Universiteit Utrecht

Bachelor Thesis

Fake it 'till you make it:

How advancing AI imitation learning could contribute to the development of Artificial General Intelligence.

B Artificial Intelligence: Bachelor Thesis (7.5 ECTS)

Author: M.E.H. Gehlmann (4293649)

Supervisor and first examiner: J.M. Broersen

Second examiner: B.H. van der Gaast

Final version: 08-07-2019

Abstract

Research into Artificial Intelligence has sought to create a human-like machine, referred to as Artificial General Intelligence (AGI). Reinforcement Learning (RL), Artificial Neural Networks, the more recent Deep and Convolutional Neural Networks and, for this thesis most importantly, Imitation Learning (IL) have all been developed for this purpose. But until now, recent AGI-research, e.g. Google DeepMind, underlines Reinforcement Learning as the key to developing AGI, as this is how biological systems learn as well. This thesis, however, argues that this current paradigm unjustifiably neglects IL for its inferior specific task performance. This literature study illustrates the latest advancements in RL, then compares these systems with pure IL and mixed approaches. Next is a discussion on the essentiality of IL in our only functional example of AGI: the human mind. Finally, all chapters are brought together in a comparison between artificial and human IL, and it ends on a brief summary of how IL currently contributes to RL-systems in the field of AI.

Contents

Abstract ... 2

Introduction ... 4

Theoretical Background ... 4

Thesis Structure ... 5

Chapter One: Defining Concepts ... 6

Artificial General Intelligence ... 6

The Artificial Neural Network ... 6

Reinforcement Learning and Google DeepMind ... 7

Google DeepMind – Deep Q-Learning ... 8

Major Problems with RL-ANN Systems for AGI-research ... 8

Chapter Two: Imitation Learning in Recent Research ... 10

Why Imitation Learning? ... 10

Methods of Imitation Learning ... 10

One-Shot Imitation Learning ... 10

Zero-Shot Imitation Learning ... 11

Reinforcement-Imitation Learning Systems ... 11

Combining the Two: DDPGfD ... 12

Combining the Two: DQfD ... 12

Chapter Three: Imitation Learning and Artificial General Intelligence ... 13

Imitation Learning in the Human Mind ... 13

IL: Essential for human learning ... 14

IL: Integral to the human mind ... 14

Contributions to current AI and future AGI research ... 15

Summary and Conclusion ... 16

References ... 18

Introduction

Since the beginning of research into Artificial Intelligence (AI), the dream has been to create an artificially intelligent 'being' that could roughly match human thinking, reasoning, problem solving, speaking and so on; to create a machine that we would be able to converse and cooperate with: a robot that one could see as an equal. Much has been achieved: some AI-systems have been created that can match or even surpass human beings in specific task environments. For example: 'Deep Blue' [1] has beaten a world champion chess player in a six-game chess match in 1997, or AlphaGo Master, which beat the world's best player at Go with a score of 100-0 in 2017 [2]. Furthermore, by using Reinforcement Learning and Artificial Neural Networks, we have successfully created artificially intelligent problem solvers, gaming masterminds, reasoners, speech processors, surprisingly accurate personal preference predictors or even systems that can tell what your future actions will be by just analysing your Facebook feed [3]. In this thesis, however, the current research paradigm, as dominated by Reinforcement Learning combined with Artificial Neural Networks, in the development of Artificial General Intelligence will be challenged.

As of yet, a machine that thrives in chaos, an 'artificial human', commonly referred to as an Artificial General Intelligence (AGI), has not been designed. Deep Blue was not able to play any strictly simpler game than chess; it could not do anything else at all. Because that generality is what no robot has ever achieved. Nonetheless, many experts do think it is possible. In fact, some researchers even accept the hypothesis that you, me and everyone we know are all machines [4]. Quite exceptionally, Friedrich Nietzsche stated in 1888 that we cease to understand human beings when we stop regarding them as machines [5]. And this makes sense: an infant can be seen as an embodied neural network, with the world as its 'dataset', and parental guidance as a 'scoring function'. What would it take to successfully create a machine like that?

Theoretical Background

The answer to that question is complex. Reinforcement Learning (RL) has been used in AI since the mid-80s: a way of programming agents to learn by reward and punishment without specifically saying *how* a task should be performed [6]. It has gained interest in the field of machine learning rapidly ever since and has been a main kind of AI-learning algorithms for the past few decades. It is most often used in combination with some kind of Artificial Neural Network.

The Artificial Neural Network (ANN) is a computer model that is based on biological neural networks. It contains programmed representations of neurons and synapses and learns in a similar way: by adjusting the artificial equivalent of a neuron's firing rate to approach the most preferable output. The ANN dates back to 1948, when Alan Turing introduced the 'Unorganized Machine', which he stated was the simplest possible model of the nervous system [7].

Regarding AGI, the problem is that most of the systems mentioned above are very narrowly programmed or trained to perform *one task* very well, while not being able to perform any other task (also referred to as: 'narrow AI'). This, however, might be counterproductive in the face of AGI, and therefore a more recent AI-learning method will be evaluated: Imitation learning (IL).

To clarify: RL and IL are both learning methods used to train interpreter systems, that get input from observations of the environment and output some form of action – in the broadest sense of the word. RL roughly stands for learning through reward and punishment, and IL for learning through imitation. The systems they train, configure, or optimise, can be either artificial (AI) or biological (e.g. the human brain). The ANN, however, is a kind of system *architecture*, a *network* of artificial neurons, that can be trained or conditioned - through some sort of learning technique - to perform transformations on an input required to achieve the desired output. These three concepts are not mutually exclusive, and can even be combined in AI-systems, but they can also be used separately; IL or RL are both used as training methods for architectures other than the ANN, and an ANN can also be trained by using some other learning method.

Since this millennium, IL has been a method to train AI [8], by giving systems the ability to imitate what they perceive. The oldest and most basic method of IL is a physical skill learner, which first requires the system to learn to move, and to deal with visually different task environments [9]. Historically, the latter required many training examples, a long training time, and has often been done through RL. Note here that this is not a state-of-the-art form of IL; new IL-methods are somewhat more complicated, require significantly less training examples, and do not make use of RL whatsoever (See: Chapter 2). When the IL-machine is trained properly, it can learn to perform new tasks from few human demonstrations [10]. These can be given to the robot in the form of an array of pictures, a physical performance of the task in front of its camera, or through teleoperation by a human in Virtual Reality. This learning method, however, has been around for a lot longer, as it is how a great deal of animals including humans acquire new skills and knowledge at all ages.

IL is therefore similarly relevant for AGI as RL is: the usage of RL in AGI-research is mainly motivated by the fact that "it is how neurons are trained, as well" [11]. But is the goal to recreate the human neuron, or networks of them, or to create agents closely resembling full-fledged human beings? Then, IL could be considered, as it is a method by which human beings are trained, consistently, albeit gradually declining in importance, from the age of two weeks until death [12]. It is often superior in terms of dealing with multiple chaotic and dynamic environments, and greatly so at generalising knowledge to new tasks. IL is a crucial learning method for human beings, but it is underrepresented in research on AGI when compared to RL. Therefore, a partial paradigm shift to the inclusion of more IL will be argued by comparing it to current attempts at AGI that solely use some form of RL, combined with some form of ANN.

Thesis Structure

This thesis consists of three main chapters. The first two will feature historical and/or contextual information on the concepts of Artificial General Intelligence, Artificial Neural Networks and Reinforcement Learning, and Imitation Learning, respectively. These chapters summarise the most relevant recent research into each of the technologies mentioned above, and briefly describe several problems with the current research paradigm in AGI-research, to form a theoretical foundation for the third chapter. The third chapter challenges this paradigm and discusses the relevance of imitation learning for AGI-research in the context of human imitation learning. It will be demonstrated that imitation learning is as essential for human beings, and as integral to the human mind, as is reinforcement learning. Furthermore, the current states of research on these learning methods will be compared and related to their human counterparts, and finally, contributions of imitation learning to current AI-systems will be discussed.

NB: in this thesis, definitions and arguments concerning the concept of 'consciousness' and what that means for AI are deliberately evaded, as even the basic definition of *human* consciousness is still a topic of heavy philosophical discussion, let alone its translation to artificial consciousness. To summarize all of the above and to conclude this introduction, the main research question is formulated as follows:

How could advancing Imitation Learning techniques contribute to the future development of Artificial General Intelligence?

Chapter One: Defining Concepts

Artificial General Intelligence

Going back to the beginning of AI, it begins with the 'Turing Test', first designed and described by Alan M. Turing in 1936 [13]. He has devised a thought-experiment that subsequently formed the beginning of research into machines that resemble human intelligence. It goes as follows: you chat with a man acting like a woman, and a woman trying to convince you that the man is lying. By chatting, you must determine who really is the man. The first time, you chat with two ordinary people, and the second time a machine takes over the role of the man. A proof of concept for AI is established if your guessing accuracy is independent of the presence of the machine. Turing concluded that there were no problems we could not overcome to successfully develop what we now call AGI. Later, in 1955, a modern definition of AI was formulated by four later Turing Award-winners. They defined AI in terms of a machine's ability to think, understand and learn in a similar way to human beings [14].

So, what has come of this definition through the past 60 years? In his research, Y. Pan mentions that there have always been two main driving forces for the development of AI: research, and the information environment, with the latter being the strongest [14]. Also, in terms of the latter, the world is very different from the one in which the definition of AI was formulated. We now have the widely used technologies of the Internet, big data, e-commerce and much more, and this is calling for a lot of rapid development and financial investment in the field of AI. This influence from the current information environment, however, leads to a fundamental problem for the creation of AI as defined in 1955. As was stated in the introduction, creating many ungeneralizable AI-systems, each specifically trained for a small subsection of the information environment, is opposing AI's original definition and thereby in a way inhibiting the creation of AGI.

The definition of AGI – as used from this moment on – does not differ from the original definition of AI, which was, in short: 'a human-like machine'. Additionally, humankind has one feat one can consider to be greatly responsible for our evolution into dominance: our independence of task environment, our generality. AGI, as a concept in this thesis, will stand for machines with human-like thinking and learning that can operate in greatly different, ever-changing and effectively infinite environments. More specifically: AGI systems will be considered to be systems that autonomically communicate, acquire knowledge and learn new physical and abstract ('verbal', immaterial) skills.

The Artificial Neural Network

This section proceeds with the most commonly used AI-learning model in successfully predictive, modern-day AI-systems: the artificial neural network [15] (ANN). In this paragraph, the ANN will be explained and two of its variants in current AI-research highlighted. "Artificial Neural Network" is the general term for networks consisting of layers of nodes with between-layer connections. The connections within such a network are 'weighted', meaning the output of the previous node is multiplied by an integer, or "weight", and passed on to the next layer of nodes, thereby giving the input value a measure of 'importance' for the next layer, much like the neural networks in a human brain. The regular ANN has three layers: an input, output and hidden layer. Deep Neural Networks (DNNs) are ANNs that have multiple layers between the input and output layer. The hidden layer(s) – in both ANNs and DNNs – perform(s) all sorts of transformations on the input. This eventually produces an output, by which the whole network is then evaluated through a scoring function based on the correctness of that output, which alters the weights in the network's connections in a way that promotes performance accuracy. This last step is often referred to as 'training the network', which for all ANNs is most often done using Reinforcement Learning (next paragraph). Methods for training a DNN i.e. multi-layered ANN are often collectively called 'Deep Learning' (DL) methods.

Adding more hidden layers to the basic ANN – creating the DNN – had a massive impact on what AI-systems could achieve. The reason for this is that a greater amount of hidden layers, allows for a greater amount of sequential transformations on the input, therefore greater levels of abstraction from the input, which has allowed us to create systems that can learn, for example, object categories from raw data [16].

A specific variant of DNNs, called Convolutional Neural Networks or Deep Convolutional Networks (CNNs/DCNs), consist of so-called 'convolutional' hidden layers. These are currently used in image classification, and work in a similar way to human receptive fields in the early visual cortex. The CNNs hidden layers use 'tiled filters', a mathematical matrix, by which it multiplies the input, thereby revealing pixel patterns in the input [16] (Figure 1). An example of a filter used by CNNs is an 'edge detection' filter, shown in Figure 1. The left filter multiplies all sets of 6x6 pixels from the input to detect broad, hard edges, and the right one is multiplied by all 3x3 sets to reveal slightly smaller, relatively softer, gradient edges. This one-by-one multiplying of all suitable sets of input pieces by a certain filter is also called 'convolution': the filters 'convolve' over the input data, hence the name.

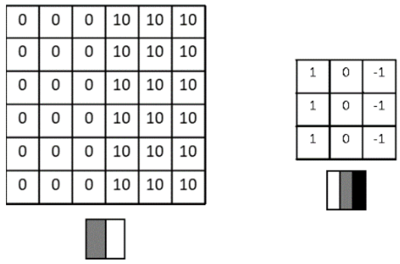

Figure 1: Matrices used by CNNs in multiplication with input-pixels for detecting edges in the image.

Reinforcement Learning and Google DeepMind

This paragraph will define and inventory the variants of Reinforcement Learning (RL), of which a specific variant, combined with a form of DNNs, is currently used in the endeavour to create AGI by Google DeepMind, which, later on, will be compared to IL-algorithms.

RL involves training (an) agent(s) in a task environment to maximize some pre-defined cumulative reward. The agent interacts with its environment in two ways: it gets imperfect observations of its environment, which it uses to build and update a model of that environment, and it can take actions based on that model. Specific actions are selected based on which one, for that specific state, leads to the highest cumulative reward as predicted by the model [6].

A specific form is end-to-end RL, which means the entire process from input to output in agents involves a single, layered neural network without modularization, trained by RL. In contrast to regular RL's learning for actions, it can develop a lot of abstract higher-level functions by incorporating the entire sensors-to-motors process in the learning algorithm.

Another variant of RL is Q-Learning, which is similar to regular RL, yet differs in terms of what it learns, updates and outputs. It is a model-free variant of RL, and, instead, it learns a policy based on the observations of the environment. The difference between a model and a policy, is that a model is some sort of simulation or substitute version of the real environment, to manipulate and check actions on, while a policy is an arguably simpler list that says what action to take in certain circumstances [6]. The value of each action, as given in the policy for achieving maximum rewards, is called the Q-value.

The application of RL in situations with defined reward functions (i.e. in games) is superior to IL, as it can be directly optimized. In robotics, however, reward functions can be undefined or sparse, which makes RL get stuck in local optima [17] i.e. states in which no action provides any further improvements on generated models or policies, in spite of the fact that there are possibilities for improvement in some other state.

Google DeepMind – Deep Q-Learning

'Google DeepMind' (or just DeepMind pre-2014) is a company that greatly contributes to research on developing AGI. DeepMind's CEO Demis Hassabis describes their main mission in two simple steps: 1: "Solve Intelligence", 2: "Use it to solve everything else" [18]. They use the framework of Reinforcement Learning in their research on developing 'learning systems' that can generalize to any task, in any environment – i.e. AGI. They motivate their use of this framework in terms of RL also being used in the human brain, even being hard-coded to the level of single neurons in the dopaminergic system [18] [19].

In context of this research, Google DeepMind patented a system/network which they called a Deep Q-Network (DQN) in 2015 [20]. This 'DQN' uses a form of Q-learning used as a DL method – coined 'deep Q-learning' – to train a CNN for continuous image recognition in Atari games, which allows it to learn to play several different Atari games from scratch. Regular Q-Learning methods are known to be unstable when used to train large neural networks, because of the potential of small updates to the Q-value to significantly change the generated policy and because of correlations in the sequence of observed data [16]. Google DeepMind innovated on this twofold. First, they used an iterative update that adjusts the Q-value toward some target Q-value that is only updated periodically, to remove correlations with the target. Second, by using 'experience replay', they randomised the observation sequence to remove any further correlation resulting from constant input sequences [16].

This deep Q-learning method is then used to generate and update a policy that maximizes action-rewards, as predicted by the CNN. As input they used an 84x84x4 image, the Atari's screen, that passed through three convolutional layers, then through two fully connected hidden layers interpreting the filtered data, and that led to an output of the best in-game action to take [16]. As a side-note, this can also be called 'end-to-end' RL, as the whole process, from input to output, is taken up into the CNN. In short, they use the raw image data, interpreted by the CNN, to generate, by deep Q-learning, a state-action policy that maximizes the cumulative reward, which in the Atari situation is equal to the game-score. They used the exact same setup for the DQN to learn to play several different Atari games. The system went on to outperform professional human players at 49 different games. This paragraph formed an introduction on Google DeepMind's AGI-research: more on this will be discussed in chapters two and three.

Major Problems with RL-ANN Systems for AGI-research

Making AI-systems learn by using ANNs, trained through usage of traditional RL-methods, constitutes two problems for developing AGI, that can be regarded as shortcomings of RL-methods. First and foremost: creating systems that perform one specific task highly accurately is a step away from and not towards AGI. It is often expensive to reprogram or tweak these systems to be able to operate in new situations, and their affordances and abilities for task performance are always limited to situations the human operator has considered [21]. The success of using these combinations for superior task accuracy, however, and the lack of immediate benefits of researching other applications, is what leads to continuous financial and computational resource investments for single-task performance. Not all agree with that, however, as many other scientists think the desire for increasing computational power is not what leads humankind to creating AGI [4] [8] [22] [23]. In fact, R. Brooks stated as early as in 1991, before the rise of imitation learning, that it would be better for reaching higher intelligence to design and test robots in the dynamic, ever-changing environment of the real world, even though that would definitely not be very rewarding at the start [24].

The second problem with these systems is that, for them to be properly trained through RL, a precise formula for what constitutes a ‘correct’ output is necessary. This is, however, nigh impossible in most everyday situations human beings encounter, that require, for example, social intelligence. There is no widely applicable, always-valid scoring function to evaluate our dealing with those situations, as there are no ‘correct’ outcomes. It can even be argued, in this example, whether a notion of ‘one outcome being better than some other’ does in fact exist. And even if one has some vague notion of how to perform a certain skill, through the – so-called - Dunning-Kruger effect [25], one will think (s)he is performing well, when (s)he actually is not. To recap; performance in a dynamic world, on a wide task variety, with imperfect observations and mostly inconsistent rewards has been persistently problematic for RL-algorithms. In the next chapter a relatively novel learning method is discussed, which might be able to resolve these problems: Imitation Learning.

Chapter Two: Imitation Learning in Recent Research

Why Imitation Learning?

To exhibit the concept of imitation learning (IL), and to underline its relevance in AGI-research, AI-research that already prominently features IL will be briefly summarised. The simplicity of learning by imitation, and integrally, of teaching by demonstration as performed by human beings, has not yet been implemented into AI-systems [8]. However, it has led researchers to develop systems capable of basic IL, and to research the concept of IL for the purpose of developing AGI.

In fact, many studies in the fields of the behavioural and psychological sciences, neuroscience, and computer science have stated the relevance of IL for AGI [26], some of which are even similar to this thesis [8] [12] [27]. These studies were mostly conducted in the late 1990s and the early 2000s, when IL-techniques for AI had mostly been only described, but in 2019 many of these descriptions have been practically realised [9] [17] [28] [29]. IL, however, has not come to be as relevant for AI-research or for AI-implementation, as many researchers thought it would. An example of this is a study from 2004 that proposed and applied IL to make computer game characters more life-like [30], but, as of today, no prominent games feature this strategy. Another example is that, in their current endeavour to create AGI, Google DeepMind mainly uses RL-frameworks [11] [16], while the fundamental problems that RL-frameworks cause for achieving generality in an AGI-system, were already extensively reviewed in a study by Stefan Schaal (1999) [8]. In light of this review, recently developed AI-systems that make use of IL-methods will be discussed. First, a review of two systems that primarily use IL-methods will be given, and secondly, a review of two primarily RL-systems enhanced with IL-methods. It will be shown that IL enables learning in dynamic environments where RL alone cannot, that IL-methods have a significantly higher learning speed for certain tasks, and that currently they can be added to RL-systems to improve training, learning and task performance.

Methods of Imitation Learning

The most basic forms of imitation learning (IL) are: Behaviour Cloning (BC) and Dataset Aggregation (DAGGER). BC is a simple algorithm that learns a policy through supervised learning on state-action pairs extracted from demonstrations, but struggles to generalise outside of demonstrated data. DAGGER overcomes this generalization-problem by intertwining expert and self-learned policies, but requires an expert to be present during the entire training stage [31]. For the purpose of this thesis, however, two more recently developed forms of imitation learning (IL) will be reviewed that overcome this problem. Many other forms of IL involve a fallacy already discussed for RL. While still able to outperform RL in terms of task variety, they remain designed to work in isolation, i.e. in a discrete environment, and require precise feature definitions or a large amount of samples to function properly [9].

One-Shot Imitation Learning

One form of IL designed to overcome this problem is ‘One-Shot Imitation Learning’ (OSIL), as demonstrated in a study by Duan et al. in 2017 [9]. The goal was to create a meta-learning framework that would enable imitation learning systems to “learn to perform any given task from very few demonstrations and instantly generalize to new situations of the same task, without task-specific engineering”. A block-stacking robot arm was created, involving an ANN that was trained to manipulate the blocks, but not to perform any task on them. For training methods, they used BC or DAGGER in different experiments.

With just a single virtual demonstration, the robot could successfully perform a block-stacking task, even when the initial position of the blocks was different from the demonstration. For a visual demonstration *click here* [10]. They concluded that: "by training such a model on a much greater variety of tasks and settings, we will obtain a general system that can turn any demonstration into robust policies that can accomplish an overwhelming variety of tasks."

The main shortcoming of this research was that it was still necessary to provide many (about 100,000) training examples to train the ANN for basic object manipulation. They were, however, successful in creating a system that could instantly generalize to new tasks within the task environment of block-stacking, in a way that RL alone could not. They provided a proof of concept for a framework that could theoretically do much more, and which has since been built upon.

Zero-Shot Imitation Learning

One particular follow-up study from 2018, that resolved the issue concerning extensive ANN-training, was 'Zero-Shot Imitation Learning' [29]. A method was proposed that differed from OSIL in that it *did not require any expert supervision* in the training stage. They demonstrated an IL-system that trained through autonomic, self-supervised exploration of the task environment. From observations it derived potential goal-directed skills, which it would recall and reorganise upon viewing a human demonstration to subsequently perform a task. The system was tested in a situation from earlier research on 'Rope Manipulation' [32], and on 'Navigation in Indoor Office Environments'. In both cases it outperformed the baselines, which were methods that still required multiple step by step demonstrations or were based on other sub-forms of the same method, respectively.

IL for AI shows similarity to imitation as performed by human beings [12]. Exploration and demonstrations are used to find new ways to manipulate task environments or to learn to perform previously demonstrated actions through imitation. This overcomes a problem that RL has persistently had difficulty dealing with; the 'exploration problem', which will be discussed in the next section. The last concept, 'Zero-shot Imitation Learning', approaches another essential factor of AGI: autonomy in the training stage. Considering IL has only been in development half as long as RL, and the methods mentioned above being researched in quick succession over the last two years, IL can be considered to be in a stage of rapid development, which is promising for future research.

Reinforcement-Imitation Learning Systems

Now comes the return to Reinforcement Learning algorithms, specifically, to RL-systems that have been enhanced to learn from visual demonstrations i.e. equipped with IL. Exploring an environment with sparse rewards has persistently been difficult for RL-systems, and finding a reward is increasingly more difficult in problems with greater task horizons or action potentialities. Many real-world tasks are therefore out of reach for RL-methods [17], but not out of reach for those that also use IL. Methods combining RL and IL learn an initial policy through demonstrations - using IL - and improve on it through RL. This overcomes the 'exploration problem' and enables RL to work in more dynamic environments. It also enables RL-systems to function properly in sparse-reward environments, which might result in RL-IL combinations being fully functional in the real world at some point, although this would require the system to continuously alter, completely change or altogether take on a different reward function based on continuous observation. Enhancing RL-algorithms with the ability to imitate starts to partly solve the main problems with traditional RL-ANN combinations as mentioned at the end of chapter one. It enables them to perform a wider variety of tasks properly, while maintaining accuracy at any single one of those tasks [33]. Imitation also, theoretically, allows for a continuous redesign of the reward function used to train the ANN through making use of human demonstration or feedback. There has been a lot of recent development in the area of RL-IL-algorithms, and what follows is an inevitably short and incomplete summary of only two recently developed methods to broadly outline how such combinations work as opposed to solely IL or RL-algorithms.

Combining the Two: DDPGfD

We start with DDPGfD, or Deep Deterministic Policy Gradients from Demonstrations. DDPGfD is an RL-algorithm designed for continuous-action-environments, combined with the ability to learn policies from demonstrations. 'DDPG' itself is an 'actor-critic' algorithm (refer to Figure 2 for specific details [34]), and is often used to increase learning stability of RL-methods. It does this through the addition of a 'replay buffer', a sort of cache storage for previous observations, which enables the algorithm to alter its value prediction and action evaluation parameters 'off-policy', i.e. without making use of information from its policies.

1. take action $\mathbf{a} \sim \pi_\theta(\mathbf{a}|\mathbf{s})$, get $(\mathbf{s}, \mathbf{a}, \mathbf{s}', r)$
2. update $\hat{V}_\phi^\pi$ using target $r + \gamma \hat{V}_\phi^\pi(\mathbf{s}')$
3. evaluate $\hat{A}^\pi(\mathbf{s}, \mathbf{a}) = r(\mathbf{s}, \mathbf{a}) + \gamma \hat{V}_\phi^\pi(\mathbf{s}') - \hat{V}_\phi^\pi(\mathbf{s})$
4. $\nabla_\theta J(\theta) \approx \nabla_\theta \log \pi_\theta(\mathbf{a}|\mathbf{s}) \hat{A}^\pi(\mathbf{s}, \mathbf{a})$
5. $\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta)$

Figure 2*: An 'actor-critic' algorithm combines policy gradient and value-function learning and consists of a 'Critic' which estimates the value-function (V_φ^π) in every state (s), using the reward (r) and the target state's (s') perceived value from V. It then evaluates the next action ($A^\pi(s, a)$) by using the action's reward ($r(s, a)$) added to the subtraction of the current-state value ($V_\varphi^\pi(s)$) from the estimated next-state value ($V_\varphi^\pi(s')$). The 'Actor' then updates the policy distribution as suggested by the 'Critic'. [34]*

This DDPG has been enhanced with the ability to learn from demonstrations to form DDPGfD [17]. DDPGfD adds the demonstrations to its replay buffer and learns and updates the policy network by sampling observations from that buffer. Before running the algorithm, the replay buffer contains only demonstrations, which are used to learn an initial policy. It has been successful at peg insertion and similar tasks, speeding up the solving process when compared to traditional RL-methods, and in recent research has been optimized to perform a wide variety of tasks unsolvable by RL alone [17]. However, the longer the runtime, the more insignificant the demonstrations become to updating the policy network, therefore making it unable to autonomically generalize to new task(s).

Combining the Two: DQfD

More recently, Deep Q-Learning from Demonstrations (DQfD) was also researched by Google DeepMind in later research on their Atari-DQN (Chapter 1). Implementing demonstrations into a replay buffer massively accelerated the AI's learning, and – in contrast to DDPGfD - included giving the perceived expert-actions higher Q-values than all other actions. To provide the demonstrations, an Atari game only needed to be played for five minutes or so [33]. It outperformed all similar methods in the same context and is a prime example of improving RL-methods by introducing IL. Nonetheless, for DQfD, it could happen that the algorithm was not able to improve on the expert's performance [17].

Chapter Three: Imitation Learning and Artificial General Intelligence

In this chapter, the current paradigm in AGI-research, in terms of it prioritising Reinforcement Learning is, challenged [16], by an illustration of the relevance of Imitation Learning for the only general-purpose learning algorithm known to us: the human mind. To do this, Google DeepMind's (GD) particularly successful contributions to AGI-research [2] [16] [18] are treated as representative of the current paradigm in AGI-research. This was chosen to be representative, since it gained enormous amounts of praise and attention in the last few years, both publicly and in the field of AI. For example, experts in both the fields of AI and Go publicly stated that GD's AlphaGo AI was at least "A decade before its time" [18]. GD's research on AGI became very popular, mainly due to AlphaGo, to the extent that it was written about by multiple prominent international news institutions [35] [36]. Their AGI-research is regarded as our most advanced effort at creating AGI, with some sources even loosely claiming the company has already created an AGI [37].

Nevertheless, their motivation for mainly using RL-frameworks in research on AGI, as given by Demis Hassabis (CEO), is challenged in terms of it mostly leaving out imitation learning. The simplest reason Hassabis provides for using the RL-frameworks, is, quote: "Our brains use Reinforcement Learning as one of our main learning mechanisms" [11]. Indeed, it seems reasonable to use a learning method that is essential to the human mind in research on AGI. Moreover, in the brain's dopaminergic pathways a form of RL called 'Temporal Difference Learning' is biologically implemented, down to the level of single dopaminergic neurons [11] [19]. For these reasons, Hassabis claims that 'if we could solve everything about RL, we know that this would be enough for AGI' [11].

However, our brains use IL as one of their main learning mechanisms as well. Moreover, IL is also implemented in our brains to the level of single neurons, specifically in so-called 'mirror neurons' [38]. So why should RL be preferred over IL in research on developing AGI? Future research for the purpose of understanding the way these learning methods are combined in the human mind might be key to the development of AGI: a view in contrast – or perhaps addition – to GD's claim that solving RL will be enough. In other words, the realisation should come that RL is *only one* of our main learning methods, and correspondingly, this view poses new questions for AGI development: can it be created with only one learning method? Or are others as vital for success? Since RL-methods are already extensively researched in terms of their potential for developing AGI, this chapter focuses on what IL-methods might contribute.

This view will be supported in two ways, in the following two paragraphs. First, by a summary of literature on human IL, it will be clarified that IL is as essential a learning method for human beings, and as integral to the human mind, as RL is. Secondly, further details are given on how IL-techniques are already being applied to RL-systems to improve on their performance, followed by a summary of prior research on the importance of incorporating IL into AGI-research.

Imitation Learning in the Human Mind

Imitation is widely seen as a fundamental part of human learning [39]. The value of IL in general is substantial; expert demonstrations can significantly reduce the amount of states and actions that need to be visited or evaluated by a learning agent, for it to reach a near-optimal strategy for solving a given problem [40]. As such, imitation can be seen as an accelerator for learning [26]: instead of working out problem-solving strategies by trial and error, human beings can simply learn these through watching - or remembering - expert demonstrations.

IL: Essential for human learning

A human being's capacity to selectively apply and learn from imitation is exceptional. Less than an hour after birth, human beings engage in imitation in the form of protrusion of the tongue as a reaction to a caregiver doing the same [41]. They gradually improve on their ability to imitate observed behaviours over the first few years of their lives. Infants start to imitate their caregivers *voluntarily* and *intentionally* when they are about a year old [42], the prior meaning they do not automatically imitate whatever they observe, and the latter meaning they imitate only if there is something to gain. Moreover, children do not copy any behaviour when they do not consider it the most rational alternative [42]. Some studies have even found that infants spend the majority of their time mimicking previously demonstrated behaviours [12]. Imitation Learning is responsible for most learning processes that occur early in a human life, for it is suggested to be "almost the only tool in the infant's toolbox" [43], as a lack of basic knowledge and self-supervisory abilities inhibit the infant to use RL. In these early years of life, by imitation, human beings learn to perform a wide variety of cognitive and physical tasks, from executing precise movements, to conversing with other individuals. In the first few months of life, this can only be done through direct imitation, yet children get increasingly more adept at 'deferred imitation' over the years [44]. Deferred imitation has been defined as 'the ability to reproduce a previously witnessed action or sequence of actions in the absence of current perceptual support for the action'. By imitation from memory, a human being is able to observe and gather knowledge in one test environment, and subsequently use that in another [44]. Imitation learning in general allows for a rapid expansion of knowledge and skill into adulthood.

Adult human beings mostly use RL-methods to improve on their task performance, which becomes possible due to a large collection of data on previous observations [4], allowing training independent of other individuals. Given that human demonstrations are usually not perfect nor sufficient for near-optimal performance, RL gives an individual the ability to improve on observed demonstrations [21]. Nevertheless, IL-methods are regularly used when learning completely new skills, or when the requirements of the task at hand are not entirely clear [43]. This could also be an explanation for the heavy use of IL by human infants, since – at their stage of development – most skills are completely new and task requirements often are not entirely clear. Thus, in the context of human learning, Imitation Learning can be regarded as mostly responsible for the initialisation and the early stages of learning, and Reinforcement Learning can be seen as mostly responsible for the later stages of improvement by repetition and observed rewards. This underlines the essentiality of *both* learning methods for the development of the only exemplary general-purpose learning algorithm i.e. the human brain, and thus for the future development of tabula rasa AGI-systems.

IL: Integral to the human mind

At the core of a human being's exceptional ability to imitate is a specific type of neuron mentioned earlier: the 'mirror neuron'. Mirror neurons are a unique set of premotor neurons – specifically in area F5; the human inferior frontal gyrus – firstly discovered in macaques in 1992 by Rizzolatti et al. [38], and named for their exclusive property to fire during both action performance and observation [45]. Whilst sitting completely still, a macaque observed another macaque performing a specific action, and later performed that same action itself. Both times, the single mirror neuron that was recorded fired as a response to each of the stimuli. This type of recording is seldom performed in the human brain, but indirect measures support the existence of a human mirror neuron system (MNS) [45]. The human MNS is suggested to be a part of a broader network that includes the inferior parietal lobule, the superior temporal sulcus and regions of the limbic system [45]. This network has the ability to map perceptions of the environment to internal sensorimotor representations, which effectively enables training of task performance *during observation*, as well as during task performance [8] [26]. To clarify: one's observation of an action that is performed by another individual activates the same (pre-)motor networks as are activated during one's own task execution. Additionally, the human MNS activates upon hearing a word uniquely associated with a specific action [46], and has been proven to distinguish between different levels of sociality [45].

Contributions to current AI and future AGI research

As mentioned in the introduction, some AI-systems using RL-techniques can outperform a human being at specific tasks, even though the human has learned to perform the task using similar techniques [1]. However, this is not the case for Imitation Learning, regardless of both techniques being similarly essential to the human mind, as demonstrated in the previous section. AI-systems that use IL often still require a great amount of training examples and preparative use of RL-techniques to learn basic movement and to get familiar with the task environment [9], and from that can only learn by imitation to perform simple tasks, such as block-stacking, rope tying and navigation in discrete task environments [10] [29]. A combination of RL and IL techniques might be more effective for most applications [17], yet, in some cases, RL-IL systems are unable to improve on expert demonstrations [33], or become unable to generalise to new tasks, due to the way in which demonstrations are stored and subsequently used in the learning process [17]. Moreover, all the IL-techniques in the field of AI currently still only allow for imitation of physical movements or behaviours [9] [10], or for imitation of virtual action i.e. gameplay [33].

In contrast, human IL-techniques are suggested to be responsible for language acquisition [38], development of social skills, transfer of behaviours and initialisation of skill learning processes [43]. Even stronger claims about human IL include statements such as: 'mirror neurons are the driving force behind "the great leap" in human evolution' [47], and that IL 'provided the [evolutionary] foundation for arguably unique human social skills such as theory of mind, empathy and language' [48]. Human beings have the capacities for both direct and deferred imitation, and the latter, to some extent, allows for generalisation of knowledge from one task environment to another [44]

Thoroughly understanding this generality – or independence of task environment – is widely suggested to be the key to successfully developing AGI [18], yet no system has achieved any generality in context of reasonable performance in multiple non-similar task environments. For example, while GD's DQN was able to outperform professional human players at 49 different Atari games, using the same system architecture and higher parameters [16], it would not have been able to perform tasks outside of the Atari environment, let alone outperform human competitors. And in subsequent research, in which IL was added to DQN to create DQfD, the addition of IL only meant it was able to use demonstrations by retrieving them from the same replay buffer as its own experiences. Another example is DDPGfD, as compared to systems using regular DDPG, in which the addition of IL allowed for performance on a wider task variety, yet its increasing use of RL-methods proportional to increasing runtimes, ultimately made it unable to generalise to new tasks. These problems can be attributed to the fact that current IL-methods have no form of selective applicability, i.e. the ability to apply IL only if it is the best alternative, while selectivity might be one of the aspects of human IL most relevant for their general learning performance [42].

However, even though AI-IL is in its infancy, when compared to human IL, additions of current IL-methods to existing RL-systems can already lead to massively increased learning speeds [33], superior policy initialisation [16], the ability to deal with greater task horizons and action potentialities [16], and better task performance in general [28]. Imitation Learning effectively 'resolves' the exploration problem posed by RL-methods. It can be used on its own to create systems without a need for expert supervision [29], or in combination with RL-methods, which might greatly increase general system performance [28]. DQfD outperformed all similar learning methods in the same problem context [33]; DDPGfD outperformed all traditional RL-methods in terms of solving speed and task variety [17].

Summary and Conclusion

In the first chapter, AGI was defined as: 'Machines that can freely and autonomically communicate, acquire knowledge and learn new abstract and physical skills, which have the ability to operate in hugely different, ever-changing and effectively infinite environments', and it was stated that creating many highly accurate single-task – yet ungeneralizable systems – will not help in creating the former. Google DeepMind's DQN was introduced as a prime example of how RL and ANNs alone currently contribute to AGI-development. Finally, two major problems for AGI-development resulting from the current paradigm as dominated by RL-ANN systems were stated: again, their non-generalisability, and the high level of specificity of the reward function that is required for the traditional version of such a system to function properly, which makes them inapplicable in real world situations.

The second chapter further argued for a paradigm shift in favour of Imitation Learning, and this thesis was placed in context of prior research on IL-methods. BC and DAGGER, two basic IL algorithms, were briefly explained, and two IL-methods developed in the last two years were reviewed. 'One-Shot Imitation Learning' was demonstrated in 2017 on a system that, after training, could perform a task with very few demonstrations. However, it still made use of RL-methods to learn basic movement, which required many training examples. That was not the case for 'Zero-Shot Imitation Learning': an IL-method that built upon One-shot Imitation Learning by adding the ability for a system to autonomically explore the task-environment and store observations as potential goal-directed skills. Demonstrations aside, it notably required no expert-intervention whatsoever. Lastly, two RL-methods that incorporated IL were discussed: DDPGfD and DQfD. They both are versions of RL-systems that incorporate imitations by adding the demonstrations to their 'replay buffer', a sort of cache memory for past experiences.

The final chapter challenged the current paradigm of AGI-research in terms of it being dominated by RL-methods, and it was clarified that this thesis' intention is not to promote the replacement of RL-methods by RL-methods, but to motivate future research on how these two work together in the human mind and on how they could in an artificial system, which, due to the current dominance of RL, means promoting research on IL-methods. Specifically, the motivation given by Google DeepMind for mainly – and for the most part: only – using RL-methods in AGI-research was challenged. GD states RL is essential for human learning and is hard coded in the brain, down to the level of single neurons. And since the human mind is the only form of general intelligence known, GD states that a full understanding of the RL-process will lead to AGI. In line with this reasoning, it was stated that a full understanding of the IL-process might be necessary for AGI development as well. To demonstrate that the two are of equal importance, it was underlined that IL is as essential to human learning, and as integral to the human mind, as RL is. In the last section, all chapters were brought together, and artificial IL and human IL were compared to reveal several discrepancies between the two. Resolving these discrepancies might give key insights for the development of AGI. Lastly it was shown how, even though artificial IL is still in its infancy, the application of current IL-methods to RL-systems can already massively improve on general task performance.

Since adult human beings make use of many different learning techniques voluntarily, selectively and intentionally, it is best to conclude that future AGI-systems are most probably neither going to be pure IL-systems, nor pure RL-systems. A greater focus on equalising the discrepancy between the two, however, could be beneficial to future AGI-research. Based on a literature study summarised in the first two chapters, I answered my research question of 'what the development of imitation learning techniques could contribute to the future development of AGI' in the third chapter. Inspired by the importance of IL for the human 'system' i.e. brain, that answer, in short, in my opinion, is *a lot*.

In terms of the current contribution of IL in research on AI and AGI, I wrote on the persisting shortcoming of RL-algorithms, in terms of what is commonly referred to as the 'exploration problem'. Being able to explore an environment autonomically is trivially crucial to the development of human intelligence and is essential for any self-learning system to function properly in the real world. IL-algorithms do not suffer from the exploration problem at all, and current RL-systems, like the ones mentioned in the last paragraph of chapter two, can overcome it by implementing some form of IL as well. In a research that compared artificial RL, IL and a combination of the two, it was found that the latter significantly outperformed both other systems at a variety of visuomotor skills [28].

This thesis has two main shortcomings. Firstly, it is in no way practical or experimental, while it attempts to answer a question that involves a lot of complex technicalities. This was intentional, however, as actually formally advancing imitation learning techniques in the foreseeable future is what was hopefully encouraged in this thesis. Furthermore, future research on the same topic would benefit a lot from actual experience within the research fields I, relatively inexperiencedly, only scratched the surface of. This being a Bachelor Thesis, it suffers heavily from a lack of experience, limited resources, time and manpower and most probably some concepts have been overstated and/or underexplained.

Nonetheless, the main statement of this thesis is sound. There is no reason for preferring research on RL-methods over IL-methods in the field of AGI: both are equally effective learning strategies, and each has its own optimal environment type, both are essential to the human mind i.e. the only example there is of a general purpose learning algorithm and they can already be combined to significantly outperform both their separate subcomponents. This should be sufficient to motivate more research on IL-methods within the field of AGI. Future research in the field of neuroscience could be done on fully understanding the workings of the human mirror neuron system, and on how such a system might be created artificially. In the field of AI, future research can be done on ways of implementing properties of human IL that are not in any way present in artificial IL, among others: selectivity and learning from written or spoken demonstrations.

References

[1] "IBM100 - Deep Blue," [Online]. Available: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/. [Accessed 17 6 2019].

[2] D. Silver, A. Huang, C. Maddison, A. Guez, L. Sifre and G. Van den Driessche, "Mastering the game of Go with deep neural networks and tree search.," *Nature,* vol. 529, no. 7587, pp. 484-489, 2016.

[3] S. Biddle, "Facebook uses artificial intelligence to predict your future actions for advertisers, says confidential document," The Intercept, 13 April 2018. [Online]. Available: https://theintercept.com/2018/04/13/facebook-advertising-data-artificial-intelligence-ai/. [Accessed 16 June 2019].

[4] R. A. Brooks, "I, Rodney Brooks, am a robot.," *IEEE Spectrum,* vol. 45, no. 6, pp. 68-71, 2008.

[5] F. Nietzsche, Der Antichrist/De Antichrist, 3rd ed., Amsterdam: Uitgeverij de Arbeiderspers, 1888/2006, p. 21.

[6] L. P. Kaelbling, M. L. Littman and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research,* vol. 4, pp. 237-285, 1996.

[7] C. S. Webster, "Alan Turing's unorganized machines and artificial neural networks: his remarkable early work and future possibilities," *Evolutionary Intelligence,* vol. 5, no. 1, pp. 35-43, 2012.

[8] S. Schaal, "Is imitation learning the route to humanoid robots?," *Trends in cognitive sciences,* vol. 3, no. 6, pp. 233-242, 1999.

[9] Y. Duan, M. Andrychowiscz, B. Stadie, O. J. Ho, J. Schneider, I. Sutskever and W. e. a. Zaremba, "One-shot imitation learning," *Advances in neural information processing systems,* pp. 1087-1098, 2017.

[10] OpenAI, "Open AI's new machine learning system," Youtube, 18 May 2017. [Online]. Available: https://www.youtube.com/watch?v=b_wG80BTj-w. [Accessed 17 June 2019].

[11] D. Hassabis, "Artificial Intelligence and the Future," 11 March 2016. [Online]. Available: https://www.youtube.com/watch?v=8Z2eLTSCuBk. [Accessed 16 June 2019].

[12] R. P. N. Rao and A. N. Meltzoff, "Imitation Learning in Infants and Robots: Towards Probabilistic Computational Models," in *Proc. AISB 2003 Convention: Cognition in Machines and Animals*, Aberystwyth, UK, 2003.

[13] A. M. Turing, "Computing Machinery and Intelligence," *Mind,* vol. 59, no. 236, pp. 433-460, 1950.

[14] Y. Pan, "Heading toward artificial intelligence 2.0," *Engineering,* vol. 2, no. 4, pp. 409-413, 2016.

[15] K. J. Holyoak, “Parallel distributed processing: explorations in the microstructure of cognition,” *Science,* vol. 236, no. 1987, pp. 992-997, 1987.

[16] V. Minh, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg and D. Hassabis, “Human-level control through deep reinforcement learning,” *Nature,* vol. 518, no. 7540, pp. 529-533, 2015.

[17] A. Nair, B. McGrew, M. Andrychowicz, W. Zaremba and P. Abbeel, “Overcoming exploration in reinforcement learning with demonstrations.,” *Proceedings of IEEE International Conference on Robotics (ICRA),* pp. 6292-6299, 2018.

[18] D. Hassabis, “Artificial Intelligence and The Future,” 22 November 2018. [Online]. Available: https://www.youtube.com/watch?v=zYII3AOSgo8. [Accessed 4 July 2019].

[19] F. Woergoetter and B. Porr, “Reinforcement Learning,” *Scholarpedia,* vol. 3, no. 3, p. 1448, 2008.

[20] Google DeepMind, “Methods and Apparatus for Reinforcement Learning, US Patent #20150100530A1,” 2015. [Online]. [Accessed 16 June 2019].

[21] J. Kober and J. Peters, “Imitation and Reinforcement learning,” *IEEE Robotics and Automation Magazine,* vol. 17, no. 2, pp. 55-62, 2010.

[22] B. R. Duffy and G. Joue, “Intelligent robots: The question of embodiment.,” in *Proceedings of the Brain-Machine Workshop*, 2000.

[23] F. Kaplan, “Neurorobotics: an experimental science of embodiment,” *Frontiers in Neuroscience,* vol. 2, no. 1, pp. 22-23, 2008.

[24] R. A. Brooks, “New approaches to robotics.,” *Science,* vol. 253, no. 5025, pp. 1227-1232, 1991.

[25] D. Dunning, “The Dunning-Kruger Effect: On Being Ignorant of One's Own Ignorance,” in *Advances in Experimental Social Psychology*, New York, Elsevier Inc., 2011, pp. 259-262.

[26] A. D. Meltzoff, P. K. Kuhl, J. Movellan, T. J. Sejnowski and ', “Foundations for a new science of learning,” *Science,* vol. 325, no. 5938, pp. 284-288, 2009.

[27] D. Kiela, L. Bulat, A. L. Vero and S. Clark, “Virtual embodiment: A scalable long-term strategy for artificial intelligence research.,” in *NIPS Workshop in Machine Intelligence*, 2016.

[28] Y. Zhu, Z. Wang, J. Merel, A. Rusu, T. Erez, S. Cabi, ... and N. Heess, “Reinforcement and imitation learning for diverse visuomotor skills,” in *Computing Research Repository*, 2018.

[29] D. Pathak, P. Mahmoudieh, G. Luo, P. Agrawal, D. Chen, Y. Shentu and T. Darrell, “Zero-shot visual imitation,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,* pp. 2050-2053, 2018.

[30] C. Thurau, C. Bauckhage and G. Sagerer, “Imitation learning at all levels of game-AI,” in *Proceedings of the international conference on computer games, artificial intelligence, design and education*, 2004.

[31] S. Ross, G. Gordon and D. Bagnell, “A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning.,” in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, 2011.

[32] A. Nair, D. Chen, P. Agrawal, P. Isola, P. Abbeel, J. Malik and S. Levine, “Combining self-supervised learning and imitation learning for vision-based rope manipulation,” in *ICRA*, 2017.

[33] T. Hester, V. Matej, O. Pietquin, M. Lanctot, T. Schaul, B. Piot, D. Horgan, J. Quan, A. Sendonaris, I. Osband, G. Dulac-Arnold, J. Agapiou, J. Z. Leibo and A. Gruslys, “Deep Q-Learning from Demonstrations,” in *32nd AAAI Conference on Artificial Intelligence*, 2018.

[34] J. Hui, “RL - Reinforcement Learning Algorithms Quick Overview,” Medium, 29 August 2018. [Online]. Available: https://medium.com/@jonathan_hui/rl-reinforcement-learning-algorithms-quick-overview-6bf69736694d. [Accessed 20 June 2019].

[35] A. Young-joon, “One Giant Step for a Chess-Playing Machine,” 26 December 2018. [Online]. Available: https://www.nytimes.com/2018/12/26/science/chess-artificial-intelligence.html. [Accessed 4 July 2019].

[36] I. Sample, “'It's able to create knowledge itself': Google unveils AI that learns on its own,” 18 October 2017. [Online]. Available: https://www.theguardian.com/science/2017/oct/18/its-able-to-create-knowledge-itself-google-unveils-ai-learns-all-on-its-own. [Accessed 4 July 2019].

[37] A. Krumins, “Artificial General Intelligence Is Here, and Impala Is Its Name,” 21 August 2018. [Online]. Available: https://www.extremetech.com/extreme/275768-artificial-general-intelligence-is-here-and-impala-is-its-name. [Accessed 4 July 2019].

[38] G. Di Pellegrino, L. Fadiga, L. Fogassi, V. Gallese and G. Rizzolatti, “Understanding motor events: a neurophysiological study,” *Experimental Brain Research,* vol. 91, pp. 176-180, 1992.

[39] Y. Demiris and A. Meltzoff, “The robot in the crib: A developmental analysis of imitation skills in infants and robots,” *Infant and Child development: An International Journal of Research and Practice,* vol. 17, no. 1, pp. 43-53, 2008.

[40] A. P. Shon, J. J. Storz, A. N. Meltzoff and R. P. Rao, “A cognitive model of imitative development in humans and machines,” *International Journal of Humanoid Robotics,* vol. 4, no. 2, pp. 387-406, 2007.

[41] A. N. Meltzoff and M. K. Moore, “Newborn infants imitate adult facial gestures,” *Child development,* pp. 702-709, 1983.

[42] G. Gergely, H. Bekkering and I. Kiraly, “Developmental Psychology: Rational imitation in prevarbal infants,” *Nature,* vol. 415, no. 6873, p. 755, 2002.

[43] A. C. Horowitz, “Do humans ape? Or do apes human? Imitation and intention in humans (Homo Sapiens) and other animals.,” *Journal of Comparative Psychology,* vol. 117, no. 3, pp. 325-336, 2003.

[44] L. M. Hopper, “Deferred imitation in children and apes,” *Psychologist,* vol. 23, no. 4, pp. 294-297, 2010.

[45] L. M. Oberman, J. A. Pineda and V. S. Ramachandran, “The human mirror neuron system: a link between action observation and social skills,” *Social Cognitive and Affective Neuroscience,* vol. 2, no. 1, pp. 62-66, 2007.

[46] F. Foroni and G. R. Semin, “Language that puts you in touch with your bodily feelings: The multimodal reponsiveness of affective expressions,” *Psychological Science,* vol. 20, no. 8, pp. 974-980, 2009.

[47] V. S. Ramachandran, *Mirror Neurons and Imitation Learning as the driving force behind "the great leap forward" in human evolution,* 2000.

[48] V. Gallese, “The 'Shared Manifold' hypothesis,” *Journal of Conscious Studies,* vol. 8, pp. 33-50, 2001.