term memory, self/social awareness, and language ability to interact with action subsystem. Additionally, emotional reaction could help you expressing yourself.

Testing AGI (Artificial General Intelligence)

Goertzel notes that there are several tests for assessing whether an agent has achieved a human level of AGI (Goertzel et al., 2012). The most classical one would be the Turing test which statistically tests if the robot could be indistinguishable from a human (Turing, 1950). A human judge would ask several, possibly connected, questions over a text chat to both a robot and a human participant. The judge would then guess which of the two participants really is human.

Another, more difficult, test would be to let a robot enroll in a university (Goertzel et al., 2012). The robot would succeed if it was able to obtain a degree given the same conditions of passing as the human students.
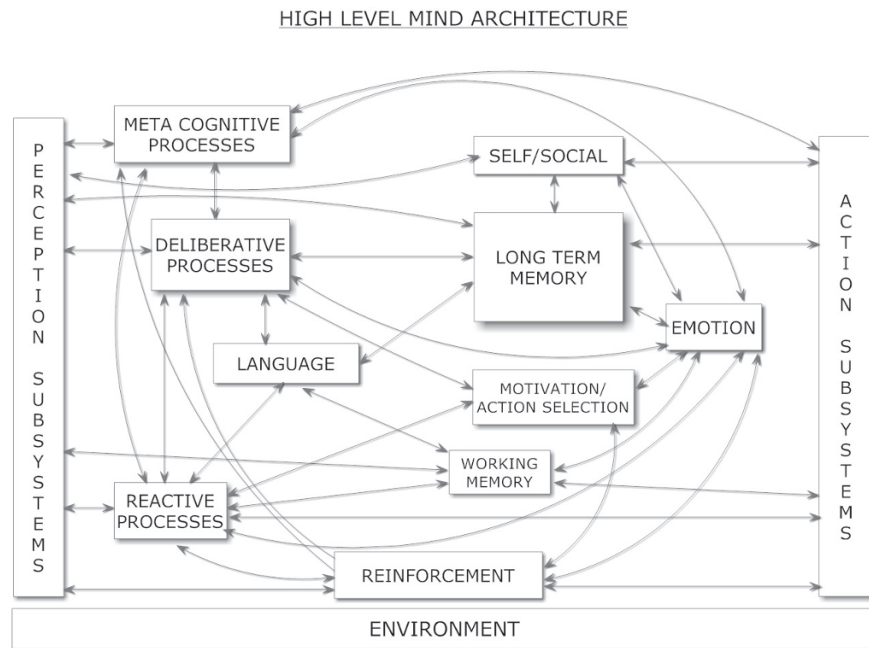
Figure 1: Model architecture of human mind as proposed by Goertzel (Goertzel et al., 2012).

Goertzel also listed some tests that could be assumed to require human intellect but could also be solved by a "narrow AI", i.e. specific to the task, (Goertzel et al., 2012). One of these tests: "to be able to play an arbitrary video game based on experience only" was recently achieved by Google on the Atari game platform (McMillan, 2015).

It is very difficult, as Goertzel also mentioned, to measure when research is halfway to the goal since all components in the architecture of the mind are so interdependent: even if only one or two parts are missing, the agent would probably not behave even close to a mind (Goertzel et al., 2012).

Understanding minds

The tests above only test the behaviour of an agent, that is AGI. When it comes to minds, consciousness and understanding others, Nagel made an interesting point (Nagel, 1974). He argues that a human can never understand what it is like to be a bat: We can try to imagine sleeping upside down, navigating with sonar and "catching insects with our mouths" but that would only result in what it would be like for oneself to be a bat, which is fundamentally different from what it would be like for a bat to be a bat.

If we accept Nagel's argument, one can draw the parallel to computers. Thus a bat expert might be able to predict and even imitate a bat while not being able to understand it internally. Likewise, a computer could possibly imitate and predict human behaviour whilst not understanding humans emotionally. For example, things like laughter, grief and art could possibly be predicted or emulated but not internally understood.

Chinese Room Experiment

In any case, even if a machine would successfully understand all abstract parts of being human, it could be very hard to tell from just a smart imitating machine. Searle illustrated this in his famous thought experiment: the Chinese room (Searle, 1980). If an agent that speaks English natively but doesn't understand any Chinese is sealed away in a room together with a list of very thorough instructions of how to manipulate some input characters (presumably) written in Chinese, he might be able to produce some Chinese text but it doesn't mean he understands what he is doing. However, the people on the outside might not be able to distinguish that agent from a native Chinese speaker, assuming the instructions are sufficiently detailed and intelligently designed. Likewise, if a computer successfully could understand

humans, we couldn't really tell the difference of that from a list of intelligent instructions executed by a thorough but non-understanding agent.

2.2 How does a simple mind arise?

What might be considered as *simple minds* can arise in, among others, animals and robotics. Of course, this depends on which definition of 'mind' is used, a question which definitely is not trivial. On one hand, these 'minds' display complex and seemingly intelligent behaviours. But on the other hand it is sometimes also possible to describe and explain these complex behaviours using rather simple explanations, like the food-finding capabilities of *E. Coli* bacteria or the navigation methods used by the desert ant *Cataglyphis Fortis* (Wahde, 2015). This falls into what is called *complex systems*: complex behaviours arising from simple components. Examples of this, some of which are described below, are social insects and even single-celled organisms.

Slime mould

Slime mould are unicellular amoeba which, when food is in short supply, can form into networks that coordinate and display behaviours similar to animals with simple brains (Bonner, 2009). This is extraordinary, since the individual cells are not known to have the ability to think. Yet somehow a large collection of these individuals are able to display a degree of intelligence when put together, something which can be seen in the example with slime mould. An experiment in which a mass of slime mould was placed in an artificial environment (representing the area around Tokyo) showed that the slime mould was able to find an efficient way of connecting the major cities in the Greater Tokyo Area (Tero et al., 2010). The generated map was strikingly similar to the rail network that already is in use in the Tokyo area.

Bee drones

The societies of social insects are known to display complex behaviours, for example when searching for food or new homes. Researchers are now trying to create robots which are inspired by social insects. At the University of Harvard, a research team is developing RoboBees, microrobots with a size and behaviour similar to those of real bees (Harvard University, 2015). The hope is that these robots will be able to work in areas such as artificial pollination, search and rescue operations, weather monitoring and so on. The RoboBees, though rather simple robots on their own, will be able to collaborate as a "colony" in order to efficiently solve tasks.

At the universities of Sheffield and Sussex, researchers are cooperating in the Green Brain Project with the aim of building a model that mimics the brain of a bee (University of Sheffield and University of Sussex, 2015). The goal is to achieve a better understanding in how honeybee brains work and at the same time find improvements for the methods used in controlling autonomous flying vehicles. So far, the researchers have been able to simulate parts of a bee's brain and upload it to an UAV (quadcopter), allowing it to fly and use simple navigation to avoid running into obstacles.

3 Conclusion

3.1 Does it matter?

Before we start, you have to know they don't care!

Simulation of intelligence or real intelligence is in many sense not important issue for AI researchers working on delivering programs that finish designated tasks. As much as you wouldn't care if a screwdriver is actually a simulated solution or a real solution as long as it fixes your IKEA beds. AI research is therefore needed as it produces more tools for solving our problems in life, and some of these problems are practically unsolvable without these tools. We now pose a nice question to ponder: a screwdriver is a handy tool but in the wrong hand could be a weapon to harm others. Did the inventor of the screwdriver think of this perspective? Wouldn't it be more safe to not release this potentially dangerous version of screwdriver and instead keep refining it until a more harmless version is invented?

Another question worthy considering would be: "If a computer, provided with the correct software, can be considered as a creation with a mind (and in some degree, consciousness), would it be a crime of murder if the plug is pulled out?"

$(\textbf{simple mind})^n = \textbf{complex mind?}$

Brains are too complex to recreate or mimic at the time of writing. Yet, if it is possible to show, our direction of work is on the right track. The examples shown in Section 2.2 can maybe offer us some confidence. One can either build very simple units (replicating slime mould or neurons) or assemble meta-complex brains like bee's. Given previous studies on composition, or to say components, of human mind, it could be possible even with smaller scale of minds, like those of bees, can lead us to some association of certain functions of mind to its corresponding circuitry. Or by focusing on single simple cell replications, like slime mould, and grouping large quantities of these simple agents for performing complex behaviour in order to help us understand how to transform basic communication of individual agents into higher level patterns which in turn trigger complex behaviours. Although there exists technological limitations, for instance a maximum computation power per single chip before causing meltdown on chip itself, we can still combine many known principles from the development of modern computer systems in attempts to explore more possibilities.

Our limits need not apply

If we regard programs as form of intelligence and mindful thing, machine intelligence shouldn't be confined (or defined) by limitations of humans. Maybe one can argue computers will never be able to understand true semantics of our natural languages, but this would be a very narrow view. Our natural languages have resulted from cultural evolution separated by geographical boundaries, and if we review history of mankind, the flaws in natural language causing misunderstanding or semantic mismatch have cost us dearly. Yet all civilisation throughout the world seems to reach one unified syntax when it comes to mathematics. The connection from its syntax to semantics are so absolute that we still use zero to represent the very same concept as when zero was invented. Computers and programs are born in the world of math, binary or not, there exists no doubt that computers are much capable than humans in field of numbers. Needless to say the speed of computation carry out by computers, easily reaching to petaflop at the time of writing, can never be achieved by our biological limits.

It is truth saying human brains are most complex masterpiece of nature, evolution made our brain excellent to handle any happening situation around us. We can see the world and filter out most critical information and process under amazing low power consumption. While machines still have huge difficulty identifying heels from sneakers, one has to consider the fundamental differences between human and machine. Analog information is an easy task for us because we've survived in such analog world for thousands of years, but you can't perceive current temperature up to three decimal precision.

Embracing such perspective, and one will find huge potential in development of machines. As mentioned, all our cultures are geologically separated, so are all the individuals of mankind. Our communication forms are still very limited, even with aid from machines. Given the power of almost instantaneous connection, the accumulated knowledge of machines must eventually exceed ours.

3.2 Final words

As we have seen, it can be hard to define what a mind really is and how to test it. As Goertzel noted, detecting partial progress is even harder. On the bright side, this may mean that we are closer to achieving a conscious machine than we think, but it can also discourage people from trying. From the top down perspective, we can look at components of the human mind and discuss how to emulate them; from the bottom up perspective, we can simulate simple brains. When the two approaches go deep enough to meet, a working mind may not be far even if it is not at a human level.

However, if a successful simulation of mind really is a mind in itself and therefore is subject to ethics, is a philosophical question. Furthermore, if a robot mind is created it will not necessarily be human or close to human, but considering human history, that may turn out to not be such a bad thing.

References

David Attenborough. 2007. Wild crows inhabiting the city use it to their advantage. Youtube: `https://www.youtube.com/watch?v=BGPGknpq3e0`, February. [Online; accessed 2015-05-08].

John Tyler Bonner. 2009. *The social amoebae: the biology of cellular slime molds*. Princeton University Press.

René Descartes. 1641. *Meditationes de prima philosophia*.

Nathan J. Emery and Nicola S. Clayton. 2004. The mentality of crows: Convergent evolution of intelligence in corvids and apes. *Science*, 306:1903–1907.

Ben Goertzel, Matt Ikle, and Jared Wigmore. 2012. The architecture of human-like general intelligence. In *Theoretical Foundations of Artificial General Intelligence, Atlantis Thinking Machines*, volume 4, pages 123–144. Atlantis press.

Harvard University. 2015. Robobees. `http://robobees.seas.harvard.edu/home`. [Online; accessed 2015-05-07].

Ray Kurzweil. 2005. *The singularity is near: When humans transcend biology*. Penguin.

N. J. Mackintosh. 1988. Approaches to the study of animal intelligence. *British Journal of Psychology*, 79:509–522.

Sally McGrane. 2013. Moscow's metro dogs. The New Yorker: `http://www.newyorker.com/culture/culture-desk/moscows-metro-dogs`, July. [Online; accessed 2015-05-07].

Robert McMillan. 2015. Googles ai is now smart enough to play atari like the pros. Wired: `http://www.wired.com/2015/02/google-ai-plays-atari-like-pros/`, February. [Online; accessed 2015-05-07].

Thomas Nagel. 1974. What is it like to be a bat? *The philosophical review*, pages 435–450.

John R Searle. 1980. Minds, brains, and programs. *Behavioral and brain sciences*, 3(03):417–424.

Atsushi Tero, Seiji Takagi, Tetsu Saigusa, Kentaro Ito, Dan P Bebber, Mark D Fricker, Kenji Yumiki, Ryo Kobayashi, and Toshiyuki Nakagaki. 2010. Rules for biologically inspired adaptive network design. *Science*, 327(5964):439–442.

Alan M Turing. 1950. Computing machinery and intelligence. *Mind*, pages 433–460.

University of Sheffield and University of Sussex. 2015. The green brain project. `http://greenbrain.group.shef.ac.uk/`. [Online; accessed 2015-05-07].

Mattias Wahde. 2015. *Introduction to Autonomous Robots*.