A Three-Component Cognitive Theory

Author:
Francisco J. Arjonilla
(pronounced [pãko̩ arxo̩nĩʎa])
f.j.arjonillagarcia@students.uu.nl
pacoarjonilla@yahoo.es

Main supervisor:
John-Jules Ch. Meyer
Co-supervisors:
Stefan van der Stigchel
Carlos Sacristán

Master's Thesis
submitted in fulfillment of the
requirements for the degree of
Master of Science
in the programme
Artificial Intelligence
at the
Faculty of Natural Sciences
Universiteit Utrecht

Utrecht, 22nd of April, 2015

Abstract

Artificial Intelligence and robotics are developing at a fast pace. However, implementation of general intelligence in computers remains elusive. Currently, there are no technologies able to solve unexpected challenges nor learning algorithms that outperform their initial requirements. This thesis takes an unconventional multidisciplinary perspective on cognitive systems to propose a universal theory of cognition that can be identified in systems exhibiting intelligent behaviour. I present a theory that is based on a formalization of trial and error that is composed of three cognitive components and has the capacity to produce intelligent behaviours. In absence of models, tentative trials to reach a specific goal are inevitable until one succeeds. Each attempt at fulfilling a goal is termed a variant. The cognitive components are: a substrate, a generator of variants and a selector. (1) The substrate holds the information that shapes each variant and it may be either physical or immaterial in nature. It is linked to environmental mechanisms that interpret the instructions conveyed by each variant. (2) A cognitive generator provides the heuristics to produce variants, and (3) the selector chooses amongst the generated variants which one is the most adequate for the pursued goal. Then, I argue that there are three families of intelligent-behaving systems that give evidence to the theory. Firstly, I propose a biologically-based cognition that relies on principles of evolutionary theory. Particularly, I give an alternative interpretation of Darwinism that diverts away from the traditional notion of evolution by chance and credits biological evolution with cognitive capabilities. Secondly, I identify a mapping between the theory and the latest advances in neuroscience and experimental psychology. Specifically, attention drives selection and variants are represented in neural modules. Thirdly, I explore methods in artificial intelligence and I justify their cognitive limitations. I discuss a comparison of these cognitive families where more analogies are drawn, including a description of a putative sequence of cognitive emergences. Finally, I deduce from the theory a novel cognitive architecture that does not rely on preconstructed models to interact with the environment.

Keywords: *Cognition, evolution, neuroscience, artificial intelligence.*

Contents

List of Figures

CHAPTER 1

Introduction

Since the birth of Artificial Intelligence, researchers have striven to make machines that learn like humans do. The diversity of proposed solutions can be categorized in two paradigms: symbolic processing and connectionism (Shi, 2011). The former assumes that symbols and reasoning constitute the fundamental building blocks of intelligence, while the latter tries to mimic the human brain by distributing knowledge in a vast array of interconnected elements. Neither of them have achieved general learning capabilities. Indeed, no single source of true intelligence has been accounted for except the human brain. Engineers and computer scientists struggle to create autonomous intelligent machines that divert us from those jobs that hinder people's self-transcendence and in that manner stand sustainable, wealthier lifestyles. Common methods include bottom-up approaches in which explicitly computer-programmed models deal with particular problems usually with some degree of learning, which can be reduced to sophisticated parameter adjusting. An alternative top-down approach is to study cognition from a theoretical perspective and apply that knowledge to particular problems. This thesis belongs to the latter approach.

It is reasonable to think that there may be some theory that encompasses all systems exhibiting some degree of intelligence, and the implementation of that general theory of cognition is adapted to each system. In this respect, computers follow strict-rule-following algorithms relying on mathematics and logics, while neural circuits are better suited to operate with ambiguous information and nevertheless yield suboptimal but adequate outputs. So, any cognitive theory attempting to encompass both kinds of intelligence should set aside these differences.

Furthermore, it is unduly anthropocentric to insist that any cognition must work at speeds resembling those of the human brain. I find no sensible reason to conclude that the thinking speed of cognitive systems should accommodate those of the human brain to be considered as intelligent as humans, or to design algorithms that aim at that speed. In contrast, the efficiency of a cognitive system refers to the speed of finding solutions to novel problems. The power of a cognitive system and specifically the capability to display ingenious, creative solutions to exposed challenges should be measured with time independence.

Behaviours produced by humans are tentatively constantly being mimicked in machines, so much that the most accepted test to compare machine intelligence to that of a human is the accuracy of a machine to mimic humans in natural written conversation (Turing, 1950). The brain is a very complex system that evolved after thousands of millions of years and resulted in an intricate net of neurons with a diffuse distribution of entangled functions specialized in survivability in a natural environment. Remains to define which of those functions are necessary for cognition, and which can be discarded. For instance, early blind subjects experience function replacement in the neurons of the visual cortex without cognitive function diminishment (Röder et al., 2002), in a similar way that visual stimuli take over activation of auditory cortex in early deaf subjects (Finney et al., 2003), suggesting that alternative functions can take over unused primary sensory areas with no significant cognitive impairment and consequently are inessential for cognition.

1.1. Problem description

Science has not been able to explain human intelligence yet. Nor how does intelligence emerge in artificial systems. Traditionally, a number of human thinking processes have been assumed essential for cognition. These include planning, reasoning, induction, language, and emotion, amongst others. However, they are based on introspection and the assumption that there is only one way to achieve intelligence. Since there are no other known cognitive systems, it is difficult to refute these assumptions. But it may be possible that those abilities are only cognitive enhancements to underlying processes, and that the essential mechanisms of universal cognitions are hidden by these abilities. It is not known if there is a common framework followed by every system that exhibits intelligent behaviours. Actually, biological evolution may be considered such a system. Evolution has been studied for centuries and its mechanisms are well known. The fact that we can explain evolution does not impede it to create biological traits that are more efficient, more complex and better suited than any technology developed by humans. This thesis claims, amongst other things, that evolution is in fact a cognitive system, and its similarities and differences with human cognitive processes can be pinpointed under a common theory whose main characteristics might be found in all intelligent-behaving systems.

It is possible that cognitive theories confined to the human brain and computers may not be enough to explain all kinds of intelligent behaviour. Maybe we should consider as well any kind of system that possesses intentionality and complex behaviour. An ultimate theory of cognition should explain all sorts of cognitive systems, ranging from the human brain to machine cognitions, but also other cognitions yet to be known. Science needs a common framework to explain all intelligent behaviours, such as human actions and in less degree machines, in order to advance in machine learning. As I will show, molecular biology qualifies as a system that processes information by manipulating organic molecules in an intentional and complex manner and cannot be left out.

Moreover, there is no common framework to determine what constitutes a cognitive system. Different fields of science use the term cognition in an incompatible way to other fields, i.e. psychological cognitive skills are quite different from cognitive processes attributed to control engineering and computer science, and have certainly dissimilar power and extent. Science fails to give indications as to what are the minimum requirements to qualify as a cognitive system.

It would be interesting to acknowledge if it is possible to lie the theoretical grounds for a *universal cognitive system*, in the same way that computation machines that are Turing compatible can be compared, run within each other and translated between each other.

1.2. Cognition: An ambiguous term

Cognition is a term that nowadays is used ambiguously among different fields, mainly neuroscience and psychology, on one hand, and artificial intelligence and robotics on the other. The ambiguity comes from the difficulty of defining the terms used to define cognition.

In neuroscience and psychology, cognition refers to *the set of processes that allow humans and many other animals to perceive external stimuli, to extract key information and hold it in memory, and ultimately to generate thoughts and actions that help reach desired goals* (Purves et al., 2013, p. 2). Cognition is also described occasionally as the processes that the *mind* carries out, but many other problems related to terminology and philosophy may be introduced with that description. In fact, due to the difficulty of defining the true nature of cognition, it remains a topic of hot debate in philosophy. In artificial intelligence and control engineering, according to Ikeuchi (2014), *a cognitive system is an autonomous system that can perceive its environment, learn from experience, anticipate the outcome of events, act to pursue goals, and adapt to changing circumstances.* Legg and Hutter (2007) wrote an extensive review of definitions and tests of human and machine intelligence.

The terms *biological cognition* and *biological intelligence* are commonly used to refer to human intelligence, but they can produce misunderstandings under the framework envisioned herein. For that reason, the term *biological intelligence* will be restricted to the cognitive processes attributed to Darwinism, whilst the term *rational cognition* will make reference to human reasoning processes.

In this thesis the term cognition will refer to *those systems that possess abilities and processes that enable them to learn to interact in new ways with the environment and ultimately to accomplish its goals.* Automatic processes that are unable to learn new abilities are left out of this definition.

1.3. Research goal

The rationale for this thesis is to infer a general theory of cognition that can be applied to machines. With that goal, I will explore how all intelligent-behaving systems can be related. The implications of these links will be explored and directions for interpreting cognition given. This work is purely theoretical with no experimentation involved. I will make use of previous scientific work in several fields to justify and provide evidence for the theoretical proposals. The research question can be summarized as:

What are the basic building blocks of any cognition, how do they combine to materialize in the implementation of each specific cognitive system, and what qualifies as a cognitive system?

To answer these questions, a theoretical model of general intelligence will be proposed. Then, it will be mapped to evolution, the human brain and methods in artificial intelligence. A novel cognitive architecture suited for cognitive robotics will be deduced from the principles of the theory.

1.4. Outline

Chapter 1, the current one, introduces the problem that this thesis aims at solving and gives a conceptual overview of the contents that will be treated in the rest of the document. Chapter 2 briefly reviews the state of the art related to cognition in the multiple fields that the topic has been discussed in order to introduce the unacquainted reader to the latest advancements and theories. It also cites some identifications of intelligent behaviours that have been largely neglected in cognitive science. Chapter 3 defines the core of the theory, describes in detail each of the three components and introduces some entailments such as the importance of adequate heuristics. Chapters 4, 5 and 6 tentatively interpret the theory over different systems, namely biological evolution, human rationality and methods of artificial intelligence. These systems possess the capability of interacting and manipulating the world in a directed manner. The interpretation of the theory over these 3 families should give indications to interpret and acknowledge potential cognitive systems that are not covered in this work. Chapter 7 reviews some similarities and differences between the three cognitions, including an account on the temporal characteristics of the appearance of the three cognitive families described in Chapters 4-6. Finally, the last chapter summarizes the thesis, pinpoints the strengths and weaknesses of the core theory and proposes ways to design smarter machines by indicating new research directions in cognitive architectures.

CHAPTER 2

Background

Intelligence has been historically a topic of major research. There have been attempts to describe intelligence in the fields of psychology, neuroscience, philosophy, systems engineering, computer science and artificial intelligence. No field has arrived to conclusive solutions or effective models, though, hence the difficulty of implementing true intelligence in computers. Many cognitive models have been proposed, but proving correctness has not been possible, for two main reasons. First, technological and ethical limitations on the study of the brain, and second, technological models of cognition that have been implemented on computers fail to surpass the narrow applications that they are devised for. This chapter presents some of the multidisciplinary knowledge that motivated this thesis, in addition to accounts on intelligence other than the human brain and computers.

2.1. Artificial Intelligence

Two main paradigms drive research in Artificial Intelligence. GOFAI (Good Old Fashioned Artificial Intelligence) aims efforts on symbolic-like processing but faces the symbol grounding problem as a major drawback. Alternatively, the newer connectionist or subsymbolic approach processes information in a decentralized manner, with neural networks as its major representative (Amoretti and Reggiani, 2010). Neither of them show signs of ever reaching the main goals of Artificial Intelligence. Much effort, time and money has been put into research to develop both paradigms, but they resist to acquire general intelligence. Troublesomely, both paradigms encompass almost every model that has been proposed.

The term Artificial General Intelligence was coined by Gubrud (1997) to refer to the kind of artificial intelligence that surpasses human abilities and is usable in essentially any operation requiring intellectual abilities where a human is otherwise needed. When and how will Artificial General Intelligence be achieved is unknown, or if it will be achieved at all. Some of the proposed models have been more successful than others, but all of them are confined to narrow tasks and abilities. Turing (1950) proposed a test that has been broadly accepted by the scientific community to test human-like intelligence. It consist of a human judge that attempts to distinguish a computer from a person by interacting with them simultaneously by means of a chat interface. *Narrow intelligence*, conversely, refers to computer intelligence devoted to specialized tasks that do not require general learning capabilities.

Sandberg (2013) explores theoretical models of the dynamics of a hypothetical emergence of a superior artificial intelligence, happening that has been termed *technological singularity*. He focuses on the growth aspect of emergences and analyses the effect on dynamics of technologies that can access and improve their own cognitive processes. Interestingly, he also suggests and reviews some analogies between these models and biologically evolutionary processes.

2.1.1. Cognitive architectures

Cognitive architectures are generally extensively inspired in the human brain, in spite of the lack of knowledge about it (Langley, 2006). Some authors have proposed characteristics that cognitive architectures should possess to be considered *cognitive* along with typical challenges that should be able to solve (Langley et al., 2009). However, the proposed requirements resemble very much traditional skills typically ascribed to human cognition, without questioning the validity of this link.

Sammut (2012) makes strong criticisms against current robot cognitive architectures. His main point is that the biggest drawback is the inflexibility of the proposed solutions. He distinguishes between reactive and deliberative control. The former is most used in robotics in the shape of planners that quickly convert sensor measurements to actuator commands whilst the deliberative control is more flexible but generally slower. He questions the need for symbolic representations and limits its utility to *creating specialised skills in offline training, prior to actual missions.* For that reason, he concludes, reactive systems are more useful to respond rapidly in dynamic environments, but symbolic representations will find more applications with growing task complexity. I will quickly digress by adding that he also reflects on the relation between progress in robotics research and evolution and concluded that both are *somewhat analogous.*

The next few paragraphs briefly introduce some of the most successful cognitive architectures.

SOAR architecture

The SOAR Architecture (State, Operator And Result) was born as a model of the theories of cognition defended by Newell (1992). The core hypothesis is that operators can be applied on the input of a problem to be solved and on states that represent knowledge to obtain a result that fulfils a desired outcome. Chunking and goal splitting are used as mechanisms for goal management. Although the ultimate goal for Soar is to achieve general intelligence, there is no claim and no evidence that this goal has already been reached.

IBM deepQA

In the field of natural language processing, *IBM Watson* is one of the most advanced frameworks. It was developed under the IBM deepQA pipeline with the purpose of competing against 2 champions at the game of Jeopardy!™, where participants receive clues in the form of answers and the response must be given in the form of questions. IBM Watson beat two other human contestants in 2011, proving that computers can outperform humans in responding to facts phrased in natural language (Ferrucci, 2012). The IBM deepQA framework takes a set of candidate answers and ranks them according to a model produced in previous training sessions that take relationships between concepts in the clues and concepts in a knowledge base as parameters. The ranking model is produced by training and testing different classifier methods, and the most successful one was selected for the contest (Gondek et al., 2012). It can be noticed that there are two loops that produce variation and selection of the best answer. The first one occurs during training of the system and involves selection of a number of ranking methods based on classifiers. The second loop occurs during the contest itself where several candidate answers are produced and the highest-ranked answer is selected.

LIDA

More recently, Franklin et al. (2014) have proposed a cognitive architecture that deals with the functional organization of cognitive processes. It takes ideas from cognitive science, neuroscience and psychology into a theoretical framework that has been partially implemented, and not yet fully tested. Some of the hypothesized cognitive properties that it can handle are: learning, decision making, action selection, feelings and emotions, working memory, and perceptual symbol systems. The emphasis that they give to cognitive cycles stands out. They hypothesize that a cognitive system consists of a continual sequence of cognitive sequences composed of three phases: understanding, attending and action selection, similarly to other cognitive architectures that also implement cognitive cycles.

2.2. Theories of Intelligence

There is no general agreement on the exact nature of human intelligence amongst psychologists and neuroscientists. There has been considerable debate over this topic and cognitive research usually focus on specific cognitive (as defined in psychology) functions.

During the history of psychology, many theories have been proposed to account for intelligence. Spearman (1904) proposed a concept he called general intelligence, or the *g factor*. It was one of the earliest attempts at measuring intelligence. By use of mental aptitude tests, Spearman could numerically express aptitude scores on a variety of narrow tasks. He found that subjects that performed well at one cognitive test tended to score well on other tests as well. Thurstone (1938) presented a theory of intelligence that departed from the view of intelligence as a single ability. According to his theory, the *primary mental abilities* are seven: verbal comprehension, reasoning, perceptual speed, numerical ability, word fluency, associative memory, and spatial visualization. Moving forward in time, the theory of multiple intelligences that Gardner (1983) presented rejected the idea of measuring human intelligence as a set of test scores because they do not depict accurately people's abilities. He also described intelligence as composed of seven different abilities: Visual-spatial, verbal-linguistic, bodily-kinesthetic, logical-mathematical, interpersonal, musical, and intrapersonal intelligences, in contrast to a single general ability of earlier models. Lastly, the Triarchic Theory of Intelligence was suggested by Sternberg (1985) and defines *successful intelligence* as comprised of three factors: Analytical intelligence, creative intelligence and practical intelligence. This theory defended that intelligence should be measured by the ability of an individual to adapt successfully to their environment throughout their lifespan.

This very brief review of theories of human intelligence models should suffice to realize that scientific theories on intelligence are focused on measuring intelligence and describing it from an anthropocentric point of view. There are no explanatory links between consciousness/cognition and neural activity, and no indications on how it can be replicated in machines.

Cognitive theories of consciousness

More recently, neuroscience is giving more promising results in the form of cognitive theories of consciousness. Baars (1988) proposed the Global Workspace Theory (GWT). This theory explains how access consciousness functions. According to Baars, a *global workspace* holds the contents of consciousness, which are broadcast to separate functional modules in the brain. He makes an analogy of his theory with a *theater of consciousness*, where a *spotlight of selective attention* marks the contents that a *central executive* selects. These contents become available to other cognitive processes. Other contents and processes remain in the dark, i.e. in the unconscious. The Global Workspace Theory has received a lot of attention, but it remains yet as an interpretation of evidence in neuroscience.

Another important theory is the Multiple Drafts Theory by Dennett (1991), who rejects the analogy of consciousness as a "theatre". He suggests that collections of sensory information, called drafts, activate conscious experiencing for updating and revising cognitive processes. The experience of consciousness as being *on-line* is an illusion. Dennett recurs to the analogy of a manuscript under constant editorial revision, elicited by sensory information entering the nervous system.

Furthermore, Tononi (2004) proposed an integrated information theory (IIT) based on the measure of integration in a system given by the so-called function Φ. IIT is built on two axiomatic pillars. The first axiom is based in the differentiation of conscious experiences and mental states. The second axiom claims that these experiences are integrated and cannot be experienced independently. Later revisions have refined the theory and added more axioms.

Cognitive theories of consciousness do not attempt to model a general theory of intelligence. On the contrary, they are focused on explaining human intelligence, consciousness, phenomenology and other features that are specific to human thinking processes. It is debatable whether human thinking is the only method for general intelligence.

2.3. Two stage models

Philosophy of free will and sometimes biology, have been described as conforming to two-stage models. In general, the two stages are approximately outlined by an initial random stage that provides different courses of action followed by another stage that selects the desired courses of action, or alternatively discards the undesired ones with similar results. Two views are introduced here: the first one has to do with philosophy of mind and tries to explain human free willed behaviours whilst the second one links evolutionary change to generation and selection of variants, suggesting that there is some connection between evolution, free will and cognition.

Two stage models in human free will

Many theories composed of two components have emerged in the history of philosophy of the mind. Typically, two stage models on the topic of human free will defend that the first stage needs of an indeterministic, partially random generation of ideas, thoughts, etc. and the second stage operates a deterministic selection based on internal desires, goals and drives.

James (1897) was attributed to be the first philosopher to give a two stage model of free will. In his manuscript, he suggested that choice needs of a previous production of differing hypothesis. Although not explicitly stated, he laid the initial sketches of what has become one of the most debated arguments and a reference position in free will, with many other models forking from his. Mixing religion, philosophy, poetry and free will in a single manuscript, his model starts by conceiving viable alternative future*s* for the *free*, followed by the will, which chooses amongst the available futures and removes chance from the first stage.

LeShan and Margenau (1983, p. 240) suggested that freedom, the essential feature of human consciousness, *involves two components: chance (existence of a genuine set of alternatives) and choice*. In their view, chance is generated by quantum processes and choice is accomplished by the mind. Again, it is possible to see a trend where thinking processes generate a set of possible choices and later select the one that best fits the goal at hand.

Two stage models in biology

Mayr (1988) described evolution as a two stage process almost a century after James' first two-stage model:

> Evolutionary change in every generation is a two-step process: the production of generically unique new individuals and the selection of the progenitors of the next generation. (Mayr, 1988, chap. 9, para. 6)

It is interesting to see how Mayr's two stage process integrates well with two-stage models of free will. In both free will and evolution, according to James for human free will and Mayr for biological evolution, the first stage involves generating alternatives with some random factor, and the second stage narrows the alternatives given in the first stage so as to follow the desires and goals of the system. The relation suggests that there might be a link between human free will and evolution. Arguably, were both systems to be indeed related by two-stage models, evolution could be credited with free will.

Heisenberg (2009) has most recently given a view on animal free will. He claimed that even unicellular organisms have a random stage followed by intentional behaviours that can be clearly seen in even unicellular organisms that move towards higher concentrations of nutrients, despite apparent random walks (Codling et al., 2008). In higher animals, trial and error behaviour can also be identified while they constantly explore the environment, continuously testing and discarding different alternatives for action. It is arguable whether these behaviours are sufficiently grounded for crediting these organisms with cognitive skills.

To sum up, many authors have given two stage models throughout history to explain free will, intentionality and intelligence in humans, animals and evolution. This relation suggests that there is a profound link between all of these systems when it comes to explain intelligent behaviours.

2.4. Intelligence in evolution

There is a minority view in the scientific community that evolution is intelligent by itself, at least of a weak sort, i.e. it appears to be intelligent. Certainly, the complexity of life may suggests that living organisms are a product of some intelligent agent. Intelligence in evolution should not be confused with the apparent intelligence that some organisms display, specially swarm organisms like ants and bees. Rather, it involves the adaptations of those organisms to deal with challenges never before confronted, and thus have no previous adaptations to deal with them. The philosopher Hume (1779) wrote an appealing argument on behalf of the Greek philosopher Cleanthes, named the *Argument from Design* (and then criticized it):

> Look around the world: contemplate the whole and every part of it: you will find it to be nothing but one great machine, subdivided into an infinite number of lesser machines, which again admit of subdivisions to a degree beyond what human senses and faculties can trace and explain. All these various machines, and even their most minute parts, are adjusted to each other with an accuracy which ravished into admiration all men who have ever contemplated them. The curious adapting of means to ends, throughout all nature, resembles, exactly, though it much exceeds, the productions of human contrivance, of human design, thought, wisdom, and intelligence. Since therefore the effects resemble each other, we are led to infer, by all the rules of analogy, that the causes also resemble, and that the Author of Nature is somewhat similar to the mind of man, though possessed of much larger faculties, proportioned to the grandeur of the work which he has executed. By this argument a posteriori, and by this argument alone, do we prove at once the existence of a Deity and his similarity to human mind and intelligence.

In chapter 4 I will admit to the intelligent nature of biological designs and I will claim that this *Author of Nature* exists and is, in fact, the process of life itself in agreement with evolutionary biology.

2.4.1. Cancer

As a first example of intelligent behaviour in evolution I will mention cancer. Cancer cells have the property of rapidly evolving into a system that can successfully confront the destruction attempts of the host immune system and medical treatments. Eukaryote cells in complex organisms are subject to Darwinian evolution once the mechanisms that control unbounded cell growth are lost. Thereon, tumour cells become their own entity whose primordial goals consist in the consumption of all available resources to survive and reproduce. Adaptability and change can occur exceptionally fast, showing in short periods the characteristics of evolution that take generations to become discernible in complex organisms (Greaves and Maley, 2012).

The growth and adaptability of cancer cells have been described as possessing a collective emergent intelligence that displays teleonomic behaviour. The difficulty of counteracting the strategies that cancer cells take have lead some authors to describe them as possessing *swarm intelligence*, able to withstand the most severe and advanced cancer treatments (Tarabichi et al., 2013). Fortunately, tumour populations cannot communicate with cells in other hosts and need to rediscover survival paths with each appearance.

2.4.2. Orgel's rules

Orgel (1973) wrote from a very interesting point of view how the creation of life emerged from primordial biochemical processes. Throughout his book he suggested what has become known informally as the two Rules of Orgel (Dunitz and Joyce, 2013):

1. Whenever a spontaneous process is too slow or too inefficient, a protein will evolve to speed it up or make it more efficient.

2. Evolution is cleverer than you are.

The second rule is most important for the claims in this thesis. Not only does it suggest that evolution is a process that shows signs of intelligence, refuting that chance alone drives evolution, but he astonishingly claims that human intelligence has a superior competitor. The rules were not stated directly in his book, but were rather inferred from passages such as the following:

> Once one gets used to the idea of natural selection, one finds it helpful in thinking about the development of many systems other than living organisms. One should not underestimate the importance of trial and error in the development of technology, for example. The development of the revolver is an object lesson in evolution. Those clumsy guns with revolving barrels are the dinosaurs; there were many small successes and many great failures on the way to the Peacemaker, and for every company that presently makes revolvers, there must be ten that have been eliminated. (Orgel, 1973, p. 183)
>
> [...]
>
> There is, thus, an exception to the rule that objects of high information content must be the direct product of natural selection; they may be the products of human ingenuity. This exception does not weaken the argument, since the intelligent "creators" in this case

> are themselves the products of natural selection. (Orgel, 1973, p. 196–197)

Hence, Orgel links human cognition with evolution, and gives an example of human inventions that have been designed with a similar process to trial and error that is ubiquitous in evolution. Selection processes can be glimpsed in his example of revolver-making companies.

2.4.3. Fogel's claims

Not only Orgel argues for evolution as an intelligent agent. More recently Fogel (2006) compared the characteristics of learning processes in evolution and intelligent systems:

> The theory of evolution is the great unifying theory of biology. But the impact of the theory extends even further: Evolution serves as a unifying description of all intelligent processes. Whether learning is accomplished by a species, an individual, or a social group, every intelligent system adopts a functionally equivalent process of reproduction, variation, competition, and selection. Each such system possesses a unit of mutability for generating new behaviors and a reservoir for storing knowledge [...]. Learning is accomplished through some form of random search and the retention of those "ideas" that provide the learning system with the greatest understanding of its environment. [...] The learning system evolves, adapting its behavior to achieve its goals in a range of environments.
>
> [...]
>
> Species, individuals, and social groups rely on the same underlying physics to generate adaptive behavior despite using various genetic, neuronal, or cultural mechanisms, just as birds, mammals, and fish rely on the same physics for achieving flight even though the specific mechanisms employed are vastly different. Every system that incorporates the evolutionary processes of reproduction, variation, competition, and selection, by whatever means, inherently evokes inductive learning. Intelligence is not the end result of evolution; rather, it is woven into the process itself. Intelligence and evolution cannot be separated. (Fogel, 2006, Section 6.1)

2.4.4. Summary of intelligence in evolution

There are clear cues that Darwinian evolution should be endowed with intelligence. Indeed, there are biological devices that closely resemble the products of human ingenuity. Such is the case of mechanical gears in biology (Burrows and Sutton, 2013), screws and blades (DeRosier, 1998; van de Kamp et al., 2011), explosions (Raghuveer et al., 1993) and jets (Jackson et al., 1972). In the end, there is no reason to conclude that evolution does not have the cognitive power to design any system that can be designed by humans as long as it is given enough time and there is selective pressure towards the feature, in the same way that we, humans, will only spend time and effort on things that we find valuable. With so much evidence that Darwinian processes exhibit intelligent behaviours, it is natural to ask ourselves if evolution carries the key to understanding intelligence and cognitive processes. If that was the case, Darwin (1859) might have not only discovered the origin of the species, but would have inadvertently started the science of what might become a description of how intelligence works in its simplest form.

CHAPTER 3

THE THEORY

There is a need to formally describe what a cognitive system is, how it can be identified in known systems and how to distinguish it from non-cognitive systems. The following theory sets the grounds to discuss the extent that systems with some kind of intentionality possess true intelligence. It will allow to compare different cognitive systems by analysing the characteristics of each system in relation to the theory and pinpoint which cognitive systems are stronger in what areas, and how they can be improved. Any physical or immaterial system with apparent intentionality is in principle susceptible to be analysed with this theory and consequently very different systems displaying intelligent intentional behaviour can be linked with each other, even if their decision processes are very different in nature.

Before digging further into details, it is convenient to define the term *variant*. A cognitive system seeks fulfilling goals and, for that purpose, a range of possible alternative solutions are generated to be selected upon, according to this theory. The solutions are stored and expressed by the substrate in the target environment, where assessment of each solution takes place. Each of the alternative solutions is termed a cognitive variant, or *variant* in short. The reason that it is necessary to assess each novel solution is that there is no previous knowledge about which solutions are the most adequate. If that knowledge existed, then an automatism that makes use of the knowledge base can be created without the need for a cognition.

In this chapter I will present in detail the core of the theory. This abstract theoretical development demands further characterization for interpretation of real-world cognitions that will be delineated in the following chapters. Three basic components constitute the core of the theory:

1. The first one refers to the **informational substrate** of the other two components below. The substrate is a reservoir for storing knowledge. It can be physical or immaterial in nature and it must be possible for other components to make modifications to the variants. Also, it must be possible to assess the variants held by the substrate either by direct variant evaluation or by evaluation of the effects of the variants on the environment. In the latter case, decoding mechanisms that interpret the variants produce these effects.

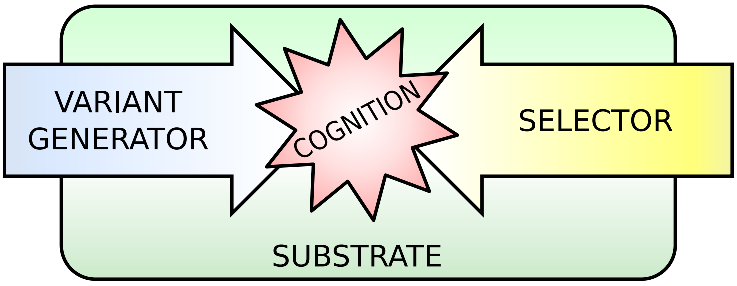

Figure 3.1

Cognition emerges from adjoining the three constituent components. The substrate stores each variant, which gets decoded in the target environment in intelligent behaviour. This graphical disposition will be used to depict the cognitive components throughout the thesis.

2. A **variant generator** comprises the second component and provides a range of different variants as alternative solutions for the problem at hand. The quality of these variants, i.e. the heuristics, is for the most part what determines the efficiency of a cognition. Variant generators are not constrained to any method, medium or location. Rather, a generator refers to the process of creating alternative variants from previously generated ones, random variants, or any other method as long as the structure of the substrate is not destroyed, but only the information that the substrate holds gets modified.

3. The last component, the **variant selector,** selects the most appropriate solution amongst those generated in the variant generator. It works by assessing some feature of the outcome of a variant and checking its compliance to the goal at hand. Variant selection occurs after variants are generated, but it may happen that some variants are filtered even before reaching the selector. Sophisticated cognitive systems may discard invalid variants in early stages even before testing them out in the target environment by using models of the predicted outcome of the variants, whilst simpler cognitions will assess every variant against the target environment.

3.1. The three components

A cognitive system needs, in one form or another, the three components just mentioned. If a system lacks any of them, there is no possibility to seek better variants in the space of possible solutions. For example, if there is no substrate, there is no way to store variant information or to actually perform any action that will lead to the fulfilment of the requested goal. Conversely, a lack of variant generators will result in no investigation of new solutions and the system would get stuck with the variants already present in the system. Finally, absence of a selector removes intentionality from the system and so, convergence to the desired goal does not occur. Each of the three components introduced above deserves a more elaborate description. Here it goes, in the same order.

3.1.1. First component: The substrate

The first factor to take in consideration in a cognitive system is the substrate, that can be physical or immaterial in nature. The substrate refers to the matter that holds elements of information that make up each variant and the mechanisms that decode that information.

In general, the substrate is determined by a discrete set of elements that can be combined to form increasingly complex variants and it must allow for the variant generator to provide a rich number of combinations. In other words, the substrate must be productive, which means that there is an infinite number of possible variants that can be generated. Were it to be finite, a knowledge base with all variants could be generated and assessed against the goal, selecting the best variant and reducing the goal to an optimization problem. Consequently any method running on that set of finite possibilities could be at most a good optimization algorithm with severely limited learning capabilities. A direct way of devising an infinite substrate is allowing for an unbounded variable-length sequence of elements that belong to a finite alphabet.

The power of the substrate lies in its potential to interact with the world. This interaction occurs by means of decoding mechanisms that produce a specific behaviour as a result of the interpretation of a variant. For instance, a sequence of bits stored in a hard drive has little capacity to produce any effect on the environment because there are no mechanisms that decode these bits. But if these bits are transferred to the memory in a computer, the processor can interpret the data as if it were a program. The system produces an effect on the environment by the combination of the information in the variants and the interpretation of this information.

The whole set of possible outcomes of the variant generator shapes the search space. The number of possible variants is, in principle, infinite, but it is also possible to have a finite search space that is infinite in practical terms, i.e. the search space is too big to fully explore it in a reasonable amount of time. Assuming the search space was finite, once it is scoured all over and the generation of new variants exhausted, the cognition would not be able to improve variants, learn from experience, etc. Any alternative variant would have been already tested and valued inferior. In the example above, the paper and the computer memory have the same search space, albeit the sheet of paper does not have any effect on the environment by itself.

3.1.1.1. Decoding mechanisms and expressiveness

It becomes necessary to express the difference between a substrate and its decoding mechanisms. Decoding mechanisms take a variant at the input and activate a chain reaction in the target environment. They are essential to produce intelligent behaviour in the environment. Decoding mechanisms can have different origins. For instance, they can be naturally present in the environment or they can be devised by a cognitive system. Expressiveness of the substrate refers to the capacity of a substrate to exploit those decoding mechanisms. Highly expressive substrates may have huge effects on the environment if the substrate sequence, or variant, adequately exploits the automatic mechanism that decodes the sequence. Thus, the potential of a cognition to control its environment depends on the expressiveness of the substrate. I will illustrate this with an example. Consider a person whose goal is to reach as many readers as possible with an idea they have written down. The person acts as a cognitive system and the letters in the book shapes the substrate. If the book is hidden deep in a library it might not be read by anyone and would have no effect on its environment, whilst the same idea may change the mindset of many people if it is well sold. Readers of the book serve as decoding mechanisms for the latter book, while the former has no readers and hence lack decoding mechanisms. Throughout this thesis, I will deal with simpler decoding mechanisms that do not involve interaction between cognitive systems, except to acknowledge the presence of this interaction.

Given a goal, several cognitive systems may provide different variants that fulfil it. As long as there are no additional requirements, these variants are all valid and may be used indistinctly. However, decodification of these variants will be each different and they will reach the goal in different ways. The secondary effects, which are those effects on the environment that are not considered for goal assessment, can vary significantly amongst variants. If the variants are not designed to minimize secondary effects, i.e. by appending a secondary goal to every goal definition such as *minimize secondary effects*, further cognitive processing may encounter that secondary effects of different variants interfere with each other's primary effects. From a goal-fulfilment point of view, variant interference is not important, but it may hinder the development of improved versions of those variants because all variants that interfere with each other will have to be modified to adapt to the changes in the effect of one of them. Consequently, a cognitive system that produces variants with few secondary effects will be more efficient. So, it is desirable to generate variants with few secondary effects, but the cognitive processes required to arrive at these variants may not be accessible to simple cognitions. That is the case of cognitions that explore the search space of the substrate randomly: it might be too expensive to explore all of it to search for variants with fewer secondary effects.

3.1.2. Second component: The variant generator

The second component of this three-component cognitive theory is the variant generator. A cognitive system needs access to alternative variants to decide which one is selected in order to follow its goals. Moreover, they need to be generated in a format that can be processed and understood by the rest of the components. This is the task of the variant generator. It will create alternative instances of actions, objects, thoughts or whatever entity shaped by the expressiveness of the cognitive system in the format compatible with its substrate, whether it is a biochemical exchange of ions as in neurons, electronic movement in electric fields as in transistors, order of nucleotides in a DNA strand, position of gears in a mechanical computing device or any other system capable of storing and processing information.

A non-comprehensive list of possible sources for generating variants follows:

- Random.
- Small random modifications to previous successful variants.
- Hard-wired from birth.

- Transformation of input data.
- Imitation of other similar agents.
- Observation, with or without simultaneous manipulation of the environment.

The simplest source of variants, which does not make use of previous successful ones, is randomly picking a variant in the search space, but then it can be argued whether the system is cognitive at all because it is not learning from previous experience. However, in absence of any kind of previous knowledge, random variants are the only available source of variants. Once the first variants that improve the fitness of the cognitive system are found, they can be evolved and improved. More complex sources of variants use information from the environment to shape the variants and then test them for accuracy and fitness. According to this taxonomy, methods that are hard-wired from birth, methods that modify previous variants and reuse them for different goals constitute endogenous sources, whilst imitation, observation and transformation of input data belong to exogenous sources because they make use of input data. Exogenous sources are not really necessary but may improve the efficiency of the generator.

With respect to when it is more appropriate to generate variants, those that accomplish the goal with no further requirements demand no improvements, so it would be a better use of resources if generation of variants were kept to a minimum until unknown challenges arise in the environment. This is possible if the pursued goal is not of the kind that asks for forever better improvements. Limited search spaces also have a global maximum that, once found, render the cognition useless as any other variant generated would be of equal or lesser value than the one corresponding to the global maximum. Consequently, operation of a cognitive system leads to learning by means of discovery of better variants, or else the features that cognitive systems provide are not really needed since the tasks can be accomplished by existing variants.

Note that there has been almost no reference to system input and output until now. The cognitive system deals with variants whose effects on decoding mechanisms may or may not operate with inputs, outputs, sensors, variables, memory, etc. What the cognitive system does is manage the variants that instruct the decoder to use the data, rather than directly transforming input data to output data. Some variants will only need to output data to reach a goal, others will need some information from the environment before outputting any data, and yet others may just operate on internal variables of the decoding mechanism.

Finally, not all generated variants should be directed towards a specific goal involving some change of state in the environment. Improvements in the way variants are generated may lead to even better and faster generation of variants. For that purpose, the algorithms that generate variants should be self-modifiable by the cognition and will result in a more efficient cognition without external intervention. Consequently, a source of variants in a generator can evolve as it finds new ways of exploring the search space, become unrecognisable in comparison with the initial method and even develop more advanced sources of variants that were initially unavailable. Even if this feedback loop is not strictly necessary in a cognitive system, it increases significantly the potential efficiency of

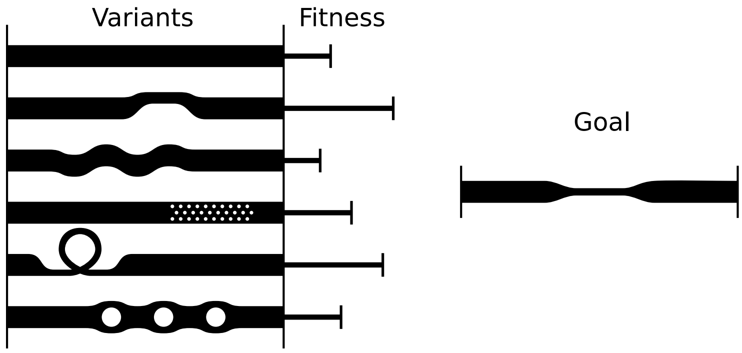

Figure 3.2

Diagrammatic depiction of a hypothetical cognition that produces lines of different shapes. On the left side a set of variants are valued against the goal on the right. The hypothetical evaluated fitness measure is represented as a line with a stop ending next to each variant. The goal is directly assessed on the variants, hence no variant expressiveness is necessary. The search space is represented by the infinite set of all the possible lines that can be drawn.

the cognitive system.

3.1.3. Third component: The selector

The selector is the last of the three components that allegedly make up every cognitive system. Selection discriminates between variants that fulfil goals and variants that do not in the target environment. Whilst the generator might have heuristics that generate variants towards the desired goal, it is the ultimate task of the selector to remove all variants that do not comply with the goal in the target environment. This way, even if heuristics are suboptimal and biased towards other goals, the cognitive system will still yield those variants that converge to the desired goal. It is necessary to narrow the outcome of heuristics since perfect variant generation cannot be expected or otherwise the intelligence would be found in the heuristic algorithm instead of the cognitive system as a whole.

Once variants are generated, testing takes over. For the sake of simplicity I will assume for now that goals are externally set and fixed. Testing involves decoding any one variant and assessing the compliance of its expression with the goal. Assessment of compliance implies the necessity for an assessment method that evaluates how closely the variant matches the goal. Assessment methods can be as simple as evaluating a variant to conform with some rules, but also as complex as evaluating the effects that the variants and associated decoder have in the environment, even if these effects are localized far away from the variant and much later in time. The assessment methods take the variant or the effects of the variant and produce a decision about how well the variant fulfils the goal. These methods can be composed of automatic processes or an additional recursive cognition. It is possible that a cognitive system devises its own assessment methods in cognitions that can represent them as variants, resulting in self-reference in the selector: The selector evaluates the variants that it will use to evaluate other variants. The *goal* in a cognitive system is more precisely defined as the effect in the environment that the assessment method accepts. The consequence of this definition is that goals are not explicitly defined in selectors, but are rather imbued in the assessment method as implicit knowledge.

Once a variant is successful at fulfilling the goal set by the selector, there is no need to generate other variants. Actually, it is counterproductive because the new variants might not fulfil the goal. Consequently, The cognitive system does not need to operate any more once the goals have been fulfilled. However, goal fulfilment should be supervised in non-static environments because there is no guarantee that the variant will remain valid. Once supervisory methods detect a failure to fulfil the goal, the generator is enabled again to produce more variants. Also, the goal can require continuous improvements of the variants, which result in continuous operation of the cognitive system.

3.2. Combining the components

Several entailments to the combination of the three components described above can be drawn. Here I describe four significant aspects that will need consideration when analysing cognitions. The first one is related to the quality of variants generated and analyses the importance but dispensableness of good heuristics. The next one discusses the theoretical possibility of embedding an independent cognitive system in the generator of another cognitive system. The third entailment explores improvements to the variant generator by storing the outcome of the assessment of variants. Lastly, cognitive cycles are discussed.

3.2.1. Heuristics

Heuristics play a mayor role in the efficiency of a cognitive system when it comes to generation of variants. The heuristics of a cognitive system are defined as the methods that create variants that lead to improved goal fulfilment under available resources. In a productive substrate, it is undoable to generate all possible variants in a search space at once, hence the need for quality heuristics. Of course, better variant generation will yield improved cognition and hence faster goal accomplishment.

I will first discuss two radically opposing heuristics. To start with, random generated variants are indeed poor heuristics. New variants are produced completely at random and variants are not recycled at all. It is the most inefficient method to generate new variants, albeit the simplest one. Exploration of the search space for goal-fulfilling variants takes a considerable amount of time and any satisfactory variant is generated by pure chance. It would take a tedious long time to reach even the simplest goal-fulfilling variants. On the other hand, generation of just one variant in perfect variant generators suffice to reach the best solution for the goal at hand. With only one variant, goal selection is redundant and consideration of the cognitive system as such becomes questionable. In these ill-defined heuristics, the true cognitive power lies in the system that created the variant generator, which shows extraordinary skills at arriving immediately at good variants. The intelligence attributed to variant generation must not be greater than the cognitive system as a whole, or otherwise it would make no sense to have a cognitive system that relies on more intelligent methods than it can produce, since the cognitive theory would rely on other, unexplained cognitive systems.

A good heuristic strategy may consist on evolving variants that represent the best solution to the goal at hand. Exploration of the search space for optimum variants is resource consuming, so it makes sense to recycle variants obtained during previous attempts to generate better solutions. A major problem of this strategy is that solutions may lie in local optima and heuristics may be incapable of escaping them. Consequently, the cognition would fixate the variants generated around a suboptimal variant. Proper cognitive heuristics should implement mechanisms to generate variants that escape local maximums. This difficulty has been tackled by Leijnen (2014), who suggests that creativity emerges in self-organizing processes by loosening the constraints that successful variants pose on the development of better solutions by adding random factors.

3.2.2. Recursion

I have just analysed some heuristics, but more complex heuristics are also possible. Heuristics can house another cognition since heuristics other than random generation of variants involve some degree of intelligence. Figure 3.3 depicts a double cognitive system with one level of recursion in the generator of the top-level cognition.

It is possible to rely on variant testing in a simulated environment to improve the heuristics of the top-level cognition. Simulation of variant achievement prior to assessment in the target environment may improve efficiency, but may instead hinder the generation of successful variants if the model that is used for the simulated substrate is inaccurate. A sandbox can be very useful when the cognition is operating in environments where there are few opportunities to test variants, although there are some risks associated. The simulated environment should possess an accurate model of the final environment, including decoding mechanisms of the substrate. It should also be supervised for errors and inaccuracies, updating it when inconsistencies are found. Reduction of resource usage in the final cognition might also prove advantageous in environments with unlimited resources.

3.2.3. Goal models

A question that might naturally show up is if it is possible to speed up selection of variants by predicting the outcome of selection before variants are tested and assessed. In order to avoid repetition of failed patterns, it would be a good idea if the heuristics could avoid them even before they go to the selection stage. In the simplest case where there is a unique unchanging goal, it would suffice to mark failed variants that have already been tested to avoid testing them again. One way to do this is to link each variant with assessment of the goal fulfilment for later retrieval. If the cognition is confronted against many goals, it would also need to store the goals that each variant fulfils so that variants that fail at a certain goal are not discarded at testing against other goals. The goal model of a cognitive system stores information about what variants are successful at what goals.

Goal models should still be tested for accuracy in case there are unexpected

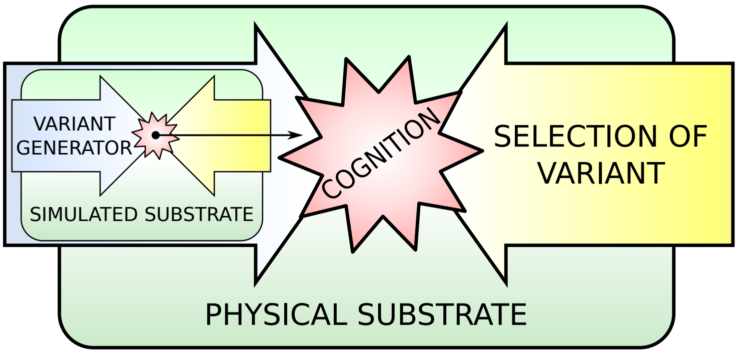

Figure 3.3

The heuristics of a cognition can be produced by the outcome of an embedded cognition in a simulated environment. The first cognition constitutes the generator of the second cognition.

changes in the environment. A variant generator with goal models may then improve its heuristics by discarding variants that do not match the goals even before reaching the selection component of the cognition, avoiding the usage of previously tested variants. Consequently, variants only require testing in cognitions with goal models if they have been generated for the first time, or if new goals are requested. Nevertheless, changes in environment conditions or in pursued goals may render a previously discarded variant the fittest under the new conditions.

A taxonomy of cognitive systems based on usage of goal models follows:

1. *Goal-model-based cognitions:* These systems evaluate the likelihood that a certain variant will accomplish the goals and select amongst the possible variants in its knowledge base the one that best fits the goal, if any is available. Also, testing of previously failed variants is avoided.

2. *Non goal-model-based cognitions*: All generated variants need testing and assessment in the target environment. Failed variants are not remembered so they are tested repeatedly, posing a huge drawback to the efficiency of the cognition.

An example of a goal-model-based cognitive architecture is depicted in figure 3.5. Note that the figure does not intend to constraint cognitive architectures to the one shown, but it just illustrates a cognitive architecture that follows the model of a goal-model-based cognitive system. Goals are used differently in each component. The assessment method of the selector implicitly defines the goal, whilst the generator uses the goal specification to model which variants may have an opportunity to be selected.

3.2.4. Cognitive cycles

Behaviourally, the three-component cognitive theory can be regarded as a system that first produces a set of alternative actions to take (or whatever the substrate expresses), followed by a reductionist phase that narrows the number

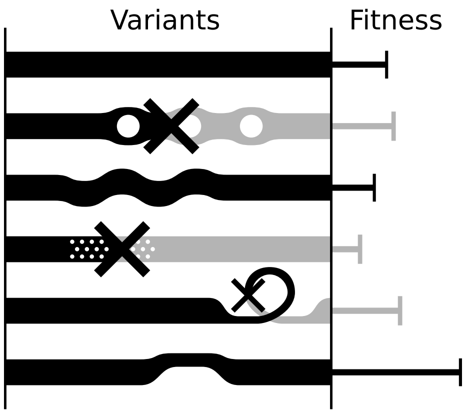

Figure 3.4

Diagrammatic depiction of lines with different shapes produced by an hypothetical goal-model-based variant generator. Crossed-out variants can be filtered before the selection stage. The goal model predicts (prediction in grey) that those variants will receive a low fitness value and aborts them before they are assessed and discarded.

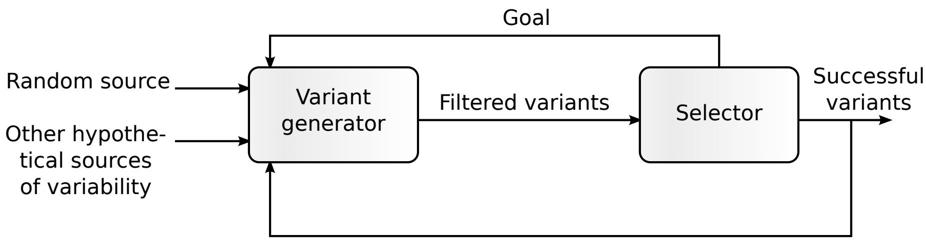

Figure 3.5

An example of a goal-model-based cognitive architecture. The variant generator uses the goal specification to filter out variants that are known not to work even before the actual testing of variants takes place.

of actions to the successful ones. Then, this process is repeated, hopefully improving the variants in each iteration. At this point it is easy to discern what can be considered a cognitive cycle. An ideal cognitive cycle consists of generation of variants, assessment of compliance to the goal and selection of the best variants in an iterative loop. Here is a sketch of a canonical cognitive cycle:

1. Generate variants. Variants are realized in the substrate.
2. Express variants in the target environment through decoding mechanisms linked to the substrate.
3. Measure some effect of each variant in the target environment.
4. Select variants that produce measures that conform to the desired goals.

Variations are of course possible, such as generation of only one variant in each cognitive cycle. The process is repeated with each iteration, perhaps using the fittest variant of the previous cycle as seed for the variant generator in the new cycle, although discarded variants may also appear in newer cycles if the cognitive system is not goal-model-based. Goal selection necessarily involves discrete steps because assessment is performed on the expression of each variant, or directly on the variant in simple cognitions. It would be inconvenient to assess a variant if it is modified during its evaluation. Figure 3.6 depicts variants in three subsequent cycles of the line generator example.

Exceptions to this scheme can be the rule: variants may not need to be generated all at once. Sometimes a cognitive cycle does not need to finish before the next one starts. Variants can be generated on overlapping cognitive cycles. In this case, the generator and the selector work continuously, as opposed to the generation of a batch of variants and selection of the best ones before a new batch of variants is generated. In any case, the duration of each cognitive cycle can be measured in order to compare cognitive systems: The duration of a cognitive cycle is the average duration that it takes for a variant to be generated and assessed.

3.3. Summary of the chapter

In this chapter I have proposed a theory of cognition that constitutes the core of this thesis. I postulated that every cognitive system is composed of three components. The first one, the substrate, shapes the informational framework where the other two components will exert its actions. The substrate stores variants of solutions to the problem at hand and is linked to mechanisms that decode and express the variants in the target environment. The second component, the variant generator, is composed of automatic mechanisms that generate variants. The specific method is dependent on the cognition. If variants are filtered with the use of on goal models before being assessed, the efficiency of the cognition can improve significantly. Finally, a selector, which implicitly defines the goal, selects which variants comply with the goal and discards the ones that do not comply. Afterwards, some entailments of the theory were presented, such as the effect of heuristics in the variant generator, how cognitions can be used as components of other cognitions, and the emergence of cognitive cycles as an inevitable consequence of iterative loops of generation and discard of variants.

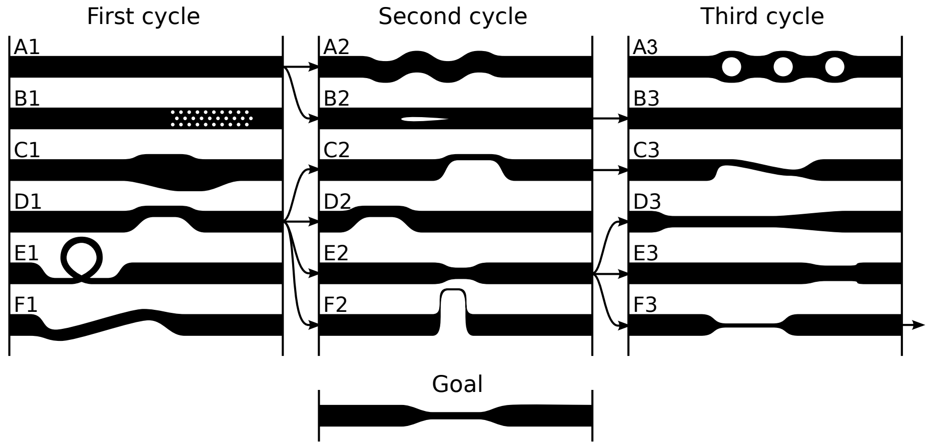

Figure 3.6

Variants that have been generated on three sequential and ideal cognitive cycles are depicted here. Variants on the second cycle are generated from the most successful variants in the previous cycle, in this case A1 and D1. The most successful variants are in turn used to generate variants on the next cycle, except for A3 which is generated randomly. Generally, after each cycle the variants are more successful. In this example, F3 has evolved from E2 and D1 to become the most successful variant after three cycles.

In some of the figures in this chapter, a series of line-with-curves variants have been illustrated. If the final goal were to produce an easy to manufacture expanding valve, such as those used in chemical engineering, the cognitive system would still be able to produce it. Neither the reader nor the cognitive system needed information about the thermodynamic properties of fluids coming into the tube and leaving the tube, which could also represent data input/output in a computer. Only the outcome from the assessment was needed. What I want to indicate with this example is that this theory does not describe what is done with data or material that goes in and out of the system. Rather, it describes how methods that manipulate data or material may be generated in any cognitive system.

To sum up, three elements need to be simultaneously taken into account to create a new cognitive system, all of them essential:

1. A productive substrate that stores and decodes variants.

2. Methods that generate variants.

3. Methods that test and assess variants in the target environment against the desired goal.

In the following chapters, I will apply the theory in real-world systems that have proved to behave intelligently, namely evolution and humans. Artificial systems will also be analysed due to the potential that they have of becoming truly intelligent, although they do not conform with the theory yet. The components defined in this chapter will be identified in each cognition and a common framework linking all kinds of known intelligence will thus be laid.

CHAPTER 4

The Darwinian Cognition

Any human engineered device in the inanimate world of machines and computers is overshadowed by even the simplest organic beings, which can create intricate molecular mechanisms that are not fully understood yet. It is disconcerting to affirm that evolution is not intelligent when human ingenuity cannot match the complexities that evolution has brought forth, despite taking millions of years of evolution. This thesis proposes a mechanism where biological entities are the iterative end product of a cognitive process. Evolution may be thought of as a cognitive system that overcomes the difficulties it confronts in nature in broader periods of time by recombination and mutation. It is based on the provision of variability on possible genetic outcomes followed by selection of the fittest sequences by natural selection, and is physically supported by large biomolecular chains. From now on evolution will be referred as a cognitive system, but it is throughout the chapter that I will justify how it relates to the cognitive theory presented in the previous chapter. I will also assume that the reader has a basic understanding of molecular biology and population genetics. Any introductory book such as (Clark, 2012) suffices to follow the argumentation. The concepts referred to in this chapter are well grounded topics in biology, so there is no need in general to recur to the state of the art.

A link between the theory and several kinds of biologically cognitive activity will be presented, denoted each as a Darwinian cognition and all together grouped in a family of cognitions closely related. The term *Darwinian* used throughout this thesis refers to the most widely accepted interpretation of evolution, that is, neo-Darwinism and the modern evolutionary synthesis. The ascription of evolutionary life forms to the three component cognitive theory will be justified and their subtle differences exposed. Several examples on how to interpret evolution will be detailed, paving the way to ascribing all forms of life to the theory. Even more, viruses may be argued as cognitive systems whose machinery for variant generation and expression is borrowed from the cognitive machinery of other lifeforms.

The reason for starting with this family of postulated cognitions is that it could have been the first cognition to appear on Earth, and most importantly, it emerged spontaneously in nature without help from any other intelligent entity (more details in section 7.5), hence the interest in describing it first. The reason for being the simplest cognition and why it could emerge spontaneously is likely

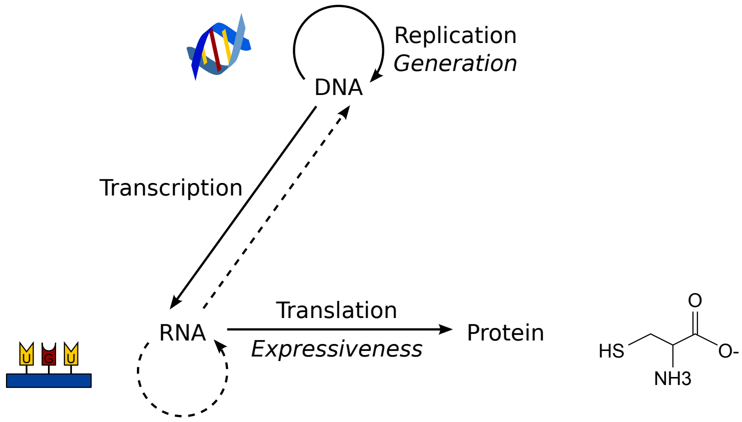

Figure 4.1

The central dogma in molecular biology. Solid lines indicate information flow from DNA to RNA to protein. These molecules are composed of unbounded sequences of nucleobases or aminoacids and correspond to the substrate in the Darwinian cognition. Aminoacid conversion from the sequence TGT (UGU) is shown, which corresponds to cysteine, the author's favourite aminoacid. Dotted lines indicate occasional information flow. Text in italics refer broadly to cognitive concepts.

that two of its components, namely the variant generator and the selector, are disembodied and occur naturally. By disembodied I mean that there is no physical matter that we can point out explicitly as the executors of the components. Rather, immaterial forces act on the cognitive substrate and follow coincidentally the functionality of the both the generator and the selector. Other complex forms of life rely on molecular mechanisms to partially implement the variant generator. More features that make this cognition so simple is that there is no simulation of the environment, no implementation of goal models and no recursion.

The first thing I want to point out in relation to cognitive processing is that it is an incredibly slow process that has been running for thousands of millions of years. In comparison, human cognition is not that patient. We have a limited life and any endeavour requires effort and some of this limited time. On the contrary, evolution runs with no deadlines.

The central dogma in molecular biology

I give now a brief overview of molecular biology before discussing the relation between evolution and the three component cognitive theory. The central dogma of biology (Crick, 1970) specifies possible flows of information between RNA molecules, DNA molecules and proteins. DNA and RNA molecules hold sequences of four nucleotides that store the genetic information of living organisms. DNA and RNA molecules replicate by complementarity of nucleobases. During replication, the original DNA strand is used as a template where nucleotides pair by complementarity of adenine with thiamine or guanine with cytosine. Other enzymes, specifically DNA polymerase, assemble the nucleotides into a new DNA strand. Sometimes, the complementarity of nucleobases does not perfectly match the original copy and a mutated copy is assembled instead.

A RNA copy can be created from a DNA strand if RNA polymerase is used instead of DNA polymerase in a process called transcription. RNA does not have the chemical stability of DNA, but it is capable of enzymatic activity. Sequences of triplets of nucleobases in RNA strands can be translated to sequences of aminoacids to synthesize proteins, which can produce much more varied enzymatic activity than RNA strands. Proteins cannot be converted back to RNA or DNA.

The leading theory that explains the emergence and evolution of the central dogma in molecular biology is the RNA world (Gilbert, 1986; Copley et al., 2007). The RNA world hypothesizes that in the beginning of life, the only molecules that existed with capacity to self-replicate were RNA strands. These molecules evolved to synthesize proteins and store their genetic information in DNA. The next sections focus on the relation between the cognitive components, RNA and DNA molecules, proteins and evolution.

4.1. Identification of cognitive components

In this section each of the cognitive components is identified in the building blocks of life. Small differences in the components across members of the family of Darwinian cognitions will be further elucidated later on in the chapter. Darwinian cognitive components are characterized by:

- Processes that modify the offspring's genetic sequence constitute the variant generator. Each individual in a species is a variant and contains the genetic sequence that interacts with the environment by gene expression.
- The selector corresponds to Natural Selection. Goals are slightly modified depending on the specific system.
- RNA and DNA strands store the genetic information of each individual in a sequence of nucleobases and consequently correspond to the cognitive substrate.

Each variant in the Darwinian cognition is comprised of the sequence of nucleobases in a RNA or DNA strand. Therefore, each individual in a species holds a variant and incarnates its expression. This claim will be justified in the following sections, since species can be very different in nature, i.e. from encapsulation of simple RNA strands in viruses to complex organisms with a DNA copy in each cell of an individual.

Cognitive cycles

Cognitive cycles in the Darwinian cognition can be hard to pinpoint, but describing them now may help to understand how the cognitive components are identified. The cognitive cycle spans from the time that a individual of the species reproduces to the time that its offspring reproduces again, matching one generation of the species. In this period, a new variant is generated, assessed and selected, which biologically corresponds to birth, growth and mate. However, individuals in a population may be at different stages of the reproduction cycle. Yet, the whole population should still be considered a single cognitive system, since the phase of reproduction cycles may synchronize.

Other species might have clearly demarcated cognitive cycles. For example, a great number of species of insects die in winter and their descendants survive in eggs that hatch on spring of the next year. This is a clear example of an ideal cognitive cycle, which features generation of variants in batches with a duration of one year and where all individuals share the reproduction stage at all times.

4.1.1. The variant generator

A variant generator is devoted to creation of new sequences of substrate elements. In the Darwinian cognition, this happens when a new genetic variant in the form of an individual is produced. For the most part, two mechanisms are involved in this process: mutation and recombination. In these mechanisms, chance plays an important role. It is involved in the generation of variants by randomly replacing elements of recycled variants in the case of mutation, or by setting the translocation point in the case of recombination.

Indeed, exogenous sources such as mutational radiation can randomly mutate RNA or DNA sequences and affect their nucleobase sequence (Pullman and Pullman, 1963) and structure (Alexander et al., 1960). However, randomness and natural selection cannot be the ultimate source of variants in a complex cognition like Darwinism because the search space of RNA and DNA chains is so big that the chances of occurring functionally effective proteins are vanishingly small (Maynard Smith, 1970). According to the three component theory, the process can be enhanced by improved heuristics, such as endogenous sources of mutation that regulate the effect of other mutational factors (Sniegowski et al., 2000) and recombinant processes, either by sexual reproduction in eukaryotes, or transduction, transformation and conjugation in prokaryotes. All in all, the variant generator is a combination of external sources that were present in the environment before the emergence of the cognition, and internal sources that were self-devised. In any case, the generator alters the sequence of RNA or DNA nucleobases of individuals in a species at conception.

4.1.2. Natural Selection

Natural selection assesses the suitability of each individual for survival and reproduction. It constitutes the cognitive selector, and manifests every time a bacteria multiplies, a predator hunts a prey or pollen fertilizes a germ cell, for example. Each cognitive variant can take anything from minutes to decades to be evaluated and selected. This time corresponds from the moment an individual is conceived to the moment it generates offspring. It may be tempting to suggest that the selection component in the Darwinian cognition is given by a fitness function integrated in natural selection, but natural selection is not a fitness function in a computational sense, nor has the Darwinian cognition a fitness function. Mathematical models may describe with good precision the behaviour and evolution of species, but describing the selector component in the Darwinian cognition requires departing from models because the assessment has to be done in the target environment.

In this cognition, the selector has no direct method to discard variants. Rather, unfit variants are unable to leave offspring before they die. With no offspring, the variant is neglected from further cognitive processing and the variant information is extinguished without cognitive intervention.

Survival is an inevitable goal

Every cognition has a goal, but that goal has to be assigned somehow. Assuming no external intervention in the Darwinian cognition, the goal assigned to the selector needs to emerge spontaneously from nature. I will explicitly argue why fitness for survival is the goal for Natural Selection and not any other for the sake of clarity.

Assume firstly that a system has a goal with a deadline. The goal can either be accomplished or not after the deadline expires, but afterwards the system does not have any use. Such a system would have a limited lifetime and then disappear from existence by, for example, ageing, decomposing and disintegrating, whether it has fulfilled the goal or not. Now assume that the goal for such system has no deadlines. Then, it is necessary that the system is provided with mechanisms that ensure its continuity so that the goal can be indefinitely pursued. Survivability becomes an important subgoal for the system. Afterwards the environment has to be taken into account. If the resources available to the system are finite, then the system will strive to use them as much as possible in order to keep the primary goal. Inevitably, the system that gives preference to the primary goal will die as a consequence of not allocating enough resources to survivability, whilst the system that postpones the primary goal survives. In the end, survivability and reproduction becomes a primary goal under resource restriction.

Research on molecular biology has proven that it is possible to manipulate artificially some biological processes. Specifically, the goals of the Darwinian cognition can easily be disrupted by interfering with the selector. Imposing artificial selection on evolution in a controlled environment leaves the Darwinian cognition under control of the human disrupter (for example, Lassner and Bedbrook, 2001; Turner, 2009). In nature, the goals may suffer slight variations depending on the complexity of the organism. In the RNA world, survivability and reproducibility of molecules were of uttermost importance, whilst in unicellular lifeforms, molecules can be discarded in favour of the cell. In the same way, complex life gives preference to whole organisms.

4.1.3. The substrate

Many different kinds of molecules have the potential to become substrates. Specifically, long chains of organic molecules can become a substrate: they are productive and the sequence of the constituting elements can be manipulated. Proteins certainly conform, since they are composed of an unbounded sequence of aminoacids in any order. RNA and DNA are sequences of nucleobases, too. RNA and DNA can also be composed of an unbounded number of elements in any order. However, it is only RNA and DNA molecules that are enabled to function as substrate because molecular mechanisms that support them have emerged and evolved to do so. I will illustrate this claim with the description of a protein-based system that could be argued as cognitive but lacks a variant generator because the substrate does not support any.

4.1.3.1. *Prions: A failed substrate*

I will start the discussion on biologically-based cognitive systems with a group of molecules that do not fully comply with the requirements of a cognitive system. In particular, I will argue that proteins lack some of the features of cognitive systems claimed in chapter 3, although they do constitute a productive substrate. Specifically, prions are infectious agents that are made up of misfolded proteins and can induce other similar proteins to denaturalise into the malign form. Often this self-replicable property of prions leads to diseases by uncontrolled growth of the population of faulty molecules. The Creutzfeldt–Jakob disease is one of the most notable diseases (Prusiner, 1998). The following argument may be used as an example of how to interpret apparent cognitive systems that are not really such.

The substrate of prions is constituted by a sequence of aminoacids that defines how is the protein going to fold to form the secondary, tertiary and quaternary structure. The native state of a protein is the folded conformation, or tertiary structure, that gives the intended final functionality to the protein. If a protein does not fold correctly, it might lead to diseases due to the inability of the protein to perform its function.

Prions feed themselves with the correct isoform of the protein and force a conformational change on the latter to the malign form (Surewicz and Apostol, 2011), providing instances of the same prion that are subject to natural selection. Arguably, the selection component of a cognition takes place when a prion replicates, otherwise at some point in time it will decompose with no copies to take over. However, the variant generator does not appear in prions, which always share the same primary structure (Gibbons and Hunter, 1967). The sequence of aminoacids in prions does not change during the lifetime of the prion or its replication. Changes to the sequence of aminoacids involve segmentation of the sequence, replacement of a single aminoacid and reassembly. There is no molecular machinery associated to prions that allow the production of variants on primary structures of prions. Another method for generating variants on a protein would be to generate different secondary and tertiary structures, that is, additional folding states. However if this were the case, the search space would be strongly limited. As a consequence, there is no adaptation to new problems; it just blindly replicates itself in a fully automatic fashion until the reserves of the healthy protein are exhausted.

To sum up, prions do have a selection component, in the context of the three-component cognitive theory, which is reproduction, similarly as living beings. They do have as well a productive substrate, i.e. the aminoacid sequence. But prions are incapable to generate variants on the aminoacid sequence due to the lack of mechanisms that manipulate the substrate. Consequently, prions are made of a substrate that is not prone to generation of variants. RNA, as will be immediately discussed, does have an appropriate substrate and its simplicity could have been the key feature that kick-started the route towards more complex cognitive forms.

4.2. The family of Darwinian Cognitions

Three cognitive instances of the Darwinian Cognition will be analysed. The first one, namely the RNA world, focuses on the simplest known system that can be attributed with cognitive capabilities. The other two cognitions enhanced the cognitive capabilities with synthesis of proteins in unicellular organisms and complex life.

4.2.1. RNA world: Ribozymes, The Minimal System

The RNA world, first proposed by Gilbert (1986), is the current leading scientific paradigm to explain the origin of life. In essence, it claims that the first stages of life on earth occurred with molecules of RNA doubling as transmitters of genetic information and enzymatic catalysts, with the genetic code and protein synthesis playing no part in replication of RNA strands, since they had not developed yet (Joyce, 1989). Earlier non-RNA-based lifeforms could have existed (Yarus, 2011), but they have not been described as well as the RNA world and RNA is also found in current life. I will argue that the RNA world poses the oldest known system with cognitive capabilities since it had very primitive forms of the three components needed for cognition.

The following introduction to the RNA world portrays the characteristics that are relevant for the thesis.

Brief description of the RNA world

It is generally accepted that the RNA world was the precursor of all kinds of life on earth, perhaps around four billion years ago. There is not much evidence on whether this was the first form of life, since there are many arguments that favour that many other molecules, now extinguished, could have been the precursor of the RNA world, all of them capable of forming unbounded sequences (Robertson and Joyce, 2012). Most of the chemical reactions involved in the spontaneous emergence of the first RNA strands have been identified and many experiments have been performed that give evidence to the plausibility that evolution, and hence life, went through an RNA-world stage. A multitude of literature can be consulted for more detailed descriptions of the RNA world and how RNA is involved in today's biochemical reactions, such as (de Duve, 2005), (Gesteland, 2005) and (Atkins et al., 2010). Given the relative simplicity of RNA compared to current biological processes, it is easier to analyse it first to understand the connection between evolution and cognition.

RNA chains with enzymatic activity are known as ribozymes. Orgel (1986) gave evidence of how ribozymes can contingently catalyse a variety of reactions involving the substrates that RNA is built upon. It is now known that it can fully replicate itself by no other means than the interaction of its nucleobases with itself, and that its evolution is directed by Darwinian-like processes dominated by master copies that degenerate into a diversity of variations, mostly resulting in catastrophic effects for the molecule (Eigen and Schuster, 1977; Szathmáry, 2006). Indeed, minimal RNA molecules have been designed under *in*

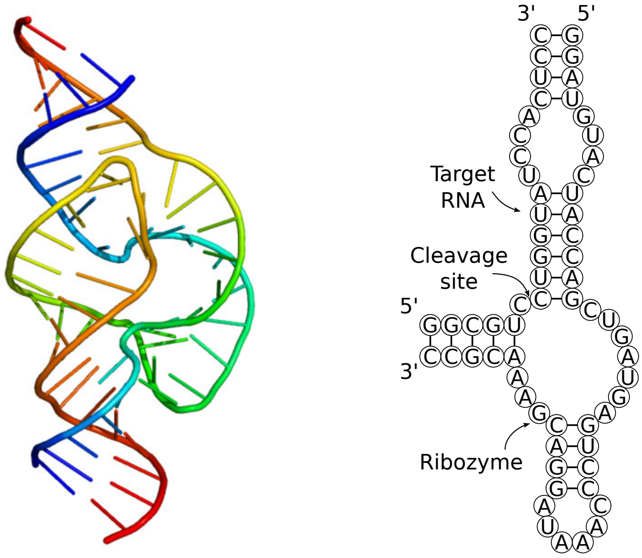

Figure 4.2

The hammerhead ribozyme (long strand) functions as cleavage of another RNA molecule (short strand) (Martick and Scott, 2006). Left: Spatial three-dimensional structure. Retrieved from http://upload.wikimedia.org/wikipedia/commons/2/28/Full_length_hammerhead_ribozyme.png on 21-12-2014. Creative Commons BY-SA 3.0 License. Attribution: Wgscott. Right: Secondary structure, showing sequences of matching bases that form spirals in the three dimensional structure.

vitro conditions with the capacity of assembling RNA portions resulting in duplicates of themselves (Paul and Joyce, 2002) and to self-sustain an exponential amplification of that ribozyme as long as the chemical substrate is not depleted (Lincoln and Joyce, 2009). After several generations, genetic errors accumulate and novel variants grow under environmental conditions set by the experimenter to dominate the population (Joyce, 2009). A review of self-replicable RNA ribozymes was published by Cheng and Unrau (2010). It has been suggested that some RNA sequences can catalyse RNA replication with an error rate of 1.2% of differing nucleobases (Johnston et al., 2001). The significance of this finding in the RNA world implies that there is no need for any other molecule besides RNA and precursors of organic material to transmit genetic information to new copies of the initial RNA sequence and to generate variation on the descendants.

Discussion

In the primordial RNA world, the components that would result in an explosion of genetic variability, but also the components for a cognitive system, appeared. Each RNA strand matches one cognitive variant. Ribozymes were the only structures that could transmit genetic information and thus provide cognitive variants based on previous ones, avoiding the generation of purely random RNA strands, which is particularly inefficient. For example, RNA strands in the experiment by Joyce (2009) were around 70 nucleobases in length, which makes a total search space of $4^{70} \sim 10^{42}$ different combinations. Random exploration of the search space would make the putative cognition too inefficient even for the low standards of the family of Darwinian cognitions. Hence, a replicable substrate is important for efficiency. The substrate of this putative cognitive system is composed of a sequence of nucleobases that compose each RNA molecule. The catalytic activity of RNA strands has the capability of joining, splicing, replicating and mutating other RNA molecules, which can serve a double function, namely as storage of genetic information and participating in enzymatic activity, as already described. Moreover, the range of different biochemical reactions on non-RNA molecules that can be induced by the sequence is huge, which confers a great flexibility and expressiveness to ribozymes. Hence, RNA molecules possess the characteristics required for a productive and expressive substrate in a powerful cognition.

Identification of the other two components remain. With respect to the selector, RNA molecules are subject to natural selection as has already been described in section 4.1.2. Again, no external entity sets the goal of the selector, but rather survival-directed goals emerge spontaneously. In this case, fitter RNA molecules correlate with faster replicating ribozymes and not necessarily with increased fidelity. Indeed, in a competitive world with limited resources and scarce prebiotic material and substrates, RNA strands capable of acquiring material faster are the ones that replicate most successfully (Joyce, 2009). The last component to be identified is the variant generator. The simplicity of the molecules do not allow for self-regulation of mutation and recombination, so the source of variants could only come from replication errors in the way mentioned in section 4.1.1 and external sources such as ultraviolet radiation (Pullman and Pullman, 1963).

Summary

Recapitulating, in the first stages of life there was a family of molecules, probably RNA, with the potential to self-replicate that preceded all life on earth. The enzymatic activity of these first molecules was initially restricted to self-duplication, but expanded its function to facilitate other reactions that contributed to faster self-replication. The RNA world represents the best candidate for the emergence of a fully fledged cognitive system, but had no actual capability to perform any function other than self-replication. New RNA variants that lead to improved replication rates in a competitive environment with limited resources would be quickly fixated by the mechanisms of natural selection. This method of acquiring new abilities suggests that the Darwinian cognition is capable of learning. The lack of sophistication in the variant generator derived from the spontaneous emergence of this cognition implies that the mechanisms leading to cognitive development and self-improvement must have been terribly slow, which agrees with timescales in the history of biology of hundreds or thousands of millions of years. In this respect, the RNA world may be considered as the minimum cognitive system: It has a great potential, but little capacity to react in the environment and extremely long processing times. The simplicity of this cognition should not mask the fact that it is particularly flexible at generating solutions. To sum up, RNA strands in the early stages of life could not synthesize many molecules, but they have the power to learn how to synthesize every organic molecule used in complex life.

TTT	Phenylalanine	TCT	Serine	TAT	Tyrosine	TGT	Cysteine
TTC	Phenylalanine	TCC	Serine	TAC	Tyrosine	TGC	Cysteine
TTA	Leucine	TCA	Serine	TAA	STOP	TGA	STOP
TTG	Leucine	TCG	Serine	TAG	STOP	TGG	Tryptophan
CTT	Leucine	CCT	Proline	CAT	Histidine	CGT	Arginine
CTC	Leucine	CCC	Proline	CAC	Histidine	CGC	Arginine
CTA	Leucine	CCA	Proline	CAA	Glutamine	CGA	Arginine
CTG	Leucine	CCG	Proline	CAG	Glutamine	CGG	Arginine
ATT	Isoleucine	ACT	Threonine	AAT	Asparagine	AGT	Serine
ATC	Isoleucine	ACC	Threonine	AAC	Asparagine	AGC	Serine
ATA	Isoleucine	ACA	Threonine	AAA	Lysine	AGA	Arginine
ATG	Methionine	ACG	Threonine	AAG	Lysine	AGG	Arginine
GTT	Valine	GCT	Alanine	GAT	Aspartic ac.	GGT	Glycine
GTC	Valine	GCC	Alanine	GAC	Aspartic ac.	GGC	Glycine
GTA	Valine	GCA	Alanine	GAA	Glutamic ac.	GGA	Glycine
GTG	Valine	GCG	Alanine	GAG	Glutamic ac.	GGG	Glycine

Figure 4.3

The universal genetic code associates three-nucleotide DNA codons (left columns) with a reduced set of aminoacids (right columns). This translation is encountered in the cells of the three domains of life: Prokaryote, Archaea and Eukaryote.

4.2.2. Unicellular organisms and the genetic code

The next lifeforms that I will review in the family of Darwinian cognitions are unicellular organisms. The relevance of the appearance of the genetic code will be analysed. For the sake of simplicity, genetic exchange processes such as conjugation, transformation and gene transfer are not considered.

The basic building blocks for a spontaneously emerging cognition were already in place in the RNA world, however, RNA sequences are limited in the amount of reactions that could catalyse. Somehow, the primordial RNA sequences managed to overcome this limitation and develop the genetic code (Forterre, 2005), which gave such an advantage that all living beings as they are known today share the code, although there are some exceptions (Barrell et al., 1979). Switching to protein-based enzymes required the genetic code and corresponding decoding biomachinery to be established. In this section, the advantages of the genetic code and the synthesis of proteins in relation to the three component cognitive theory will be discussed.

Expressiveness

DNA and protein synthesis can be interpreted as cognitive enhancements to the primordial cognition based on the RNA world. DNA strands became the transmitter of genetic information and could improve stability in RNA strands (Lesnik and Freier, 1995), which was crucial to maintain longer sequences. Protein synthesis posed a great advantage compared to ribozymes. Augmentation of the number of building blocks from 4 different nucleobases to 20 aminoacids (figure 4.3) increased the number and efficiency of biochemical reactions catalysed by living forms, while retaining the mutational and recombinant capabilities of RNA. The genetic code increases significantly the cognitive expressiveness of DNA/RNA sequences and it allowed the Darwinian cognition to increase control over its own mutability and recombination factors (Tenaillon et al., 2001) resulting in self-improvements to the efficiency of the variant generator in this putative cognition. Enzymatic activity of proteins is only bounded by the potential interactions between the many types of aminoacids and other molecules, whereas the expressiveness of the RNA world is limited to the potential interactions between the 4 nucleobases with themselves and with other chemicals present in primordial RNA. Proteins can react in many more ways with other molecules than RNA. With the genetic code, RNA found a way to improve expressiveness by translating its sequences into proteins. Thus, the evolutionary capacity of RNA was maintained while the versatility to manipulate the biochemical environment was greatly increased. In cognitive terms, the expressiveness of the substrate was enhanced and enabled the Darwinian cognition to improve its own variant generator. The cognitive substrate remained largely the same since DNA and RNA have very similar functionality with respect to evolution of sequences.

Subgoals in natural selection

Mechanisms of protein biosynthesis are encapsulated inside cells. In the case of unicellular organisms, DNA or RNA strands share the same space inside each cell with proteins, ribosomes and other molecules. If the mechanisms that decode RNA and DNA are lost, or the RNA or DNA sequence is corrupted, the cell dies. Consequently, it is necessary to duplicate the gene expression mechanisms with each duplication of RNA/DNA. Not only is natural selection acting on RNA and DNA, but also in the gene expression mechanisms and the capacity to replicate the whole cell in a controlled manner.

For that reason, the goal of natural selection in the RNA world is not valid any more. Replication of RNA molecules is substituted with replication of the whole cell. Despite this goal substitution, with respect to the cognitive theory, natural selection is still the cognitive selector and the ultimate goal of survivability remains valid, although application of replication and survivability needs to be extended to the whole cell, rather than to single RNA strands. The Darwinian cognition spontaneously achieved more sophistication in its selector mechanisms through application of natural selection to gene expression mechanisms as well as the bare RNA/DNA strand.

4.2.3. Multicellular organisms

Complex organisms are the last biological system that will be analysed. Evolving unicellular lifeforms into complex organisms required further sophistications to the putative cognitive processes of life. These systems and what it meant in terms of cognitive enhancements will be compared against those of unicellular organisms.

In the same way that unicellular organisms kept the cognitive substrate in the RNA world, namely RNA or the more stable DNA, the Darwinian cognition in multicellular organisms shares the same substrate as unicellular organisms.

Both share the genetic code and similar decoding mechanisms, but the former devised even more sophisticated methods to increase the expressiveness of the substrate and become a relevant player macroscopically. Not only does gene expression control protein synthesis in complex organisms, but it also controls cell replication and cell function in a context of inter-cell collaboration.

Another goal switch is encountered in multicellular cognitions that increases sophistication of the assessment method with respect to unicellular organisms. Natural selection remains as the cognitive selector. In this case, however, complex organisms are assessed as single entities, in contrast to unicellular organisms and RNA strands in the RNA world, where natural selection is applied to individual cells and RNA strands, respectively. In multicellular organisms, the goal of the selector becomes survivability and reproduction of the organism. The goal of survivability and reproduction is dropped for individual cells, which become dispensable for the sake of the whole organism.

Generation and expressiveness

DNA sequences in complex organisms code instructions to grow zygotes into complex individuals. These instructions include arbitration mechanisms between cells that decide on cell function and thus instructs the cells which proteins to produce and when. All cells in complex organisms share the same genetic information, although gene expression differs between them. It is only in the germ cells where gametes are produced with the intention of generating genetic variety. Somatic cells do not contribute to generate genetic variability, and consequently, neither do the DNA strands that they contain. Indeed, multicellular cells have two distinct cell division methods: mitosis and meiosis. Mitosis consists in dividing a cell in two with identical genetic information, whilst meiosis enables recombination and chromosome splitting for the purpose of combining to another gamete of a different individual and create a zygote with unique genome. Mitosis enables expression of a unique genetic sequence to grow zygotes into adult individuals and extends natural selection to assess the skills of an individual to survive and reproduce in a macroscopic environment.

Cognitively speaking, evolution found a way to separate the cognitive correspondence of individual growth (expressiveness) and reproduction (variant generation), enabling the cognitive system to interact macroscopically with the environment. A new variant is generated with each zygote. Then, zygotes differentiate in germ cells and somatic cells. These two categories of cells are specialized in different cognitive functions. Somatic cells do not intervene directly in cognitive processing of variants. Rather, somatic cells decode and express the variant that was born in the zygote. Conversely, germ cells are the ones specialized in producing new cognitive variants, and have little involvement in cognitive expressiveness because they are only involved in producing gametes, rather than ensuring survivability of the individual and ensuring that reproduction takes place.

4.3. Summary of the chapter

This chapter has suggested indications to apply the three-component cognitive theory to biological evolution. The correspondences between cognition and evo-

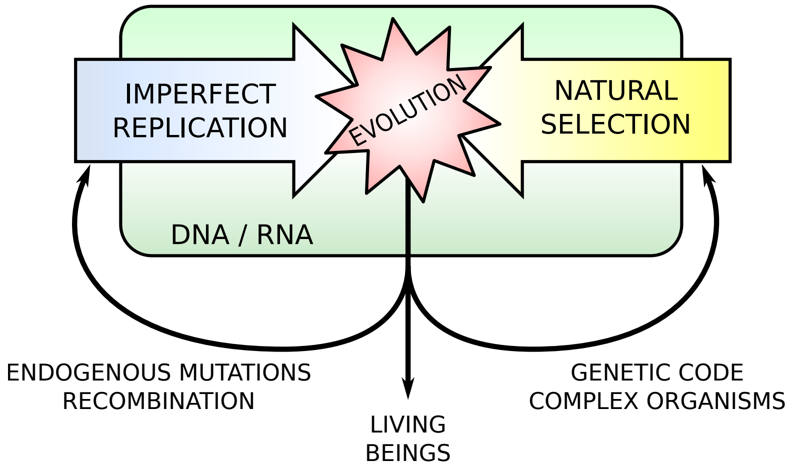

Figure 4.4

Overview of the Darwinian cognition. The three cognitive components correspond to reproduction, natural selection and DNA/RNA as shown. Development of the genetic code and complex organisms infused natural selection with more sophisticated goals. Endogenous mutations and recombination enhanced the generation of variants. Genes express in the target environment, in this cognition corresponds to a natural ecosystem.

lution suggest that, assuming that the cognitive theory is valid, life is indeed a cognitive system.

The components required for a cognitive system emerged spontaneously in nature during the early stages of life. Mutational radiation may have been involved in producing variants on long organic molecules that formed the substrate of the cognition. Once self-replication occurred in these molecules, natural selection emerged inevitably. The RNA world is a candidate to be the first cognition to appear, and also the simplest one due to the simplicity of the cognitive components.

RNA strands may have developed a genetic code and the ability to synthesize proteins. These skills significantly enhanced the expressiveness of the substrate by augmenting the range of reactions and also optimized the efficiency of the cognitive system by developing molecular mechanisms that enabled self-regulation of mutation and recombination. Finally, complex life emerged and extended the expressiveness of the substrate to macroscopic interactions with the natural environment. These notions are summarized in figure 4.4.

The interpretation of the three component cognitive system is not limited to the biological systems described in this chapter. Other lifeforms that exhibit intelligent behaviour may also be subject to interpretation. Prions have been described as well as an example of a biological system that does not constitute a cognitive system.

To sum up, evolution can be considered as a cognitive system with very long processing times. The disembodiment of the generator prevents the system from implementing goal models to improve the efficiency. It is a non-recursive cognition, but capable of self-enhancing its generator and able to attain increasingly complex goals.

CHAPTER 5

The Rational Cognition

The human brain is the only system scientifically recognized as truly cognitive. I will now discuss how human intelligence relates to the three-component cognitive theory. Other cognitive theories generally focus on inferring human cognitive processes and describing them. These models are many times implemented in computers to mimic human cognition in an attempt to replicate its cognitive abilities. I will take a different perspective in this chapter to give evidence for a possible mapping between the three-component cognitive and data from neuroscience: this link is the rational cognition.

I will reserve the use of the term *cognition* to refer to the three-component cognitive theory, unless I explicitly do otherwise, i.e. by using the term *traditional* cognitive processes. The rational cognition is considerably more complex than the Darwinian cognition. In fact, it is necessary to creatively define links between the theory and experimental data. This should not be taken as an argument to refute the theory, but rather as a demonstration that the mechanisms of the human brain are not trivial. I will also differentiate between goals in the cognitive system and goals of a human:

Goals refer to the goals of the cognitive selector.

Desires refer to the volitional goals of the individual.

Nobody knows exactly how the human brain works. The lack of scientific knowledge is due to several reasons. Firstly, there are technological and ethical limitations to experiment with it. Secondly, identification of the basic building blocks of human cognition, if ever possible, has proven to be one of the hardest challenges of scientific enquiry due to the brain's complexity. For these reasons, every possible theory of cognition that has been proposed is tentative, if not close to pure conjectures. The link presented in this chapter is no different: it will remain inconclusive until more convincing evidence is gathered. Due to the openness of the questions related to neuroscience, there are many ways to interpret scientific data. The interpretation I give here of the rational cognition is the most controversial of all the cognitive systems discussed in this thesis, without doubts. Not only is any tentative interpretation of human intelligence inconclusive, but there is no possibility, for now, of refuting it.

I will limit the claims of this chapter to a review of theories and findings of neuroscience that conform with the theory. I do not intend to propose a new cognitive theory of human rationality. Rather, the purpose of this chapter is to find analogies between findings in neuroscience and the three-component cognitive theory, as a way of reinforcing the claims of the latter. Supporting the claim that the human brain does indeed follow the three-component cognitive theory is outside the scope of this thesis. This claim would require a much deeper review of the available data. The evidence is quite abundant, so I will only cite what I consider the most relevant data for my claims and at the same time are amongst the most influential in the field. Nevertheless, the theory makes some predictions that result from mapping the components to the neuroscientific literature.

The nature of the cognitive theory and that of traditional cognitive abilities do not seem to match well at first sight. Indeed, traditional cognitive abilities recede to a secondary role for a successful identification of the cognitive components in the human brain. However, I am not denying the importance of these abilities. I will propose that there are more fundamental processes that are needed to build up such abilities. Besides, some of these traditional abilities can be found in the cognitive components, as will be shown throughout the chapter. These ambitious claims come from interpretation of the human brain as conforming with the three-component cognitive theory.

Outline of the chapter

Firstly, I will show that it is possible to identify the cognitive components from a behavioural perspective, that is, without the need to recur to neural knowledge. Next, I will propose a distinction between the terms cognition and intelligence. This distinction is necessary to differentiate what is truly cognitive, as defined in chapter 1, from the neural mechanisms that exhibit intelligence behaviour but do not have learning capabilities. Afterwards, the cognitive components will be identified in neural processes. The following section maps the components to perceptual and motor systems. Lastly, I present some other considerations, summarize the chapter and give final discussion points.

5.1. Knowledge cycles

Some authors have subjectively described their thought processes that they use to build knowledge. For example, Poincaré (1908) stated that he discovered mathematics by selecting amongst competing proof paths and then writing them down formally to assess its validity. For him, inventing was selecting. Moreover, according to Popper (1978), knowledge of the external world comes from production of conjectures, testing for its validity and rejecting unfit ones in a method of critical selection. Curiously, he also notes the resemblance of this process with that of natural selection. These statements suggest that the three-component cognitive theory can be identified without recurring to evidence from neuroscience, since the authors admit to the generation and posterior selection of proofs and conjectures. I will start the discussion on the rational cognition by identifying cognitive components in iterative engineering and in the scientific method. Note that due to the flexibility of the rational cognition, identification of cognitive components is more subjective than in the Darwinian cognition.

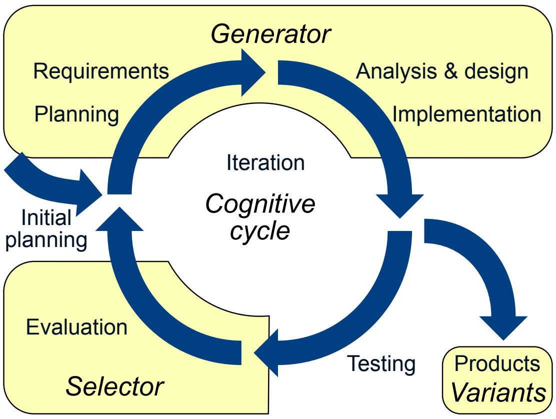

Figure 5.1

Identification of cognitive components in iterative design. New products (generator) are tested and evaluated (selector). See the text for more details. Cognitive correspondences are indicated in italics.

Iterative engineering

Iterative engineering is a design process that gradually improves a marketable product. In every iteration, the product is improved to meet better the expectations and to adapt to the requirements of the customers. It has been applied successfully for decades and allows companies to learn from previous product deficiencies (Larman and Basili, 2003).

Identification of cognitive components in iterative engineering is quite flexible.

1. First of all, variants are marketable products. The cognitive system, in this case the designer, produces these variants in an environment where the ultimate goal is to design products that are valuable for customers. The substrate is loosely defined because the variants can take many different forms and they do not need to transfer any knowledge to the next cycle; this information is stored in design documents.

2. The generator is also loosely defined. There are multiple levels of cognitive recursion (section 3.2.2) to consider here. For example, a brainstorming session may generate many designs (one level of recursion), then these designs are refined after product simulation and prototyping (two more levels of recursion). A customer that suggests an improvement is also generating a new variant (yet another level of recursion).

3. The ultimate selector is the free market. Products that do not succeed commercially are extinguished. Other selectors are in effect with each recursion: in brainstorming sessions, failed simulations, failed prototypes, etc.

In iterative engineering, designs are commonly carefully crafted following some models. Cognitively speaking, this means that the generator does not use chance as a source of variability, which contrasts to the heavy use of chance in the gen-

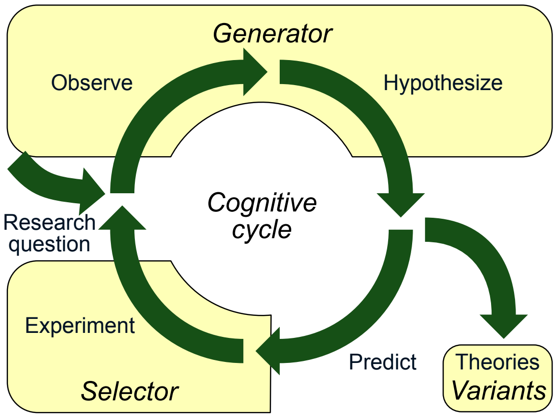

Figure 5.2

Identification of cognitive components in the scientific method. Hypothesized theories (generator) are tested and discarded (selector) if falsified. Cognitive correspondences are indicated in italics.

erator of the Darwinian cognition. Lastly, cognitive cycles are almost explicit. They can be easily identified in figure 5.1.

Scientific method

The cognitive components are also present in the development of scientific knowledge. The scientific method is an iterative process that leads to better theories and hypothesis. The method is not exactly the same as iterative engineering (Cross et al., 1981), but the cognitive components can be identified in a similar way. This time, the cognitive substrate is immaterial. Figure 5.2 depicts the classical procedure for scientific experimentation. The cognitive components can be identified as follows.

1. The variants are scientific theories and hypotheses.
2. The generator is formed by researchers who observe an event and make hypotheses to explain the event. This generator is recursive, since the elements that form it, the researchers, are cognitive systems themselves.
3. Experiments are devised to assess the theories. Falsified theories are discarded, or negatively selected.

Cognitive cycles are present in the scientific method as well. As with the iterative engineering process, the generator does not depend on chance as much as the generator of the Darwinian cognition, which makes it much more efficient. Multiple theories may be supported simultaneously by different researchers, but in the end the most parsimonious theory gets selected.

To sum up, there is no need to recur to experimental psychology and neuroscience to admit to some analogies between rationality and the three component cognitive theory. However, it is still possible to be more precise in the identifications of cognitive components. The rest of the chapter is devoted to this task, which is significantly more complex.

5.2. Split between cognition and intelligence

The terms *cognition* and *intelligence* are used ambiguously in the field of artificial intelligence and psychology. Before moving forward, I want to propose here a distinction between these terms to reconcile them with the definition of cognition given in the Introduction in chapter 1. This distinction will be most helpful to identify the cognitive components in the rational cognition, but can be applied to artificial systems and neural processes alike.

Firstly, *intelligence* would be composed of the processes and methods that read data from (neuronal) inputs, transform it, and output the result to other intelligent processes or to effectors. In a computationalist theory of intelligence, all methods would belong to this notion of intelligence. Intelligent processes are reflex, automatic behaviours. On the other hand, *cognitive* processes would not know anything about inputs, outputs or the transformations that can be applied to them. On the contrary, cognition would manage the processes ascribed to intelligence, devising and combining them. Pure cognitive processes do not have access to the environment. Cognition would therefore delegate interaction with the environment to intelligent processes. Failure of the intelligent processes to fulfil the goals of the cognitive system would trigger cognitive processes to devise other alternative intelligent methods. In relation to the three-component cognitive theory, intelligent processes correspond to decoding mechanisms of the cognitive substrate, whilst cognition conforms with the activity of the three cognitive components. Again, in the scope of the three-component cognitive theory,

Intelligence refers to those mechanisms that enable a system to interact with the environment in a directed manner and can process arbitrarily-complex information.

Cognition refers to those processes that create and manipulate intelligent processes and assess them against a defined goal.

This definition of cognition is compatible to previous definitions. It merely clarifies the notion of cognition in the current context.

In the human brain, intelligence and cognition are entangled. Both use neuronal mechanisms to operate and both depend on each other. In order to clarify the distinction, I will refer to evidence of reflex behaviour in humans. Pure automatic behaviours are most easily seen in reduced states of awareness, such as during sleep (Mutani et al., 2009). Cartwright (2004) has described sleepwalking as automatic behaviours that are not subject to volitional control. In some cases, it has led to difficult legal cases because the degree of automaticity of the human brain during reduced states of awareness is not clear yet (Arboleda-Flórez, 2002). Broughton et al. (1994) described the case of a man that committed homicide of his beloved wife during an episode of sleepwalking. He assured that he could not remember any of the acts he committed, and that he would have never done that voluntarily. The defence claimed that the aggressor had unconsciously perceived his wife as a threat during his sleepwalking and automatically responded to confront a life threatening situation using only automatic mechanisms. This case might be an example of automatic mechanisms in perception and motor control interacting with each other to produce complex behaviour with no cognitive control whatsoever. The actions that the aggressor had to carry were indeed complex and require a certain degree of intelligence: navigation in his property, physical confrontation of the threat, relocation of a corpse and return to the dormitory. Assuming that the claims of the aggressor were true, the case is a good illustration for the above definition of intelligence.

Another less tragic example of automatic behaviour is that of automatic driving under the effect of Z-drugs (Hoque and Chesson, 2009). These drugs affect alert states and may led to sleep-driving-like behaviour in experienced drivers (Pressman, 2011). Sleep-drivers perform complex tasks, such as perception of traffic signals, crash avoidance with other vehicles, keep the car in the lane, changing gears, etc. However, during conscious states, there is a continuous supervision of these automatisms (Groeger and Clegg, 1997). It is precisely in a cognitive component, the selector, where a supervisory process is required (see section 3.1.3), suggesting that cognition is not available during sleep states, but intelligent processes might nevertheless activate. Sleep-driving provides more evidence about automatic, intelligent behaviour, suggesting a dissociation between cognitive control and intelligent action as defined in this section.

5.3. Cognitive components

Identification of the cognitive components in the rational cognition is not as clear cut as in other cognitions. I will introduce a tentative mapping by delving into neural mechanisms, and I will delineate them more precisely later.

5.3.1. Variants and the cognitive substrate

It is difficult to identify precisely the cognitive substrate because the underlying processes are not well understood. I have previously identified variants as marketable products, but also as ideas and theories. These variants are very dissimilar, yet the rational cognition can operate equally on them. This is possible thanks to the use of representations of concepts in neural circuits. Martin (2007) suggested that representations of object concepts are stored in partially distinct sensory and motor neural networks, and that they are not explicitly found, but rather emerge from weighted activity in neural circuitries. Not only object concepts may be represented, but also abstract, action-related and immaterial concepts. Zahn et al. (2007) argued that social concepts are stored in the anterior temporal lobe. Miller et al. (2003) claimed that humans and many other animals can classify stimuli and store them as generalized classes. They also explored neural correlates of different categories of concepts in several brain regions, including representations of visual and auditory stimuli, motor action, and rules. Current hypotheses indicate that spiking neuronal patterns realized in neural circuits encode information in the brain (Rolls and Treves, 2011).

Baars (1988) described the brain as a conglomerate of highly specialized but unconscious, automatic processors, or modules, working together (also Fodor, 1983). Each of these modules may classify stimuli, control motor activity, and perform other computational tasks, but may also modulate other neural circuits (Valencia et al., 2009). Each of these processors can be considered a variant in the scope of the three-component cognitive theory. The computations that each variant performs correspond to decoding mechanisms of the cognitive substrate and thus conform the intelligence of the system. The substrate is shaped by the neural circuits that support the representation of concepts. For example, the neural circuit that controls muscle synchronization for automatic walking is a variant located in (pre-)motor areas. Variants that represent abstract concepts and/or manage other variants are also possible; these will be referred as meta-variants.

Consequently, variants are located throughout the cortex depending on their function and are realized as neural modules. Representation of environmental entities in variants allow for cognitive manipulation of these entities.

5.3.2. Generator

The variant generator in the rational cognition defines what variants are available for selection. Generation of variants comes in two flavours:

1. Variant creation. We have seen that knowledge is represented by neuronal spiking patterns, which depend on how neurons are connected. Emergence of new spiking patterns therefore require adaptation of neural circuits. Indeed, Choquet and Triller (2013) reviewed the latest findings in dynamics of neural networks and confirmed that synaptic plasticity in the cortex underlie learning even in adults. It is not well known what are the exact mechanisms that drive synapses to create new connections, but in any case, these mechanisms are essential to create new spiking patterns and thus, essential to learning.

2. Variant recall. Due to the high flexibility in the goals that can be pursued, each of the variants in the rational cognition are dedicated to different goals. Not every variant is appropriate for every task, so the generator elicits only those variants that are most relevant for the goal at hand. The rational cognition requires goal models (section 3.2.3) in order to recall the appropriate variants.

Here goes an example to illustrate these concepts: Consider a person that learns how to ride a bicycle. At first, riding the bicycle requires cognitive processes that combine sensory inputs and motor outputs. Relevant variants are recalled in sensory and motor regions during attentive learning. After several training sessions, synaptic plasticity creates a new meta-variant that automatically connects the relevant sensorimotor variants in a coherent way to ride the bicycle. Once learning is over, the new meta-variant can be cognitively recalled and will automatically trigger the use of other relevant variants to ride the bicycle, releasing the cognition for other tasks. Riding a bicycle becomes the cognitive expressiveness of the sensorimotor variants that are automatically recalled. Consequently, intelligence is involved in riding the bicycle, whereas the task of the cognition is to learn the skill. The goal of the cognition switches from “turn the handlebar to the same side that the bicycle is falling to” to “ride in a straight line”.

Chunking

There is evidence in sequence learning experiments that confirm the capability of the human brain to combine several distinct neural circuits. Verwey and Eikelboom (2003) and Jiménez (2008) designed experiments based on the discrete-sequence production task that involved movement of fingers in easily predictable sequences to defend that the brain treats these patterns of sequences as blocks of motor commands, or *chunks*. The fact that easily recognizable patterns of finger movements were carried out faster than more complex patterns suggests the presence of chunking. Furthermore, Miller (1956) defended similar chunking mechanisms in a classical paper where he called this process *recoding*, in analogy to communication theory.

5.3.3. Selector

Attention is the best candidate for the cognitive selector. Many definitions of attention have been proposed, and most of them convey selection of concepts, motor programs, qualia and other representations. (Anderson, 2004) defined attentional systems as those that *select information to process at serial bottlenecks where it is no longer possible to do things in parallel*. What information is selected is not random at all; it is closely related to its relevance to desires (recall that I am using the term *goal* to refer to goals of the cognitive theory, and the term *desire* to refer to volitional goals of the individual), but there are other functions associated to attention (Petersen and Posner, 2012). How exactly does attention exert its selectivity is not well known, but it may be controlled by desires (Dijksterhuis and Aarts, 2010). Conflict monitoring and disengagement mechanisms may be involved in the selector as well (discussed in section 5.4). Gazzaniga et al. (2009) classified systems that regulate attention in two categories: Voluntary attention is a top-down approach that is influenced by desires. Reflexive attention comes from bottom-up influence of stimuli detection. Stimuli can seize voluntary attention and direct it towards the stimuli. In cognitive terms, stimuli may be momentarily seizing control of cognitive goals to redirect them momentarily towards factors that may be relevant for the system.

Every cognitive system requires a goal. Goals in the rational cognition are peculiar, since they can be set by the same mechanisms that will use them, suggesting a recursive cognition in the selector. Moreover, the selector is not devoted to a unique goal, which makes the rational cognition much more complex than cognitions with only one goal. According to the proposed mapping of the three-component cognitive theory, goals are associated to variants to form a goal-model-based cognition. Goals are related, but distinct from desires. Sometimes we talk about our life desires, other times we discuss how to achieve some shorter-term desires. Yet other times instinctive impulses drive our desires. The relation between both is that desires ultimately drive the goals of the cognition. Goals drive the immediate behaviour, whilst desires are more persistent and guide the selection of the goals that will be used by the selector, effectively configuring self-reference on the selector.

5.4. The components in action

I review here some evidence that suggests the presence of complex interactions between cognitive components in the rational cognition. I will start with a mechanism in the auditory system for variant recall, followed by indications of a cognitive selector in motor-breaking circuits. Then , I will mention some

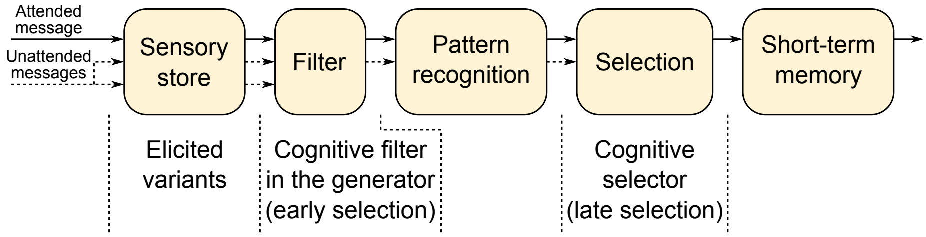

Figure 5.3

Relationship between models of selection and the cognitive selector. Early selection models conform with filters in cognitive generators. Late selection models conform with cognitive selectors. See also figure 3.5.

research on visual attention, where the generator and the selector can be seen working together.

5.4.1. *Variant recall and selection in the auditory system.*

Putative variants that correspond to auditory perceptions are localized in the temporal lobe. The rational cognition does not necessarily dictate how and when to generate variants: they can be recalled automatically. So, it seems plausible that external auditory stimuli may trigger elicitation of auditory variants. Simultaneous stimuli may trigger many variants at the same time, thus they have to be selected.

Many models of selection are available for the auditory system. Broadbent (1958) defended an early selection model where all information is analysed at a basic level and only attended stimuli is analysed deeper. Conversely, Treisman (1964) defended a late selection model where unattended stimuli are merely attenuated and processed up to the semantic meaning. The mapping of the cognitive theory to the auditory attentional system supports both models (see figure 5.3).

Indeed, in early and late attentional models, attention modulates what messages are selected. We saw in section 3.2.3 that the cognitive generator in goal-model-based cognitions may use knowledge of the goal to filter variants before arriving to the selector. Early attentional models conform with these generators: stimuli are blocked even before cognitive processing is available. Cognitive filters in the generator also conform with the preattentive stage in the feature-integration theory of attention of Treisman and Gelade (1980). This stage features parallel search, in agreement with the parallel character of cognitive generators. Selection in late attentional models conform with the cognitive selector: stimuli (variants) are fully processed and then selected. Moreover, the focused attention stage of the feature-integration theory features serial searches, like the sequential selection of the cognitive selector.

5.4.2. *Variant discard in motor breaking mechanisms.*

Motor breaking mechanisms feature action interruptions that can illustrate the mapping between brain processes and the cognitive selector. This analysis focuses in the cognitive consequences of selecting a motor behaviour that does not

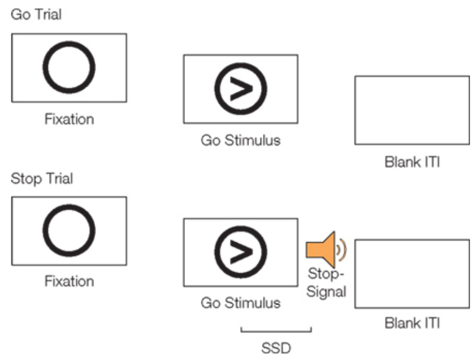

Figure 5.4

In the stop signal task, participants are required to press one of two buttons depending on the direction of the arrow in the cue. In some trials, a stop signal is issued and participants must inhibit the response. Figure retrieved from (Cohen et al., 2011).

fulfil the desire of the experiment.

Experimentation with breaking mechanisms in motor control commonly recurs to the stop signal paradigm. Briefly stated, a participant performing the stop signal task must press one of two buttons as indicated by a visual cue, except when an auditory signal instructs them to withhold their response. The auditory signal is present only in a minority of trials, so the best strategy for participants is to plan to press the buttons and inhibit the response when the signal is present. Logan (1994) found that the number of failed stops was significantly greater when participants had to respond quicker, suggesting that motor areas need some time to disengage motor actions.

Cognitively, the stop signal task can be interpreted as follows. Variants related to sensory perceptions and motor behaviours involved in the task are recalled and prepared for action. When the stop signal is perceived, the cognitive goal changes from pressing the correct button to pressing no button. Variants that were recalled in the preparation stage are no longer valid for the new goal. Goal supervision mechanisms detect the incongruence between the new goal and the goal expectancy of the variants recalled, and trigger discard of the variants by the cognitive selector. In the rational cognition, discarding a variant does not mean removing it from the system, but rather disabling it.

Fitness supervision in the anterior cingulate cortex

Kenemans and Ramsey (2013, s. 4.3) reviewed evidence for action monitoring and suggested that the anterior cingulate cortex (ACC) is involved in detecting contradictions between automatic actions and voluntary intentions. Cognitively, the role of action monitoring mechanisms conform with the notion of assessment functions in cognitive selectors, and may be triggering motor variant discard and selection of a different motor variant (or none) in order to adjust performance. The ACC is also activated during incongruent trials in the Stroop task (MacDonald et al., 2000), which reinforces the claim of the presence of breaking mechanisms due to the nature of the task. Disengagement of action conforms with negative selection of corresponding variants. (figure 5.5).

Red	Green	Orange	Blue
Yellow	Blue	Green	Red

Figure 5.5

The Stroop task involves naming the foreground colours of the words. Naming words with incongruent colours elicits activation of the anterior cingulate cortex, which is a region commonly linked to contradiction-monitoring mechanisms. The existence of such mechanisms suggests that the human brain may be operating a cognitive supervisor, which is also required in cognitive selectors.

Disengaging mechanisms are not only restricted to action inhibition, but social behaviours like racial attitudes are inhibited too. Kubota et al. (2012) detected stronger activations in the ACC in participants that were unconsciously more aversive to out-group members and voluntarily inhibited this automatic feeling. However, breaking mechanisms in perception involve other areas, mainly the right inferior frontal gyrus and the right temporal-parietal junction (Corbetta and Shulman, 2002), so it is not clear that there is a dedicated region for monitoring conflict. In any case, the existence of neural mechanisms for conflict monitoring supports the concept of a fitness supervisor in cognitive selectors (introduced in section 3.1.3).

5.4.3. Goal-models and variant discard in the visual system.

Experiments in visual attention hold evidence that simultaneously support the analogies of variant recall, selection and discard that I presented in the previous two sections. I will focus on the additional singleton search task to analyse how the generator and the selector operate in an integrated manner in the visual system of the human brain.

The additional singleton task is a visual search experiment where participants search for a specific stimulus in their visual field. Theeuwes (1994) designed two experiments to test the effect of distractors. In the first experiment, the stimuli consists of three or six separated segments surrounded by colour circles arranged around the fixation point. At stimulus onset, one of the circles changes colour. Participants respond by indicating the orientation of the segment inside the colour singleton. In some of the trials, an additional singleton appears, constituting an abrupt onset distractor. The set-up of the second experiment is the same, except that all circles except one change to red. Analysis of reaction times suggested that even if subjects retain a attentional set for a colour singleton, abrupt onsets capture attention involuntarily. Colour singletons also captured involuntary attention.

I will now consider a single trial with a distractor singleton and analyse it with respect to the three-component cognitive theory. There are three stages to consider in each trial: preparation, goal usurpation, and stabilization.

1. *Preparation.* Before stimulus onset, top-down control areas prepare the visual attentional system. There is evidence that it is not possible to prime singleton features at this stage other than with respect to its localization in the visual reception field by adjusting the size of the attentional window. Directing spatial attention seems to alter the sensory gain for features at locations primed, resulting in an apparent increase of stimulus contrast

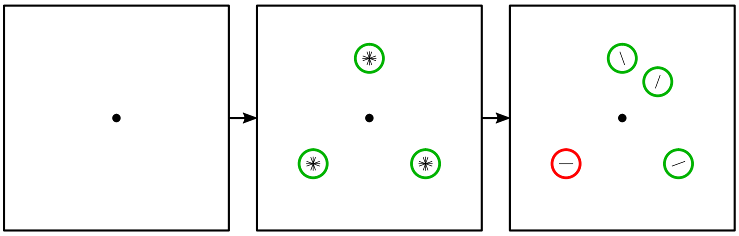

Figure 5.6

In the additional singleton task, participants must indicate the orientation of a segment that is either horizontal or vertical. A colour circle cues where the segment is. In some trials, distractors are present, such as incorrect cue and additional singletons.

(Theeuwes, 2010). The cognitive correspondence is that visual variants are primed in the generator, but not yet recalled, in anticipation of relevant stimuli at attended locations. Stimuli at these locations are predisposed to be positively filtered by the generator, and thus are more easily recalled than variants that represent stimuli at other locations.

2. *Goal usurpation.* At stimulus onset, the variant that represents the distractor singleton is salient enough to bypass the filter in the generator and is recalled at the expense of expected stimuli. The effect is that the cognitive selector experiences an involuntary goal switch that directs attention towards the unexpected singleton, effectively usurping the cognitive goals set by the participant's desires. Theeuwes et al. (2006) suggested that top-down attention cannot vary the initial selection priority, indicating that the filter bypass of the distractor variant is an automatic mechanism that is not under volitional control. Moreover, Wentura et al. (2014) suggested that in the additional singleton task, attention is directed towards stimuli that may constitute life-threatening situations or other positive opportunities for the individual. The evidence suggests that attentional capture is natured and thus, goal usurpation may be a mechanism devised by the Darwinian cognition that momentarily and automatically seizes the goals of the rational cognition.

3. *Stabilization.* After the first swipe of information, the cognitive selector evaluates the decision of directing attention towards the distracting stimulus. A mismatch between the goal pursued by the selector and the desires of the individual triggers an additional goal switch, this time self-controlled by the cognition. Mayer et al. (2004) claimed that exogenous facilitation activated very few areas in the brain, whereas endogenous facilitation is a more effortful task that involves a large cortical network. This claim agrees with the hypothesis that bottom-up attention capture is triggered automatically by goal-usurping variants, whilst top-down attention is directed by recursive mechanisms in the cognitive selector that overwrite their own goals. Top-down control mechanisms make use of the three cognitive components for the task. The result is a shift of attention from the goal set by the usurper to the genuine desires of the individual, producing a deactivation of reflex visual variants that compete to define the goals. This deactivation is probably encoded spatially and can extend

to subsequent trials (Klein, 2000). The resulting effect of bottom-up attentional mechanisms is that it warns the cognitive system about potential environmental factors that may go against its desires. Once the cognition evaluates the actual irrelevance of the stimuli, it triggers a shift of attention to the original task, or otherwise proceeds to deal with the event that caused the distractor.

5.5. Other considerations

I will complete the analysis of the rational cognition with some additional considerations.

Parallel and serial

The human brain has been described as a massively parallel network. On the other hand, human cognition is commonly viewed as a serial integrator of information (Baars, 1993). The three-component cognitive theory supports both views. Parallel processing is accomplished in the generator, where sources that recall variants, such as external stimuli and preattentive processes, compete to recall the variant that will be presented to the selector. These variants are assessed serially and disengaged if their goal model and the goal of the selector do not match. Indeed, in perceptive systems, stimuli are processed in parallel in earlier stages, but selected for further processing serially, as was shown in previous sections.

The role of the prefrontal cortex: Meta-variants.

The prefrontal cortex (PFC) has been linked to problem-solving and planning tasks, but is in general poorly understood (Koechlin et al., 1999). Miller et al. (2003) claimed that the neural correlates for learning and applying rules are located in the PFC. Certainly the PFC does not encode sensory experiences (Haynes, 2015). But it does encode stimulus-response associations (Boettiger and D'Esposito, 2005). More evidence of neural correlates of complex behaviour in the PFC is available, but an extensive review is outside the scope of this thesis.

Therefore, the PFC may be involved in engaging other perceptual and motor circuits in complex tasks. This point is compatible with the view that sensorimotor-related variants can be modulated by other meta-variants to form complex sequences, abstract concepts and associations between perception variants and motor variants. The evidence reviewed so far indicates that the PFC houses a collection of highly specialized and independent neural circuits devoted to goal selection and complex tasks. Consequently, variants that manage desires and decompose them in actual cognitive goals may be located in the PFC. Indeed, Koechlin and Hyafil (2007) suggest that the PFC serves a protecting role for long-term volitional goals (desires). Desires may be represented in variants in the PFC, which could direct cognitive goals in smaller timescales.

The contribution of emotions to cognition is not well known. My wild guess, with respect to the three-component cognitive theory, is that emotions are involved in the cognitive selector as fixed meta-variants that have been designed by the Darwinian cognition (Ekman, 1992), and define the ultimate goals of the rational cognition (Nelissen et al., 2007). In the same way that salient stimuli engage attention in the additional singleton search task, strong emotions may redirect the goals pursued by the cognitive selector unless they are disengaged by the cognition.

To sum up, executive functions such as planning, decision making and problem solving might be performed by meta-variants in the PFC. Hence, these abilities might just belong to human *intelligence*, but not human *cognition*. Together with emotions, they might only play a secondary role in the rational cognition.

Recursion and self-reference.

I have conveyed that the rational cognitive components may involve recursion. In section 5.1, I suggested that rational generators may be recursive. Determination of goals in selectors can also be attributed to a recurrent cognitive system. In this case, variants themselves represent goals. But, recurrence in the generator and the selector is problematic because the homunculus problem emerges. This problem is solved by recurring again to the notion of flexibility. The cognitive substrate in the rational cognition is extremely flexible. Through neuronal representation of physical and abstract variants, the substrate can take any form. Not only can objects in the environment be represented and treated as variants, but also methods that make up the cognitive components and even meta-variants can be represented. For example, goals can be defined by the same cognitive selector that will make use of them. The cognition may also devise methods that specify what variants are recalled and under what conditions. The rational cognition achieves in this way self-reference in the three components.

5.6. Summary and discussion

In this chapter I have presented some evidence that suggests that there is a link between the human brain and the three-component cognitive theory. The putative mapping involves complex features in every cognitive component, such as self-reference, goal models and storage of unused variants for later retrieval. Curiously, the existence of recursive cognitions has been refuted, for a good reason. Flexibility of the substrate takes over cognitive recursion in a much more efficient way, since it enables the application of cognitive mechanisms to a wide range of systems, both internal and external to the cognition. Entanglement of cognition and intelligence comes from a circular relationship between each other. Intelligent methods make up the cognitive components, and cognitive mechanisms update those same intelligent methods.

To sum up, in the rational cognition, variants correspond to neural modules in the human brain that can categorize stimuli, activate motor programs, manage goals and elicit activation of other modules. Variants can also correspond to entities in the environment thanks to mental representation of entities. The cognitive selector is realized by attentional systems and conflict detection circuitry. Goals for the cognitive selector are, at least partially, self-selected by the rational cognition. The generator creates new variants for new problems through

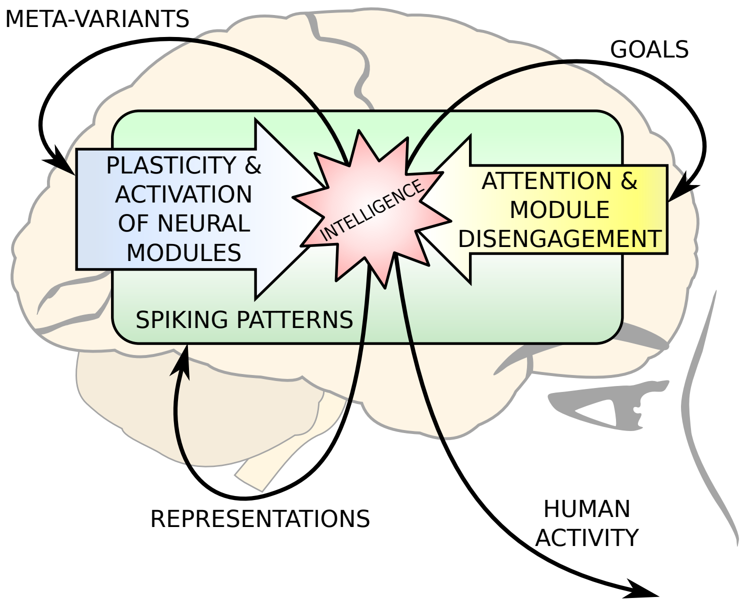

Figure 5.7

Overview of the mapping between the three-component cognitive theory and the human brain. Intelligent behaviour (as defined in section 5.2) is the product of the rational cognition, which exhibits self-reference in the three cognitive components.

synaptic plasticity and recalls older variants for known problems. Figure 5.7 illustrates the proposed mapping.

I have made some conjectures about the nature of cognitive processes that rely on tentative predictions of the mapping between the cognitive theory and the brain. Planning, decision making, problem solving, logic, reasoning, induction, deduction, etc. are certainly involved in human intelligence, but are not central to the three component cognitive theory. The open hypothesis is that these abilities can be explained as a complex network of meta-variants. Moreover, restriction of the definition of cognition and intelligence to traditional human abilities is anthropocentric and may hinder the development of potential cognitive systems that do not think like humans do. Indeed, (Johnson, 1992) and (Legg and Hutter, 2007) support this position and propose new methods to measure intelligence based on the performance of the system. Other classical human abilities such as memory, free will and consciousness have not been explored either. These discussions are left open.

Another important aspect to human cognition is language. Mutani et al. (2009) consider sleep-talking as a motor automatism. On the other hand, Roser and Gazzaniga (2006) found evidence that basic thought processes do not depend on language, but rather language interprets thought to construct a coherent narrative. Moreover, it is evident that language is essential for social communication, but it is arguable whether it is essential for thought or not. So, it might as well happen that language emerges from another complex network of meta-variants that enhance, but are not essential to, the rational cognition. I wildly speculate that language might emerge in the scope of the three-component cognitive theory from assigning a word to each variant. Other variants and meta-variants would then define the rules of grammar, produce utterances and recognize speech.

The evidence for all the analogies presented in this chapter is in most part circumstantial but, I believe, persuasive. Neuroscience is in its infancy and more research is needed to reveal the mysteries underlying human intelligence.

CHAPTER 6

Technological cognitions

In the history of artificial intelligence, many models have been proposed that attempt to create machines that exhibit learning capabilities beyond their original scope. Recent advances in Artificial Intelligence are reducing the amount of preprogrammed models that an artificial intelligence needs, allowing them to operate in environments with increasing uncertainty. However, it is unknown whether existing paradigms are powerful enough to keep this trend indefinitely. In this chapter, I will justify with the three-component cognitive theory why these methods fail at achieving the same learning performance of other cognitive systems. Three technologies that are commonly employed in artificial intelligence will be analysed:

1. Symbolic processing systems: The Prolog programming language will be analysed as representative of the symbolic paradigm.

2. Connectionist paradigm: Classification algorithms will be analysed, specifically neural networks and weight optimization, which are inspired by neuronal models.

3. Evolutionary computation: It is interesting to analyse genetic algorithms since they are inspired on biological evolution.

General remarks

For the sake of simplicity, I will assume that the systems discussed here are fully operated in immaterial domains. Therefore, any cognitive system and its decoding mechanisms will be immaterial in nature.

Cognitive generators described here tend to be the kind of fake generators described in section 3.2.1 when referring to generators with perfectly or almost perfectly efficient heuristics. Recall that knowledge in these heuristics come from the cognition that has devised the generator of the putative cognitive system, in this case a human designer that devises methods in artificial intelligence.

The notion of *goals* must be clarified in the current context. Goals are commonly used in a very broad sense in computer science. Many times, it can be exchanged with the term *input*, particularly when goals are indicated as reference values in control systems. Other times they are used as a cue to select which algorithm should be used from a pool of available algorithms in order to transform some input data. Frequently, it is not easy to distinguish a legitimate cognitive goal from an input to the system that has been arbitrarily termed *goal*. In this chapter, *goal* is only used to refer to the goals that the cognitive selector pursues. Since there is no self-reference in this family of cognitions, goals are fixed and set by the human designer.

6.1. Declarative programming

I will firstly discuss the Prolog programming language, which belongs to the subfamily of logical programming languages. In this language, the programmer writes down in a language-specific syntax the set of rules that defines the problem to compute and the compiler operates the rules to work out all possible solutions. Prolog is one of the many programming languages commonly used to automate reasoning in the scope of artificial intelligence. It belongs to the class of declarative languages, which use formal logic declarations to decide what actions to take.

The Prolog language

Figure 6.1 shows a canonical program typically used to demonstrate the basic characteristics of Prolog. A program starts with a section that declares some truths, or facts. In this example, it defines the relationships of a three generation family composed by 5 men, 5 women and 3 marriages. A graphical representation of the family is shown in figure 6.2. The second section defines the rules that can be applied to the truths stated in the first section in order to generate new theorems. Queries can then be issued interactively or appended at the end of a program file. Queries are theorems that Prolog needs to prove from the initial truths and rules. The language could in principle build a knowledge base with all possible truths, i.e. all family relations, but in more complex programs this is not possible due to the high amount of time and memory that would be needed to store all facts.

There is no need to show off more characteristics of the language that would only create confusion and would be superfluous to the argument I want to present. I will only add that it is a Turing-complete language, so any program that can be coded in any other language can also be programmed in Prolog. For a more in-depth introduction to declarative programming and Prolog, see (Clocksin and Mellish, 2013)

The Prolog programming language attempts to solve the solution to a query K by proving the validity of K from a set of facts L and a set of logical rules R that altogether form what is known as the knowledge base. In simple examples, the solution can be trivial, such as in the query `parent(lois, carter)` on the family specification in figure 6.1, where the rule in line 20 is directly applied to reach the fact in line 13. In other examples, the complexity is such that a programmer can easily get lost of the logic operations that are applied to the antecedent L to reach the consequent K. Let us trace the execution of the somewhat more complicated query K : `mother(stewie,lois)`. Briefly stated, what goes behind the scenes is that rules that match the query K are checked recursively against the primary facts and discarded if not found. The trace of

```prolog
% Facts
man(mickey).
man(carter).
man(peter).
man(chris).
man(stewie).
woman(thelma).
woman(barbara).
woman(lois).
woman(carol).
woman(meg).
parents(peter,  mickey, thelma).
parents(lois,   carter, barbara).
parents(carol,  carter, barbara).
parents(chris,  peter,  lois).
parents(meg,    peter,  lois).
parents(stewie, peter,  lois).

% Rules
parent(C, P)       :- parents(C, P, _); parents(C, _, P).
father(C, F)       :- parent(C, F), man(F).
mother(C, M)       :- parent(C, M), woman(M).
sibling(X, Y)      :- parent(X, Z), parent(Y, Z), X\==Y.
brother(C, B)      :- sibling(C, B), man(B).
sister(C, G)       :- sibling(C, G), woman(G).
grandparent(C, G) :- parent(C, P); parent(P, G).
grandfather(C, G) :- grandparent(C, G), man(G).
grandmother(C, G) :- grandparent(C, G), woman(G).
uncle(C, A)        :- parent(C, P), brother(P, A).
aunt(C, A)         :- parent(C, P), sister(P, A).
nephew(T, C)       :- sibling(T, S), parent(C, S), man(C).
niece(T, C)        :- sibling(T, S), parent(C, S), woman(C).
```

Figure 6.1

Prolog code listing with the specification of a random stereotyped family. The first section defines the primary facts of the system and the second section defines the set of rules that can be used to build up new facts. Queries have been left out but can be introduced during the interactive execution of the program.

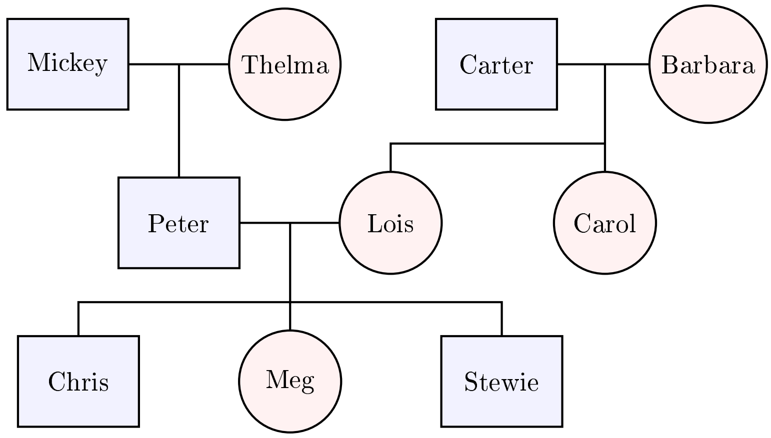

Figure 6.2

A family tree consisting of three generations depicting the relations coded in 6.1. Members of the same generation are placed within the same height and children on the next row immediately below. Women are indicated with rounded shapes and men with square shapes.

K can be seen in figure 6.3. In this example, rule `mother/2` on line 22 in figure 6.1 defines that variable M is *mother* of variable C if it is their *parent* and is also a *woman*. The second condition is directly satisfied, but checked last in the tracing, whilst the first condition needs some more elaboration by the program. In line 5 of the trace output listing shown in figure 3.5, there is a failed attempt to assess that `parents(stewie,lois,_)` exists, where `_` specifies an anonymous variable, and subsequently backtracks to check other possible ways of proving the query as specified for the rule `parents`. The program succeeds on the next attempt, line 7, and proceeds. Multiple solutions can be found for each query as long as their truth is asserted.

Discussion

The power of declarative languages has been acknowledged in multitude of applications, such as mathematical theorem provers, natural language processing, symbolic equation solvers, etc. (Clocksin and Mellish, 2013). In the field of artificial intelligence, probably the most significant recent use of Prolog is in the IBM deepQA architecture that was used in the Watson computer (section 2.1.1, Lally and Fodor, 2011). However impressive, these applications remain in the domain of narrow artificial intelligence. I will make use of what I have considered the most favourable position to interpret Prolog programs as cognitive systems. Even so, identification of cognitive components proves to be difficult.

In order to aid in identifying the cognitive components in declarative programming languages, I will directly point them out in the example above. In Prolog, cognitive variants can be considered to be instantiated by theorems that the compiler generates for a query. The cognitive generator uses the source code as a template to generate variants, that is, the rules defined in a program dictate the cognitive generator how to create new variants from the theorems available. It is the role of the human designer to implement a complete set of

```
1  | ?- mother(stewie,lois).
2        1      1  Call: mother(stewie,lois) ?
3          2      2  Call: parent(stewie,lois) ?
4          3      3  Call: parents(stewie,lois,_65) ?
5          3      3  Fail: parents(stewie,lois,_65) ?
6          3      3  Call: parents(stewie,_65,lois) ?
7          3      3  Exit: parents(stewie,peter,lois) ?
8          2      2  Exit: parent(stewie,lois) ?
9          4      2  Call: woman(lois) ?
10         4      2  Exit: woman(lois) ?
11         1      1  Exit: mother(stewie,lois) ?
12 yes
```

Figure 6.3

Example of a Prolog trace listing on GNU Prolog 1.3.0. Output of the command *trace* for the query `mother(stewie,lois)`. The program is able to successfully reach the queried statement from the facts and rules in figure 6.1.

rules, thus a big part of the intelligence in any Prolog program relies on the designer. When analysing the query `mother(stewie,lois)`, two variants are generated. The first one is rendered unsuccessful at line 5 in figure 6.3 and the second succeeds and asserts the query. More variants are possible with a more liberal interpretation, such as considering that testing each family member against being a woman also generate variants. Prolog extinguishes further processing in branches that lead to no valid partial theorems. Finally, substrate is immaterial and is shaped by the set of every possible theorem that can be deduced in each program.

In declarative programming languages, goals are explicitly introduced and the exact method to reach the goal is omitted, which must be figured out by the implementation of the programming language (Lloyd, 1994). Actually, the method that accomplishes the intermediate steps between the set of rules and the desired goal is previously determined by the designer of the programming language.

Cognitive reach of declarative languages

There are fairly strong limitations in considering declarative programming languages as cognitive. The three components are limiting factors. Firstly, selection of variants is hard-wired into the language. There is no possibility of doing other operations than solving logical queries. However, this is not such an important limiting factor because there are other powerful cognitive systems that can perform well with reduced goals, such as the Darwinian based ones. Secondly, the substrate, in this case the space of immaterial variants, is frequently limited and hence non productive. The set of possible variants is indirectly hard-wired by the programmer by the set of rules and facts. Finally, variant generation operates on a relatively limited set of possible variants and it is possible to walk through all of them, making heuristics irrelevant in the long run and defeating the purpose of a cognitive system once the search space is fully explored. By far, the biggest drawback is that variants are not associated to any decoding

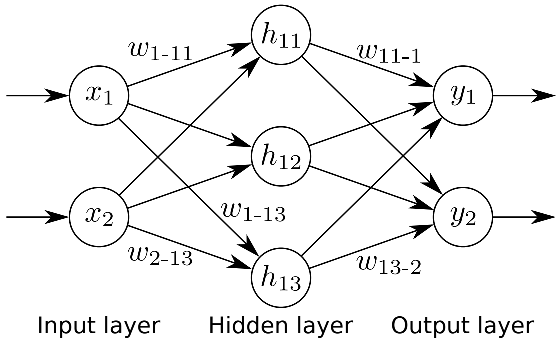

Figure 6.4

Dependency graph of a conceptual neural network and typical layers. This neural network is composed of an input layer with 2 units, a hidden layer with 3 units and an output layer with 2 units. Some weights have been labelled.

mechanisms. The selector evaluates the variants rather than the expression of the variants.

All in all, declarative programming languages are a good approximation to the design of intelligent systems when the set of facts and rules are known, but the cognitive components lack the necessary power to make these languages a true cognitive system, even in the most favourable interpretation. Maybe that is the reason why it has a long history in artificial intelligence but has failed to create truly intelligent systems as with every other attempt in the field.

6.2. Neural networks

Neural networks are the major representative of the connectionist paradigm in artificial intelligence. They were designed with biological neurons as inspiration, although only some resemblance remains with actual neurons. Here I review and discuss the most basic form of an artificial neural network. The conclusions that follow can be generalized to other kinds of artificial neural networks.

An artificial neural network is constituted by a set of interconnected units, each with several inputs and one output, in imitation with biological neurons which generally have many dendrites — inputs, and one axon — output. Each unit performs operations on the inputs to calculate a single output, which can be connected to many other units simultaneously. The inputs to each unit are multiplied by a factor called the *weight* of the connection $w_{i\rightarrow \mathrm{j}}$, that can be adjusted to the relative importance of each input. Positive weights are said to be excitatory whilst negative weights are inhibitory. The output is filtered through an activation function ϕ. Units are grouped in *layers* such that only units from previous layers have connections with the next layer. Topological variations are also possible (Graves et al., 2009; Fukushima, 1980). See figure 6.4 for a diagram of a simple non-recurrent neural network.

A mathematical description of neural networks is necessary to understand them. Each unit j holds a value y_j that represents that unit's state and can be calculated using the formula

$$y_j = \phi\left(\sum_{i=1}^{n} w_{i\to\mathrm{j}} \cdot x_i + w_{0\to j}\right)$$

where x_i is the activation value of the previous layer, n the number of units in the previous layer, $w_{i\to j}$ the weight associated to the connection between unit i in the previous layer and unit j, $w_{0\to j}$ the bias of the activation and ϕ the activation function. The logistic sigmoid function is frequently used for the activation function:

$$\phi_\sigma(x) = \frac{1}{1+e^{-x}}$$

The whole mathematical definition for the neural network depicted in figure 6.4 using only input units as independent variables is:

$$y_1 = \phi\left(\sum_{j=11}^{13} w_{j\to 1} \cdot \phi\left(\sum_{i=1}^{2} w_{i\to\mathrm{j}} \cdot x_i + w_{0\to j}\right) + w_{0\to 1}\right)$$

$$y_2 = \phi\left(\sum_{j=11}^{13} w_{j\to 2} \cdot \phi\left(\sum_{i=1}^{2} w_{i\to\mathrm{j}} \cdot x_i + w_{0\to j}\right) + w_{0\to 2}\right)$$

How the output of y_1 and y_2 is interpreted is defined by the human designer. For example, each output unit may control the joint angles in a humanoid robot (Noda et al., 2014), outputs may be interpreted as an index in classification on data mining applications (Kotsiantis, 2007), or it may host a universal Turing Machine (Siegelmann and Sontag, 1991).

Training a neural network involves adaptation of weights applied to input connections. Here I briefly introduce back-propagation in neural networks for supervised learning. The way in which coefficients are modified follows some strict rules which are selectively applied depending on the conformance of the output with the goal for a set of training data. Input vectors are presented to the neural network and the weights adapted with each vector by increasing the weight of each connection that strengthens the desired output, and lessening otherwise. Neural networks are classifiers that learn from experience. After training takes place, weights are fixed and data that needs classification can be supplied at the input units, one vector at a time.

Discussion

The most favourable interpretation of neural networks with regard to the three-component cognitive theory is that each possible combination of the set of weights constitutes a cognitive variant. Cognitive processing takes place only during training of the neural network, and the expressiveness depends heavily on the devices connected to the output units. The heuristics applied to select the most appropriate combination of parameters is fixed and allows for one possible adaptation of the weights. Alternative weight combinations are not tested nor checked for appropriateness. Consequently, with just one variant that evolves during training, there is no possibility for neural networks to consider other options in the weights: There is only one predefined state that the system can evolve to. In neural networks, the cognitive power lies in the heuristics of selecting the unique combination of coefficients in each training cycle. A cognitive selector is thus redundant, and is actually unidentifiable in backpropagated neural networks. Therefore, cognitive capabilities are located in the designer of the heuristics of the training algorithm. The advantage of neural networks is that they release the burden of manually adjusting each parameter in the neural network, but the rules that drive the evolution of the weights is still determined externally. Enhancements are possible, though. It is viable to make size-variable neural networks that will grow the number of units as needed, yet it will barely improve the capabilities of the cognitive components associated to neural networks, since the effect is that the cognitive substrate becomes somewhat more productive. Application of genetic algorithms to parameter searching may improve cognitive capacity, but there will still be limitations inherent to genetic algorithms that will be delineated in section 6.3.

Neural networks are reducible to optimization of classification problems. Many other statistical classifiers exist (Theodoridis and Koutroumbas, 2008). The cognitive interpretation for other classifiers does not differ significantly from neural networks. Consequently, they are not the definitive learning technology for machines. Noda et al. (2014) attempted to fuse visual information in speech-processing applications and suggested it improves recognition rates, but they admit that *cautious selection of sensory features is crucial for attaining high recognition performance*. Other fields that have successfully applied neural networks also suggest that selection of sensory features is crucial (for example, Cardamone et al., 2009). An intelligent machine willing to use neural networks would need to select which sensors are essential for reaching the goals and discard those sensors that will just hinder the performance of the neural network. Current artificial intelligence fails to give an answer to this problem, and selection of inputs still relies on the common sense of the human designer and empirical results from previous experimentation. Moreover, neural networks are far from giving the same performance as human vision but under controlled conditions. Indeed, Nguyen et al. (2014) recently pointed out that neural networks trained with deep learning (Hinton et al., 2006), which is a method that improves training results in deep neural networks (those with several hidden layers), could be fooled by inputting specially crafted unrecognisable images. The neural network proceeded to classify those images with better confidence than images containing real objects. Nevertheless, other studies have reservations and claim that neural networks can actually improve human performance in some tasks (Cadieu et al., 2014). Hence, it is not clear that neural networks can reproduce human capabilities, even for classification problems.

To summarize, neural networks can hardly be regarded as cognitive systems, in spite of all interesting and practical applications available.

6.3. Evolutionary Computation

Evolutionary computation in general, and genetic algorithms in particular, are inspired by evolutionary models in biology. In contrast to population genetics that focus in model accuracy of biological evolution, evolutionary computation takes the concepts of mutation and recombination to perform trial and error on optimization problems with unknown structure. For that reason, evolutionary computation is considered a different system than biological evolution and must be analysed separately. I will argue that genetic algorithms comply with the requirements for a cognitive system but is impractical due to efficiency reasons. This section assumes that the reader is already introduced to genetic algorithms. A good introductory book is (Sivanandam and Deepa, 2010). I will start with an overview of genetic algorithms and then draw analogies with the three-component cognitive theory.

Brief description

In the most general form, genetic algorithms evolve a fixed-size set of solutions in several steps. Solutions have task-specific formats that embed parameters to be optimized, individually and in combination. Each step is called a generation and involves selection of a fraction of the solutions from the previous generation, application of search operators that modify the solutions and checking for termination of the algorithm. The initial generation is termed the parent population and the modified solutions are called the offspring population. Search operators vary in complexity and efficiency. Following the evolutionary model, the two canonical operators are recombination and mutation. Recombination starts from two parent solutions and combines them to generate offspring. Mutation randomly modifies a solution and allows to escape suboptimal local maxima. Once the offspring is generated, it is evaluated against a fixed task-specific fitness function. The solutions that maximize the fitness function are selected for the next generation. Termination occurs when a predefined criterion is met, i.e. no offspring solution has entered the next generation.

Population sizes can be subject to evolutionary search to withdraw size-related decisions from the human designer. This is done by taking the population size as an additional parameter on each solution that must be optimized with the rest of parameters. The parameters that drive search operators can also be integrated in the parameter set of each solution in what is known as evolutionary strategies. This way, genetic algorithms may control their own stability, efficiency and search operators, i.e. adjusting probabilities of mutation. Hence, they effectively perform self-adaptation to improve convergence speed to the optimal solution. Generally, the parameters associated to self-adaptation of each solution are mutated and recombined first, and then these parameters adjust mutation and recombination on the subset of solution parameters that are evaluated against the fitness function.

The biggest concern in genetic algorithms is efficiency, therefore population size must be carefully determined. A population size too small is under risk of premature convergence towards suboptimal solutions in local maxima whilst bigger population sizes are computationally inefficient. The most important factor when considering efficiency is the number of calls to the fitness function, since fitness evaluation can be very costly. The relatively low number of evaluations contrasts with biological evolution, where a single species may be formed by many millions of individuals.

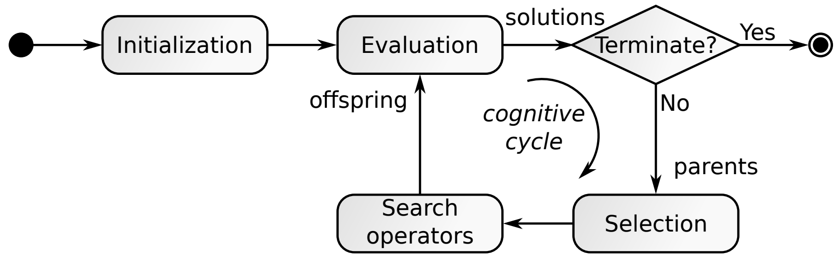

Figure 6.5

Activity diagram for a general genetic algorithm. Solutions are selected, recombined and mutated, and evaluated cyclically until the termination condition is met. Each generation adheres to a cognitive cycle.

Discussion

All three cognitive components can be identified in genetic algorithms straightforwardly. To start with, search operators, i.e. recombination and mutation, generate new solutions from previous ones just like the variant generator predicts. Afterwards selection comes into play, which corresponds unsurprisingly to the cognitive selector. Selection does reduce the number of solutions generated by search operators in agreement with the selector on variants, in this case by discarding unfit solutions, similar to the Darwinian cognition. The method then continues to produce the next generation iteratively. Each generation shapes accurately a cognitive cycle. Genetic algorithms are probably the systems where cognitive cycles are most easily discerned of all systems discussed in this thesis. Actually, cognitive cycles are hard-wired into the algorithm (see figure 6.5). The last component, the substrate, holds the information represented by each solution. The substrate is an immaterial one such as those in other technological systems already described, but cognitive expressiveness can be extended to the real world. For example, The design of actual devices can be guided with parameters that are evolved with genetic algorithms. Indeed, Chatterjee et al. (1996, sec. 2) describe genetic algorithms as precisely comprising "*three parts: Solution representation; operators which produce altered solutions; and fitness selection.*" in resemblance to biological evolution and the three-component cognitive theory.

Genetic algorithms and biological evolution share many traits. However, the accomplishments of one and the other differ significantly, probably due to constrained resources in genetic algorithms. Genetic algorithms are based on methods of biological evolution and natural selection applied to computer model optimization problems, and thus suffer similar obstacles. The differences in the number of evaluations may be another factor that determines why evolutionary computation fails to be as successful as biological evolution in the design of complex solutions to problems. Whitman et al. (1998) estimated a prokaryote world population of 10^{30} cells which differs many orders of magnitude from the world's fastest computer at $33.86 \cdot 10^{15}$flop/s= $1.21 \cdot 10^{20}$flop/hour (Top500.org, 2014). This value must be increased considerably because a computer makes many operations to generate, evaluate and select each individual. Thus, the ratio between the number of prokaryote individuals and the number of fitness evaluations in a genetic algorithm per hour is more than 10^{10}. Consequently, genetic algorithms might need much more computing power to demonstrate their true cognitive capabilities.

Scalability problems also affect the limitations of genetic algorithms (Thierens, 1999). The generator may be responsible for the scalability problems. The heuristics associated to the generation of new solutions depends much in random events and not so much in previously acquired knowledge. There have been attempts to improve the use of knowledge acquired by better than average solutions. Such is the case for the Building Block Hypothesis (Goldberg, 1989) and probabilistic model-based evolutionary algorithms. These methods protect highly fit schemata from disruption by search operators using clustering methods such as the Linkage Tree Genetic Algorithm (Thierens, 2010). Yet, efficiency is still the greatest limitation.

Evolutionary Strategies reduce the assumptions made on the problem. Most importantly, they describe pathways to increase the reach of genetic algorithms. Given enough flexibility in the parameters associated to self-adaptation, it could potentially lead to improvements in its own algorithm. Again, efficiency limits application of genetic algorithms.

Evolutionary computation has yielded interesting results in many fields. Jain and Gea (1996) devised a method for efficient component placement in electronic PCBs. Mathematical problems have been approached successfully such as graph colouring (Galinier and Hao, 1999), location-routing problems (Derbel et al., 2012) and graph bi-partitioning (Kim and Moon, 2004). Fusion of evolution and neural networks has been explored (Cardamone et al., 2009; Stanley and Miikkulainen, 2002) as well as software development through constructive grammars (Galván-López et al., 2010). A detailed review of genetic algorithms applied to diverse fields in engineering can be found in (Suresh et al., 2015). In all of these applications, the problem is bounded by the adjustment of parameters on the system that is evaluated, which generally results in limited expressiveness. Moreover, the number and complexity of the parameters are relatively simple, and selection of parameters is defined by the human designer.

As an illustration of the good results but poor efficiency, I will mention the works by Vowk et al. (2004) who developed competitive Corewar programs using genetic algorithms. Corewar is a competitive game created in the 1980's where players write programs using a Corewar custom instruction set and make them compete against each other with the goal of forcing the opponent to execute invalid instructions. The genetic algorithm created better-than-human Corewar programs, and devised strategies in program lengths of just 8 instructions that were *never seen in a human coded individual before.* The downside is that *it will take days or weeks to run each pool to completion* (Vowk et al., 2004) for programs even as short as those 8 instructions. Consequently, even though genetic algorithms can display intelligent behaviour, the efficiency is too low for practical purposes except in the most simple problems. When it comes to solve human-like common tasks, genetic algorithms fail.

Final words

Genetic algorithms have a clear correspondence with the three cognitive components. However, I stated at the start of this section that genetic algorithms are not fully fledged cognitive systems. There are several limitations on genetic algorithms that hinder their consideration as cognitive systems. Firstly, the substrate in practical applications of genetic algorithms is not productive in a linguistic sense, which reduces evolutionary computation to optimization problems. Unbounded substrates are also possible, but performance is already a major hog in non-productive substrates. Hence, the number of possible solutions, although large, is finite and consequently there is an optimal solution represented by the maximum possible value of the fitness function. Once the best solution is found, or alternatively a solution that fulfils conditions for termination, the genetic algorithm stops and the solution can be exploited. Genetic algorithms are computationally intensive, and only the most simple problems can be optimized. They are thus limited to narrow problems where variations on a very limited set of parameters are explored. There is no possibility for genetic algorithms to explore parameters other than the ones chosen by the designer.

To sum up, evolutionary algorithms do have the three components needed to progress towards an artificial cognition. Nevertheless, high computing resources are required to solve even the most simple problems. In real scenarios, genetic algorithms are used to solve narrow, partial problems on bigger engineering challenges. These methods make use of trial and error on a set of parameters to find local optima. Advanced variations of genetic algorithms are able to protect partial solutions from disruption by search operators, improving the speed of convergence. Even so, the computational requirements of genetic algorithms makes impractical to tackle more complex problems. Currently, evolutionary computation comprises the closest technological approach to artificial cognition according to the cognitive theory presented in this thesis. Nevertheless, major outbreaks of heuristic performance are needed to improve efficiency of genetic algorithms to generalize their applicability, for instance by introduction of cognitive models as described in section 3.2.3.

6.4. Summary of the chapter

Three artificial systems have been analysed in this chapter for the purpose of cognitive systems in current technology. Firstly, declarative programming was acknowledged as a cognitive system but the poor expressiveness of its solutions hinder its cognitive power. Even so, declarative languages have played an important role in developing the symbolic paradigm in artificial intelligence. Secondly, neural networks were reviewed and found to lack a generator of solutions that could be qualified as cognitive. Neural networks are regarded as classifiers with automatic and predictable coefficient adjustment that constitutes the only solution produced in each putative cognitive cycle. Thirdly, genetic algorithms have been argued as the most cognitive of all artificial intelligent methods, but efficiency limits any practical application except for the simplest ones. In all three methods, the goal of the selector is manually fixed by the human designer, and the expressiveness of the substrate is limited. The substrate is in all cases an immaterial one, where information is stored as bits in computer memory and manipulated without physical attachment. Technological generators range from fixed methods that generate unique solutions, where the cognitive power is found elsewhere, to purely trial and error methods that lack efficient heuristics able to tackle complex problems.

In conclusion, current methods in artificial intelligence comply better with *intelligence* than with *cognition*, as defined in section 5.2. Hence, in the scope of the three-component cognitive theory, it is not surprising that there is still no clear candidate for artificial intelligence, despite all the efforts and investment.

CHAPTER 7

The three cognitive families, together

This chapter is devoted to connect the Darwinian, Rational and Technological cognitions together. The links between the three cognitive families intend to strengthen the hypothesis of the existence of shared mechanisms underlying their behaviours. This chapter assumes that the previous chapters have been read and understood.

The topics covered are an aggregation of features that relate the three families of cognitions under a framework that describes the links from a more general perspective:

1. Identification of purely cognitive and purely intelligent systems.
2. Activation of cognitive processing: Exploration vs. exploitation.
3. Differences on how variants are managed: Cognitive architectures.
4. Transmission of variants between cognitive agents: Communication.
5. Bootstrapping and enhancing intelligence in the universe: Sequence of cognitive emergences.

7.1. Cognitive capabilities and intelligent behaviour

The split between intelligence and cognition proposed in section 5.2 is applicable in principle to any cognitive system. The relation with the cognitive theory is that intelligent processes correspond to the expression of the decoding mechanisms of the substrate, whilst cognitive processes are those that follow the three-component cognitive theory. But, is it possible to identify purely intelligent systems and purely cognitive systems? I think so.

Two distinct performance patterns can be distinguished in RNA and computer programs that suggest the validity of the split between cognition and intelligence described in section 5.2. Firstly, Advancements in Artificial Intelligence find more and more applications each day that were previously only attainable by humans, such as speech recognition and computer vision. Some of these applications look intelligent indeed and have already surpassed human abilities, for example at chess. While these applications may convey a subjective sensation of intelligence, they are still very far from doing other tasks that they were not specifically designed for. These methods use models that need to be studied in detail and implemented by a human designer, who injects rational models into computer programs. A computer program can demonstrate very intelligent behaviours, but it has little power to learn new skills and create new models. Secondly, RNA strands in the early stages of life had the capability to devise the genetic code and complex life, but their capacity to synthesize molecules was very limited. The latest findings claim that there was a stage in the development of life where RNA was only able of self-replication (section 4.2.1) and unintentional reactions with simple molecules present in the environment. It seems difficult to credit these RNA strands with more intelligence than, say, a computer program that makes copies of itself. However, self-replicating RNA strands have cognitive capabilities that narrow artificial intelligence programs lack, namely to learn how to synthesize molecules of arbitrary complexity, i.e. proteins. They have actually been credited as a potential precursor for all life on earth.

Consequently, there are cognitive systems incapable of doing much but with great potential, and intelligent systems that can do much but not more. Cognition can exist without intelligence and intelligence is possible with the absence of cognition. They are two distinct notions that can be analysed separately. Cognition takes care of finding new solutions, while intelligence materialize those solutions and runs them. Separately, neither intelligence nor cognition completely understand the problem at hand. But together, contingent solutions are possible for virtually any challenge. The human brain has cognition and intelligence deeply entangled due to generalized self-reference, which tremendously increases the difficulty of analysing it. Intelligent methods constitute the cognitive components, and the cognitive components devise intelligent methods. The advantage is that it can perform truly intelligent behaviours, in the traditional sense of intelligence. Of course, there are also systems that are neither cognitive nor intelligent, i.e. any inanimate object.

7.2. Exploration vs. exploitation

Exploration vs. exploitation behaviours have been described in evolution, psychology and artificial intelligence, contributing to the number of analogies between the three systems. In terms of variants, exploration tasks involve generation of new variants and selecting them, whilst exploitation makes use of existing variants and thus does not require cognitive processing.

Daw et al. (2006) discussed exploratory decisions in humans and suggested that exploration and exploitation tasks involve different cortical areas in the brain. It is not clear how the three-component cognitive theory can be linked to their claims. Cohen et al. (2007) discussed how the balance between exploitation and exploration is managed from a behavioural point of view, suggesting that the decision is cognitively accessible and thus that the rational cognition may control activation of its own cognitive processes.

With respect to evolution, exploration strategies may be described as those with high mutability whilst exploitation strategies involve low mutation rates. Sniegowski et al. (2000) delved into how natural selection can adjust the rate of mutation of a population, suggesting that the Darwinian cognition can control the balance between exploitation and exploration. Adaptation of the rate of mutation is sometimes called second-order selection. Tenaillon et al. (2001) argue that it does indeed improve adaptation to the environment. Cognitively, this is analogous to stating that self-referential mechanisms of the Darwinian cognition can control when and how to activate cognitive processing for faster evolution.

The efficiency of exploration vs. exploitation tasks has been analysed in machine learning as well, where trial and error is explicitly used for reinforcement learning. Mahadevan and Connell (1992) devised an architecture that could convert complex learning tasks to simpler subtasks that learned the desired behaviours of each individual subtask by trial and error. Cully et al. (2014) applied successfully trial and error on a 4-legged robot. They enabled exploitation strategies until the actuator program ceased to yield the desired robot movement., i.e. due to leg injury. Exploration strategies were then activated to find a new actuator program that yielded the desired movement under the new leg geometry.

The bottom line is that creation of new models and learning involves exploration, which is tightly linked to cognition. In contrast, exploitation makes use of those models, which is related to intelligence. This is applicable to evolution, human behaviour and computer learning.

Trial and error

Intelligence houses models for interaction with the environment, whilst cognition creates and assesses the models. In absence of models, the cognitive system must enter exploratory modes to generate new models, and trial and error is the way out. Indeed, Nakamura and Ohsawa (2009) found that students with less *insight* tend to revert to trial and error. Reinforcement learning in machines is based on trial and error as well (Sutton and Barto, 1998). And evolution is all about trial and error (chapter 4). Once successful models have been devised, they can be exploited to control the environment. Pure exploitation needs no model creation: All interaction with the environment can thus be delegated to purely intelligent methods.

On the other hand, learning to interact in unknown environments is where the three-component cognitive theory excels. The theory is based on the generation of models of arbitrary complexity, in the form of variants, that are assessed and selected. The variant and decoding mechanisms comprise the implicit models of the environment. Once variants that lead to the cognitive goals are found by trial and error, even if goal fulfilment is partial, they can fixate into the model of the world as implicit knowledge in variants. Note that these models are different from goal models. The former store information on how to successfully deal with the environment to pursue a certain goal, whilst the latter stores the relationship between variants and goal accomplishment.

The corollary is that trial and error is inevitable when it comes to creation of new variants in a cognitive system. This fact is shaped by the variant generator, which generates the trials, and the selector, that discards mistakes. The fewer the errors the more efficient the cognition, but making some errors is inevitable when exploring unknown solutions to problems.

7.3. Architectural differences

There are significant differences in the way that the Darwinian cognition and the rational cognition are built, which affects the way that they store and manage their variants. How variants are distributed in each cognitive system have different consequences in the efficiency and resiliency of the cognition.

In the Darwinian cognition, all variants have the same function. Most variants are duplicates or have small differences with other variants. I have stated earlier in chapter 4 that this distribution of variants is not specially effective, specially because random factors play a big role in variant generation. Moreover, all variants are designed to fulfil a single goal: survivability. There are, however, important advantages that other cognitions lack. Specifically, this cognition is very resilient to destruction of variants. The cognitive processes can proceed as long as there is a minimal population size without loosing knowledge (Shaffer, 1981). Actually, the Darwinian cognition makes use of variant destruction events as a mechanism to select the variants. It is designed to take advantage of apparently detrimental events.

In contrast, the rational cognition functions very differently. Every variant constitutes a neural module in the brain and has a different function (chapter 5). So, every variant fulfils a different cognitive goal, and it may rely on other variants to fulfil its sub-goals. The greatest advantage of how the rational cognition instantiates the cognitive theory is that it attains flexibility on the goals that it can pursue with the use of specialized variants for each task. On the other hand, the biggest disadvantage is the vulnerability of the variants. If a region of the brain is destroyed, the variants that are stored there disappear with little chances of recovery. That was the case of Broca's patient, for example, who lost a small region in the left frontal lobe that left him with impaired speech.

In sum, the rational cognition is more sophisticated and efficient than the Darwinian cognition, but more vulnerable as well.

7.4. Communication

Distribution of knowledge between individuals of a cognitive family can be an advantage. Communication events between cognitions transmit variants, and maybe information about goal accomplishment, to share a solution to a problem. Therefore, communication between cognitive individuals account for an additional source of variability for cognitive generators. The effect is that the emitter of the message becomes momentarily a recursive generator of the cognition that receives the message (section 3.2.2; recursive cognitions). Curiously, both cognitions constitute the generator of each other in a bidirectional communication event (figure 7.1).

Natural language in humans is obviously necessary for cultural and scientific transmission from parents to offspring. On the other hand, it is reasonable to think that a future technological cognition will communicate with other technological cognitions using available transmission lines, such as cables, radio and

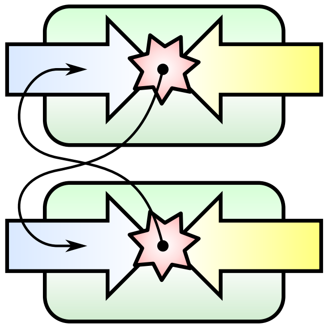

Figure 7.1

Bidirectional communication between cognitive systems. A communication event transfers a variant from one cognition to the generator of the other cognition. The former becomes a recursive cognition for the latter. The recursion is double in bidirectional communication. Variants may require transformation to adapt to and from the communication channel.

optical fibres. Communication in human and machines is well understood and documented, so I will focus in the analogy of communicational events between Darwinian cognitions and with other cognitions (Figure 7.2). Considering that a cognitive individual that belongs to the Darwinian cognition is the summation of all organisms of the same species that live close enough to interbreed, inter-population, inter-species and inter-family communicational processes can be identified:

Inter-population

Communicational events between distanced populations may be as rare as important. Trakhtenbrot et al. (2005) acknowledged the rarity of long-distance dispersal events and emphasized the importance that they have in facilitating exchange of genetic information and conservation of biodiversity. For example, in trees, a genetic discovery evolved in one island may be transmitted to the population in another distant island in the form of pollen particles under the transportation effect of periodic and infrequent strong winds. On arrival to the receptor island, the genetic information in pollen particles is integrated in the genome of the tree population on the second island, importing the genetic knowledge acquired by the individuals on the first island. Transmission of goal information is redundant because Darwinian cognitions only pursue survivability. Variants are transmitted with no transformation of information and they can be immediately integrated in the decoding mechanisms of the target cognition. In humans, the analogy is a natural conversation between two people. The rational cognition must incorporate methods that transform variant information to and from the communication channel because there is no direct way to transmit the neural circuitry that comprise each variant. Actually, written and spoken natural language may be interpreted as the realization of variant transformation. As for potential technological cognitions, the immaterial nature of digitally encoded variants might be also transmittable with no transformation. Darwinian and technological communication events are very similar

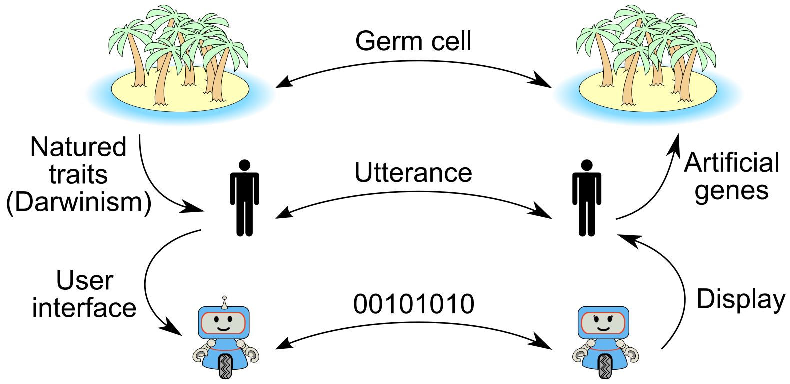

Figure 7.2

Communication events between Darwinian, rational and technological cognitions. Cognitions can communicate with each other by transmitting information about variants and the goals that they accomplish. Variants that proceed from other cognitions are a source of variability for the generator of the receptor. The goal is implicit in communication events between Darwinian cognitions. In the rational cognition, variants require a transformation before and after transmission to adapt to the communication channel. Technological communication is characterized by the transmission of programs, libraries or functions that are executed in the target processor.

indeed, since the variants are stored and transmitted in base-4 (corresponding to the 4 nucleobases) and binary encodings, respectively. Assuming that hypothetical technological cognitions will run on computers, each program, library or function may be instantiating a variant. These variants are currently designed by software engineers, but they might be designed by technological cognitions in a foreseeable future. Thus, transmission of these software components may underlie communication events in future technological cognitions.

Inter-species

There is evidence of genetic material passing between organisms outside the context of parent to offspring. Species tend to refuse genetic information from other species (Gomez-Lazaro et al., 2004), making a biological barrier for transmission of genetic knowledge between species. Even so, horizontal gene transfer is generalized between all biological kingdoms (Heinemann and Bungard, 2006). A weak analogy can be made with human conversations where speakers use different languages. In this case, the language barrier poses a difficult barrier, yet communication is possible. In both cases, transmission of knowledge becomes much more difficult and infrequent. I admit that this analogy is somewhat arbitrary, but paves the way to the next category of communicational process.

Inter-family

Inter-family communication events are those that occur between individuals of different cognitive families. For example, rational to Darwinian communication events are possible. Artificial genes can be created and inserted in a living population with the use of biotechnology. But it is more frequent that humans tamper with Darwinian cognitive components, which is not a communication event, but more an interference in the Darwinian cognitive mechanisms. This happens in the three cognitive components: Natural selection is easily substituted by directed selection (For example Turner, 2009), artificial changes to the substrate have been made (Meggers et al., 2000), and variations to the rate of mutations have been documented (Pullman and Pullman, 1963). The opposite direction is also possible in the form of variants created by the Darwinian cognition and made available to the rational cognition. This was discussed in sections 5.4.3 (goal usurpation) and 5.5 (emotions). These variants may be the result of genetic expression and drive partially the behaviour of the rational cognition (Jessell, 2000). This way, the rational cognition is endowed with knowledge about surviving in the natural environment that was developed by the Darwinian cognition, constituting a communication event from the Darwinian cognition to rational cognitions.

Communicative processes between humans and a future technological cognition might occur through user interfaces, or even natural speech, just as human-machine communication is made available today.

7.5. Cognitive emergences

There is one last think to discuss, and that is how cognitive systems emerged. Until now, existent cognitive systems have been described. It is time to describe their origin. Assuming that the three-component cognitive theory is valid, additional cognitive systems can be designed by previous cognitions once the latter acquires the knowledge to create cognitions. But there is still the problem of the emergence of the first cognition. The only way that an intelligent system can appear without recurring to other intelligent systems is by spontaneous emergence of the mechanisms needed for cognitive processing. The probability of ever occurring is at first sight extremely low, so only the simplest cognitive mechanisms would stand a chance of appearing spontaneously.

In section 4.2.1 the RNA world was described as the most basic form of the family of Darwinian cognitions. A striking convergence of events may have produced the spontaneous emergence of the first self-replicable organic molecules that may have been the precursors of life and hence, of the Darwinian cognition. An interesting and detailed account of the process of the creation of life can be found in (de Duve, 2005). Indeed, the three cognitive components met together naturally in early stages of the emergence of life. Firstly, basic organic building blocks that could react with each other to conform bigger molecules set the grounds for the physical substrate (Miller, 1953). Secondly, radiation and other random causes of molecular breakage that induced replication errors, preceded the variant generator of the first putative cognition. Third and last, natural selection inevitably joined the team (section 4.1.2). Once the three cognitive components emerged together as evolving self-replicable molecules, the first cognitive system was bootstrapped.

After millions of years of Darwinian cognitive processes, the Darwinian cognition could finally formulate the correct DNA sequence with the potential to build up cognitive brains leaning on neuronal impulses and biochemical molecular exchange. The rational cognition appeared as a means to improve fitness of humans in complex environments where fast adaptability to the environ-

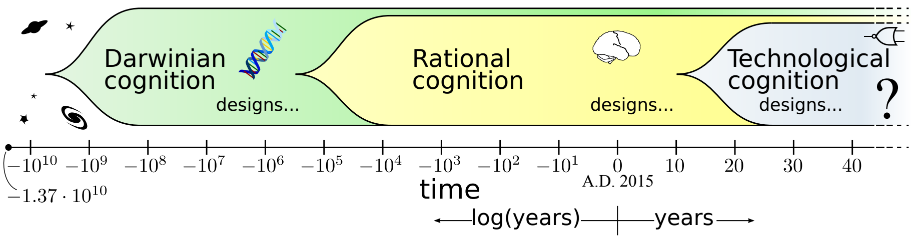

Figure 7.3

Conceptual diagram of the sequence of cognitive emergences, starting from the Big Bang up to 50 years in the future. Dates are approximate and/or tentative. The Darwinian cognition emerged spontaneously from nature and lead to the design of the Rational cognition, which in turn is hypothesized to devise future technological cognitions. The question mark makes reference to unknown future emergences, hopefully designed under the supervision of the rational cognition.

ment proved to be essential for survivability. Curiously, the goal of natural selection was not to create intelligence with improved performance, but was a by-product of designing control systems with better prospects for survivability. Consequently, the rational cognition was designed by the Darwinian cognition, in agreement with the vast amount of scientific literature that links human creation with evolution.

Cognitive emergences might not stop here. Intelligent machines that outperform the intellectual capabilities of the human brain have been suggested since immemorial times. However, the emergence of a technological cognition is essentially speculative. Despite the latest advancements in artificial intelligence, there seems to be no clear candidate method to create artificial general intelligence. Current limitations in artificial intelligence will have to be overcome in order to devise technological cognitions. Anyhow, were truly intelligent machines ever conceived, they will be designed by humans. So, that is the rational cognition designing technological cognitions in the third cognitive emergence on Earth.

Chalmers (2010) suggested additional cognitive emergences. He claimed that more advanced cognitions will be created by machines once a technological cognition emerges in the form of an arguable *technological singularity* that would outstrip human capabilities beyond control. According to Chalmers, *we should expect a sequence of ever more intelligent machines.* He does limit the sequence to artificial beings, yet the sequence may also be extended back to the first cognitions spanning the Darwinian and Rational cognitions in a similar way as posterior cognitions to the technological one in the sequence of cognitions

The sequence of cognitive emergences is delineated in figure 7.3, starting from the spontaneous emergence of the Darwinian cognition, up to the first hypothetical cognition to be developed by machines.

CHAPTER 8

Conclusion

Cognitive science lacks a reference model. In this thesis, a theoretical approach has been proposed in an attempt to remedy this situation. An abstract model of cognition that does not depend on any underlying physical mechanism has been presented. The theory describes a model that every cognitive system may ultimately follow to interact with the environment in an intelligent and directed manner. It justifies why current technologies in artificial intelligence fail to achieve general intelligence within a theoretical framework that not only integrates human cognition, but also credits other systems that exhibit intelligent behaviour, i.e. evolution, with cognitive capabilities (figure 8.1).

Summary

Firstly, leading theories of intelligence were reviewed in chapter 2, which also explored evidence of a remarkable convergence of intelligent behaviours in evolution and human intelligence that has been largely neglected in science. Afterwards, the theory was presented in chapter 3. According to the theory, cognitive systems provide tentative solutions for a given problem and then discards those that fail to achieve the goal. Each of these solutions is termed a variant. Cognitive systems are composed of three major components: A generator of variants, a selector of those variants and a physical or immaterial substrate were variants are held. The selector adheres to specified goals, which can be fixed, externally defined or self-determined. The variant generator does not need to know the goal pursued, although it might come handy in model-based cognitions to improve the efficiency of the system. Conversely, the selector needs to know nothing about how to create variants, but can decide which ones are valid and which ones are not. Later, biological evolution was linked to the proposed cognitive model. Each species individual is a variant. Imperfect replication, natural selection and DNA/RNA instantiate the cognitive components in the Darwinian cognition. This cognition has limited self-optimization and poor efficiency. Anyhow, the theory describes evolution from a framework that can be extended and optimized. Afterwards, evidence that the human brain conforms with the theory was provided. In the rational cognition, neural modules may instantiate cognitive variants. Synaptic plasticity, module recall, attention, module disengagement and spiking patterns constitute the cognitive components. However,

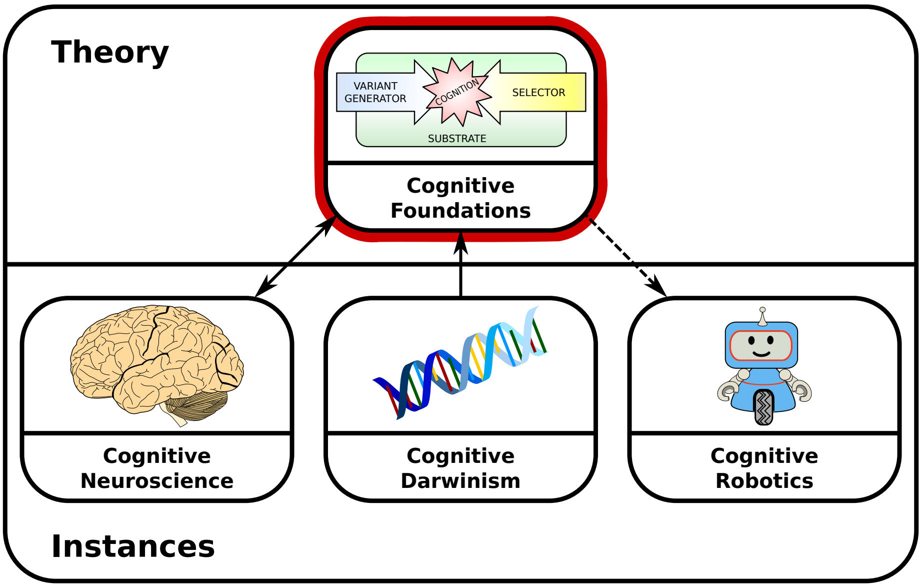

Figure 8.1

Scope of the thesis. The three-component cognitive theory, in red, proposes a global theory of intelligence. Evidence is found in biological evolution and in human intelligence. Arrows indicate the direction of contributions of each field over others. Dotted lines indicate links yet to be defined.

the difficulties of experimenting with the brain and the lack of strongly proven neural theories render any tentative theory as a conjecture for the most part. Moving on, some artificial intelligence related technologies were briefly described and also linked to cognitive components as possible. No artificial-intelligence method has been found to date that can surpass, or even match, the general cognitive capabilities of the human brain. The difficulties of identifying cognitive components in these methods may explain the limitations that they have to achieve general intelligence. Finally, chapter 7 presented more evidence that linked the three cognitive families with each other under the proposed framework.

Final discussion

This is not the first time that a many-component cognitive model is proposed (see section 2.3), but it is the first one to link it to intelligent behaviour in evolution, neuroscience and technology at the same time.

An important consequence of the theory is that every cognitive system will inevitably recur to trial and error at some point. Indeed, when there are no models to describe an event, it is difficult to predict the outcome. Human rationality has an outstanding capacity to disaggregate events and entities into smaller models and meta-models, reducing the use of trial and error to the bare minimum.

An unanswered question is if there exists a meta-model that can devise any kind of models. But, in this case, it should be capable of modelling itself as well, entering a circular problem with difficult solution. Hofstadter (1980) explored these circularities, which are well supported by human thought but fall apart in formal languages like mathematics and logic (Gödel, 1931). The theory I have presented circumvents this circularity by avoiding the use of preconstructed models in the methods that interact with the environment. Cognitive systems are only limited by the expressiveness of the substrate. Moreover, every cognitive component is prone to modification through self-referential mechanisms.

The theory I have presented circumvents this circularity by avoiding the use of preconstructed models in the methods that interact with the environment. It is only limited by the expressiveness of the substrate and does not require the use of preconstructed models to interact with the environment.

Another consequence is that the terms *biological cognition* and others alike become confusing because they have been traditionally used to differentiate rational thinking from putative technological thinking. With the notion of the Darwinian cognition, however, rational thinking processes and Darwinian cognitive processes should also be differentiated, for example by reserving the adjective *biological* to refer to the Darwinian cognition and *rational* for human thinking.

A distinction between cognition and intelligence was proposed in section 5.2. Intelligence is composed of the methods that interact with the environment, whilst cognition is characterized by development of new methods. To distinguish both in a cognitive system, the following rules can be considered. Attempts to reach a goal by trial and error are cognitive. Each attempt does not necessarily have to be designed using brute force. Sophisticated cognitive systems may avoid any attempt that is predicted to be a failure. On the other hand, behaviours that are perfectly directed towards a goal, i.e. with no mistakes, are better ascribed to intelligence. The fact that intelligent methods constitute the cognitive components in the rational cognition, and that these methods are devised cognitively, conforms an entanglement that may justify why the distinction between cognition and intelligence had not been identified until now.

A socially debated issue is that of machine control. Intelligent machines able to compare or surpass human intellectual abilities raises concerns about their control. Were the theory to be functional in machines, artificial cognitions can remain under human control even if its cognitive capabilities surpassed those of humans. The easiest method to guide the behaviour of artificial cognitions is to impede modifications to the goals in the selector. This implies that the assessment methods should only be designed by humans, which can limit the cognitive power of technological cognitions. Thus, the variant generator can run freely and even self-improve with no supervision. Should smarter machines require goal self-readjustment, self-modification to the selector should be restricted. Long term goal convergence and stability of goals become critical to ensure appropriate functionality and behaviour. Actually, the Darwinian cognition might have already devised this mechanism to control the rational cognition. In section 5.5, I hypothesized that emotions might ultimately define the goals in the rational cognition. Certainly, our desires ultimately dictate us to survive and seek pleasure. Hence, control of the technological cognition by the rational cognition might follow the same cognitive principles that the Darwinian cognition employs to partially control the rational cognition, in the terms described in this thesis.

To conclude, I have suggested three main points:

1. Trial and error may be the essence of cognition, although error should be minimized.

2. Biological evolution may hold simple cognitive abilities.

3. Traditional cognitive abilities might only play a secondary role in human cognition.

8.1. Weaknesses

8.1.1. *Difficult to attempt falsification*

The biggest scientific drawback to the theory is the difficulty of confirming or refuting it. There is currently no cognitive system that is available to scientific analysis, except maybe for the Darwinian cognition. Each cognition has its own difficulties for confirmation:

- **Darwinian cognition:** Assuming cognitive abilities in Darwinism as a valid argument may be difficult to accept. If it is assumed that Darwinism can never be cognitive, then we are doomed to deny biology as part of a broader inference of intelligence. It is partially a matter of faith as in Kuhn's (1962) beliefs on scientific paradigms. Once it is assumed, the theory can be easily modelled on the description of evolution and natural selection. The novelty of this thesis resides in the theoretical extensions that generalize the Darwinian cognition to the rational cognition and justify the absence of technological cognitions. If Darwinism is accepted as cognitive, then we can prove so with many different arguments such as the theory presented in this work, otherwise the arguments in favour of cognition in Darwinism fall into mere conjectures.

- **Rational cognition:** Even if the brain surrenders to reveal its mysteries, it could happen that the circuits could be so complex and numerous that the prospects of finding mechanisms such as the one described here are as good as any other theory of cognition. There is a risk that many theories with clashing arguments are simultaneously valid to explain intelligent behaviour in the rational cognition. Therefore, this theory would fall in the pool of theories that give no definite answer. Besides, the theory of rational cognition will need continuous refinement to catch up with the state of the art in neuroscience.

- **Technological cognition:** Every attempt at devising algorithms for artificial general intelligence has failed. This theory has not described any proven method to build artificial cognitive systems, so its usefulness on devising novel methods in artificial intelligence remains unknown. Moreover, once a technological cognition is devised, there will be no need for theories such as the one described here and consequently confirming or rejecting the theory will not be important any more. The theory falls in a deadlock: it needs a truly technological cognition to be confirmed, and a technological cognition needs a valid theory to be devised.

8.1.2. Inconclusive interpretations

I have analysed three families of putative cognitive systems as representative of different kinds of cognitions. Many other systems could be argued as cognitive, if enough arguments are given that show evidence that the system conform to the theory. The arguments for considering Darwinism, the human brain and contingently artificial intelligence as cognitive are just some that could be used to interpret the theory. Many more interpretations are possible.

The theory describes the cognitive components in their ideal form, but there might be deviations. Practical matters related to instantiation of the cognition can account for these deviations, as well as cognitive enhancements. The theory also allows for many complex features of the cognitive system, that is, self-reference, recursion, asynchronous cognitive cycles and goal models, and many ways to combine these features. There is not a clear way of making analogies between the components and the main features of any system with intelligent behaviour. This renders the correspondence between the cognitive components and the cognitions analysed here as not definite and prone to subjective evaluation from the person making the analysis. Starting from the basic principles of the three-cognitive component, different evaluators may arrive at different mappings.

In this respect, improvement in precision would greatly increase the credibility of the theory. For example, inferring mathematical models that quantify cognitive expressiveness across any kind of cognition. In each of the three cognitions, only a small subset of literature in each field was reviewed. A cognitive theory should describe the most important aspects of the cognitions analysed, but a more detailed study of scientific data may reveal inconsistencies or even refute the possibility of a mapping between the theory and any of the cognitions.

8.2. Prediction: Epistemic Cognitive Architecture

This thesis has proposed a theory that opens several interesting paths, both theoretical and practical. In the scope of artificial intelligence, the theory would have no value if it could not predict new methods to attempt artificial general intelligence. For that purpose I will propose a novel cognitive architecture that is based on the application of the three cognitive components to control systems and computer science. Recall that the rationale of this thesis was to infer a general theory of cognition that can be applied to machines. By means of the following cognitive architecture the aim is fulfilled. Hopefully, it will allow machines to construct their own intelligence from the ground up. This proposal makes use of the cognition-intelligence split suggested in section 5.2.

A schematic diagram of the architecture can be found in figure 8.2. Each puzzle piece represents a module of intelligence, which is an executable portion of code that manipulates data to produce a specific output and correspond to variants in the scope of the cognitive theory. The model starts from basic variants that can be evolved and combined into more complex variants. Alternatively, it is also possible to create completely new variants from random, however it is not desirable because it is inefficient. Models of goal fulfilment are added to successful variants and used to improve learning efficiency. No actual

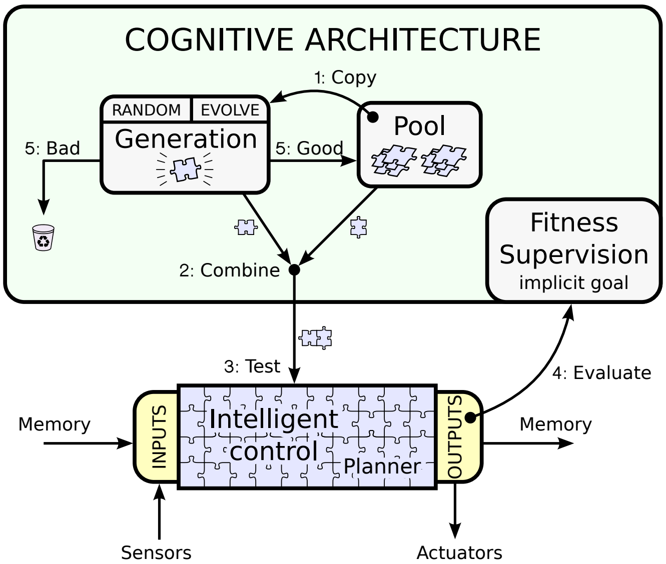

Figure 8.2

Proposed basic epistemic cognitive architecture with intelligence/cognition split. No self-referentiality is evident yet.

data is provided to the cognitive architecture, which just manages (assemble, split, merge, ...) the variants that actually manipulate inputs and generate outputs.

The cognitive components may be implemented in the following way. Firstly, the substrate requires storage and decoding of variant information. In computers, the most obvious elements that do so are memory and processors. A CPU interprets executable sequences of bits and performs operations on data as instructed by the executables. Therefore, functions, libraries and programs should be considered variants. Size of variants can vary from single operations to complex applications that run autonomously. There are huge advantages to this approach. Existent methods of artificial intelligence and other algorithms can be introduced in the cognitive system without cognitive overhead. In other words, existent software can be transferred to the cognitive system and reused as variants. In analogy to the mapping between the theory and the rational cognition, a goal should be assigned to each variant. Meta-variants that manage when other variants are executed are also desirable. Decoding mechanisms of variants should give access to data input, data output and references to other variants, which correspond to data input, data output and pointers. Multiple variants can run simultaneously as threads, and together compose the intelligence of the machine.

Next, we need a generator. In this case, the generator provides sequences of bits that will comprise the variants. Initially, the generator can be configured to generate variants in simple ways. For example, random generation of variants and random variation of previous variants may initially constitute the generator, in analogy to the generator of the Darwinian cognition. The generator is the cognitive component that will most benefit of self-reference in the initial stages of a technological cognition. If the cognition has access to improve its own

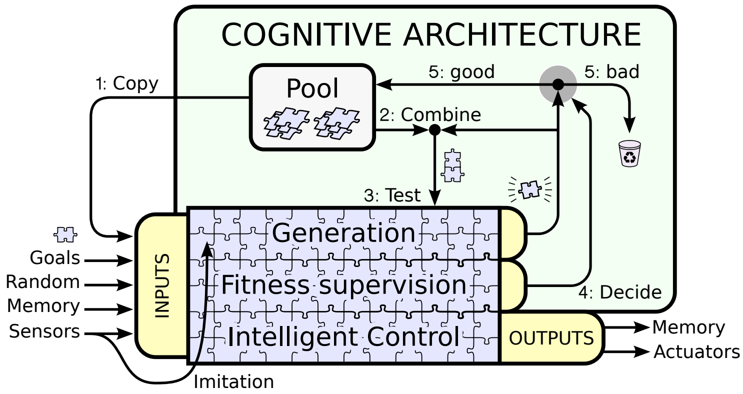

Figure 8.3

Proposed advanced epistemic cognitive architecture. The border between intelligence and cognition is not that clear in this example of an advanced cognition built upon the one in figure 8.2. Due to self-modification of the fitness supervision function, stability is critical for successful goal accomplishment. A potential informational pathway for imitating behaviours has been depicted as an example of advanced cognitive capabilities.

methods that generate variants, the quality of variants generated will greatly improve with time. The downside is that we may loose knowledge about how exactly are the variants generated, but it should be considered a time-saving feature for the human designer of the cognition.

Thirdly, the selector should start evaluating simple things to demonstrate the capabilities of the generator and the substrate. An assessment function for a more complex goal can be used thereafter. Once the correct functionality of the system is confirmed, more advanced assessment functions should be used, such as models based on reinforcement learning. Ultimately, the technological cognition should adjust its own goal assessment methods to decompose abstract goals into simpler ones. The new assessment methods must be assessed to ensure convergence to the same goals and stability of the cognition, realizing self-reference in the selector as well (figure 8.3).

Ultimately, all methods that make up the cognitive components should be accessible to the cognition (figure 8.4). This way, there is no assumption about how the cognitive components should be designed and how they interact with the environment. The only models that need an initial instantiation are those of a basic cognitive system, but still the cognitive system has access to a complete redesign of itself. Successful functionality of such a cognitive system remains for now a conjecture, but the analogies that have been drawn with other systems exhibiting intelligent behaviour in this thesis renders it a plausible cognitive architecture that is worth to research further.

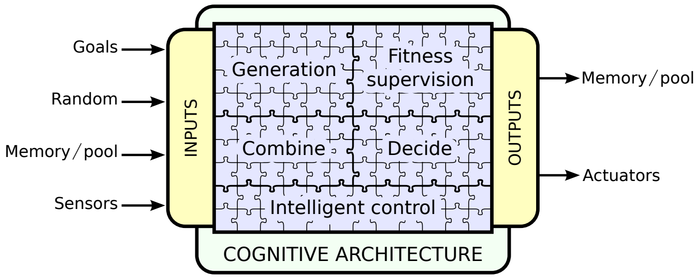

Figure 8.4

A candidate for a fully self-referential epistemic cognitive architecture. If all components of the cognitive architecture in figure 8.3 were moved to the intelligent segment, the border between intelligence and cognition falls apart due to complete self-reference in the generator (generation, combine) and the selector (fitness supervision, decide). Self-reference in the substrate is also enabled by allowing representations of modules as concepts in a similar way as it was described in the rational cognition (section 5.3.1). Sophistication reaches the maximum in this cognitive architecture, that could potentially self-develop without external intervention.

Bibliography

Alexander, P., Lett, J. T., Moroson, H., and Stacey, K. A. (1960). Changes produced by ionizing radiations and some related agents in DNA in vitro and in vivo. In *International Journal of Radiation Biology*, number Suppl, pages 47–69. 4.1.1

Amoretti, M. and Reggiani, M. (2010). Architectural paradigms for robotics applications. *Advanced Engineering Informatics*, 24(1):4–13. 2.1

Anderson, J. R. (2004). *Cognitive psychology and its implications.* Worth Publishers, 6th editio edition. 5.3.3

Arboleda-Flórez, J. (2002). On automatism. 5.2

Atkins, J. F., Gesteland, R. F., and Cech, T. R. (2010). *RNA Worlds: From Life's Origins to Diversity in Gene Regulation.* Cold Spring Harbor Monograph Series. 4.2.1

Baars, B. J. (1988). *A Cognitive Theory of Consciousness.* 2.2, 5.3.1

Baars, B. J. (1993). How does a serial, integrated and very limited stream of consciousness emerge from a nervous system that is mostly unconscious, distributed, parallel and of enormous capacity? *Ciba Foundation symposium*, 174:282–90; discussion 291–303. 5.5

Barrell, B. G., Bankier, A. T., and Drouin, J. (1979). A different genetic code in human mitochondria. *Nature*, 282(5735):189–194. 4.2.2

Boettiger, C. A. and D'Esposito, M. (2005). Frontal networks for learning and executing arbitrary stimulus-response associations. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(10):2723–2732. 5.5

Broadbent, D. (1958). *Perception and communication.* New York: Pergamon. 5.4.1

Broughton, R., Billings, R., Cartwright, R., Doucette, D., Edmeads, J., Edwardh, M., Ervin, F., Orchard, B., Hill, R., and Turrell, G. (1994). Homicidal somnambulism: a case report. *Sleep*, 17(3):253–64. 5.2

Burrows, M. and Sutton, G. (2013). Interacting gears synchronize propulsive leg movements in a jumping insect. *Science (New York, N.Y.)*, 341(6151):1254–6. 2.4.4

Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., Solomon, E. a., Majaj, N. J., and DiCarlo, J. J. (2014). Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. *PLoS computational biology*, 10(12):e1003963. 6.2

Cardamone, L., Loiacono, D., and Lanzi, P. L. (2009). Evolving competitive car controllers for racing games with neuroevolution. *Proceedings of the 11th Annual conference on Genetic and evolutionary computation GECCO 09*, pages 1179–1186. 6.2, 6.3

Cartwright, R. (2004). Sleepwalking violence: A sleep disorder, a legal dilemma, and a psychological challenge. 5.2

Chalmers, D. J. (2010). The Singularity : A Philosophical Analysis. *Journal of Consciousness Studies*, (17):7–56. 7.5

Chatterjee, S., Laudato, M., and Lynch, L. a. (1996). Genetic algorithms and their statistical applications: an introduction. *Computational Statistics & Data Analysis*, 22(6):633–651. 6.3

Cheng, L. K. L. and Unrau, P. J. (2010). Closing the circle: replicating RNA with RNA. *Cold Spring Harbor perspectives in biology*, 2(10):a002204. 4.2.1

Choquet, D. and Triller, A. (2013). The dynamic synapse. *Neuron*, 80(3):691–703. 1

Clark, D. P. (2012). *Molecular Biology, Second Edition.* 4

Clocksin, W. and Mellish, C. S. (2013). *Programming in Prolog: Using the ISO Standard*. Springer, fifth edit edition. 6.1, 6.1

Codling, E. A., Plank, M. J., and Benhamou, S. (2008). Random walk models in biology. *Journal of the Royal Society Interface*, 5:813–834. 2.3

Cohen, J. D., McClure, S. M., and Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 362(1481):933–42. 7.2

Cohen, J. R., Asarnow, R. F., Sabb, F. W., Bilder, R. M., Bookheimer, S. Y., Knowlton, B. J., and Poldrack, R. A. (2011). Decoding continuous variables from neuroimaging data: basic and clinical applications. *Frontiers in neuroscience*, 5(75). 5.4

Copley, S. D., Smith, E., and Morowitz, H. J. (2007). The origin of the RNA world: Coevolution of genes and metabolism. *Bioorganic Chemistry*, 35(6):430–443. 4

Corbetta, M. and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews. Neuroscience*, 3(3):201–215. 5.4.2

Crick, F. (1970). Central Dogma of Molecular Biology. *Nature*, 227(5258):561–563. 4

Cross, N., Naughton, J., and Walker, D. (1981). Design method and scientific method. *Design Studies*, 2(4):195–201. 5.1

Cully, A., Clune, J., and Mouret, J. (2014). Robots that can adapt like natural animals. *arXiv preprint arXiv:1407.3501*. 7.2

Darwin, C. (1859). *On the Origin of the Species*, volume 5. 2.4.4

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095):876–9. 7.2

de Duve, C. (2005). *Singularities: Landmarks on the Pathways of Life*. 4.2.1, 7.5

Dennett, D. (1991). *Consciousness Explained*. Boston: Little Brown. 2.2

Derbel, H., Jarboui, B., Hanafi, S., and Chabchoub, H. (2012). Genetic algorithm with iterated local search for solving a location-routing problem. *Expert Systems with Applications*, 39(3):2865–2871. 6.3

DeRosier, D. J. (1998). The turn of the screw: the bacterial flagellar motor. *Cell*, 93(1):17–20. 2.4.4

Dijksterhuis, A. and Aarts, H. (2010). Goals, Attention, and (Un)Consciousness. *Annual Review of Psychology*, 61:467–90. 5.3.3

Dunitz, J. D. and Joyce, G. F. (2013). Leslie E. Orgel. Technical report, National Academy of Sciences. 2.4.2

Eigen, M. and Schuster, P. (1977). The Hypercycle: A Principle of Natural Self-Organization. *Die Naturwissenschaften*, 64:541–565. 4.2.1

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, 6(3):169–200. 5.5

Ferrucci, D. (2012). Introduction to "This is Watson". *IBM Journal of Research and Development*, 56(3.4):1:1–1:15. 2.1.1

Finney, E. M., Clementz, B. A., Hickok, G., and Dobkins, K. R. (2003). Visual stimuli activate auditory cortex in deaf subjects: evidence from MEG. Technical Report 11, Department of Psychology, University of California-San Diego, La Jolla, CA 92093, USA. 1

Fodor, J. A. (1983). *The modularity of mind*. MIT Press, Cambridge, MA. 5.3.1

Fogel, D. (2006). *Evolutionary computation: toward a new philosophy of machine intelligence*. 2.4.3

Forterre, P. (2005). The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. In *Biochimie*, volume 87, pages 793–803. 4.2.2

Franklin, S., Madl, T., D'Mello, S., and Snaider, J. (2014). LIDA: A Systems-level Architecture for Cognition, Emotion, and Learning. *Autonomous Mental Development*, 6(1):19–41. 2.1.1

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202. 6.2

Galinier, P. and Hao, J.-K. (1999). Hybrid evolutionary algorithms for graph coloring. *Journal of combinatorial optimization*, 3:379–397. 6.3

Galván-López, E., Swafford, J. M., O'Neill, M., and Brabazon, A. (2010). Evolving a Ms. PacMan controller using grammatical evolution. In *Applications of Evolutionary Computation*, pages 161–170. Springer Berlin Heidelberg. 6.3

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* 2.2

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2009). *Cognitive Neuroscience: The Biology of the Mind.* 3rd editio edition. 5.3.3

Gesteland, R. F. (2005). *The RNA World.* Cold Spring Harbor Monograph Series, third edit edition. 4.2.1

Gibbons, R. A. and Hunter, G. D. (1967). Nature of the Scrapie Agent. *Nature*, 215(5105):1041–1043. 4.1.3.1

Gilbert, W. (1986). Origin of life: The RNA world. 4, 4.2.1

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38(1):173–198. 8

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning.* Addison-Wesley Longman Publishing Co., Inc. 6.3

Gomez-Lazaro, M., Fernandez-Gomez, F. J., and Jordán, J. (2004). p53: Twenty five years understanding the mechanism of genome protection. *Journal of Physiology and Biochemistry*, 60(4):287–307. 7.4

Gondek, D. C., Lally, A., Kalyanpur, A., Murdock, J. W., Duboue, P. A., Zhang, L., Pan, Y., Qiu, Z. M., and Welty, C. (2012). A framework for merging and ranking of answers in DeepQA. *IBM Journal of Research and Development*, 56(3):399–410. 2.1.1

Graves, A., Liwicki, M., Fernández, S., Bertolami, R., Bunke, H., and Schmidhuber, J. (2009). A novel connectionist system for unconstrained handwriting recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):855–868. 6.2

Greaves, M. and Maley, C. C. (2012). Clonal evolution in cancer. *Nature*, 481(7381):306–13. 2.4.1

Groeger, J. A. and Clegg, B. A. (1997). Automaticity and driving: time to change gear? *Traffic and transport psychology. Theory and application.* 5.2

Gubrud, M. A. (1997). Nanotechnology and International Security. In *Fifth Foresight Conference on Molecular Nanotechnology.* 2.1

Haynes, J.-d. (2015). An Information-Based Approach to Consciousness: Mental State Decoding. *Open MIND*, 17:19. 5.5

Heinemann, J. A. and Bungard, R. A. (2006). Horizontal Gene Transfer. In *Encyclopedia of Molecular Cell Biology and Molecular Medicine.* 7.4

Heisenberg, M. (2009). Is free will an illusion? *Nature*, 459(7244):164–5. 2.3

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554. 6.2

Hofstadter, D. (1980). *Godel, Escher, Bach.* 8

Hoque, R. and Chesson, A. L. (2009). Zolpidem-induced sleepwalking, sleep related eating disorder, and sleep-driving: fluorine-18-flourodeoxyglucose positron emission tomography analysis, and a literature review of other unexpected clinical effects of zolpidem. *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine*, 5(5):471–6. 5.2

Hume, D. (1779). *Dialogues Concerning Natural Religion.* 2.4

Ikeuchi, K. (2014). Cognitive System. In *Computer Vision: A reference guide*, pages 100–106. 1.2

Jackson, M. B., Morrow, I. B., and Osborne, D. J. (1972). Abscission and dehiscence in the squirting cucumber, Ecballium elaterium. Regulation by ethylene. *Canadian Journal of Botany*, 50(7):1465–1471. 2.4.4

Jain, S. and Gea, H. C. (1996). PCB Layout design using a genetic algorithm. *Journal of Electronic Packaging*, 118:11–15. 6.3

James, W. (1897). The dilemma of determinism. In *The Will to Believe and other Essays in Popular Philosophy.* 2.3

Jessell, T. M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature reviews. Genetics*, 1(1):20–29. 7.4

Jiménez, L. (2008). Taking patterns for chunks: Is there any evidence of chunk learning in continuous serial reaction-time tasks? *Psychological Research*, 72(4):387–396. 5.3.2

Johnson, W. L. (1992). Needed: a new test of intelligence. *ACM SIGART Bulletin*, 3(4):7–9. 5.6

Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E., and Bartel, D. P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science (New York, N.Y.)*, 292(5520):1319–25. 4.2.1

Joyce, G. (1989). RNA evolution and the origins of life. *Nature*, 338:217–224. 4.2.1

Joyce, G. F. (2009). Evolution in an RNA world. *Cold Spring Harbor symposia on quantitative biology*, 74(1993):17–23. 4.2.1

Kenemans, L. and Ramsey, N. (2013). *Psychology in the brain: Integrative cognitive neuroscience*. Palgrave Macmillan. 5.4.2

Kim, Y. H. and Moon, B. R. (2004). Investigation of the fitness landscapes in graph bipartitioning: An empirical study. *Journal of Heuristics*, 10:111–133. 6.3

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, 4(4):138–147. 3

Koechlin, E., Basso, G., Pietrini, P., Panzer, S., and Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature*, 399(6732):148–151. 5.5

Koechlin, E. and Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science (New York, N.Y.)*, 318(2007):594–598. 5.5

Kotsiantis, S. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 31:249–268. 6.2

Kubota, J. T., Banaji, M. R., and Phelps, E. a. (2012). The neuroscience of race. *Nature neuroscience*, 15(7):940–8. 5.4.2

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. 8.1.1

Lally, A. and Fodor, P. (2011). Natural Language Processing With Prolog in the IBM Watson System. 6.1

Langley, P. (2006). Intelligent Behavior in Humans and Machines. *Technical Report. Computational Learning Laboratory, CSLI, Stanford University*. 2.1.1

Langley, P., Laird, J. E., and Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10:141–160. 2.1.1

Larman, C. and Basili, V. R. (2003). Iterative and incremental development: A brief history. *Computer*, 36(6):47–56. 5.1

Lassner, M. and Bedbrook, J. (2001). Directed molecular evolution in plant improvement. 4.1.2

Legg, S. and Hutter, M. (2007). Universal Intelligence: A Definition of Machine Intelligence. 1.2, 5.6

Leijnen, S. (2014). *Creativity & Constraint in Artificial Systems*. PhD thesis. 3.2.1

LeShan, L. and Margenau, H. (1983). *Einstein's space and Van Gogh's sky*. Harvester. 2.3

Lesnik, E. A. and Freier, S. M. (1995). Relative Thermodynamic Stability of DNA, RNA, and DNA:RNA Hybrid Duplexes: Relationship with Base Composition and Structure. *Biochemistry*, 34(34):10807–10815. 4.2.2

Lincoln, T. A. and Joyce, G. F. (2009). Self-sustained Replication of an RNA Enzyme. *Science*, 323(5918):1229–1232. 4.2.1

Lloyd, J. W. (1994). Practical advantages of declarative programming. In *Joint Conference on Declarative Programming, Invited Lecture, GULP-PRODE*, volume 94. 6.1

Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In *Inhibitory processes in attention, memory, and language*, pages 189–239. Academic Press. 5.4.2

MacDonald, A. W., Cohen, J. D., Stenger, V. A., and Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science (New York, N.Y.)*, 288(5472):1835–1838. 5.4.2

Mahadevan, S. and Connell, J. (1992). Automatic programming of behavior-based robots using reinforcement learning. *Artificial Intelligence*, 55(2-3):311–365. 7.2

Martick, M. and Scott, W. G. (2006). Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell*, 126(2):309–20. 4.2

Martin, A. (2007). The representation of object concepts in the brain. *Annual review of psychology*, 58:25–45. 5.3.1

Mayer, A. R., Dorflinger, J. M., Rao, S. M., and Seidenberg, M. (2004). Neural networks underlying endogenous and exogenous visual-spatial orienting. *NeuroImage*, 23(2):534–541. 3

Maynard Smith, J. (1970). Natural Selection and the Concept of a Protein Space. *Nature*, 225(5232):563–564. 4.1.1

Mayr, E. (1988). *Toward a New Philosophy of Biology*. Belknap Press of Harvard University Press. 2.3

Meggers, E., Holland, P. L., Tolman, W. B., Romesberg, F. E., Schultz, P. G., and Jolla, L. (2000). A Novel Copper-Mediated DNA Base Pair. *Journal of the American Chemical Society*, 122:10714–5. 7.4

Miller, E. K., Nieder, A., Freedman, D. J., and Wallis, J. D. (2003). Neural correlates of categories and concepts. *Current Opinion in Neurobiology*, 13(2):198–203. 5.3.1, 5.5

Miller, G. A. (1956). The Magical Number Seven. *Psychological Review*, 63:81–97. 5.3.2

Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science (New York, N.Y.)*, 117(3046):528–9. 7.5

Mutani, R., Chiarot, E., and Cicolin, A. (2009). Parasomnias: sleep disturbances with automatic behaviours. *Confinia Psychiatrica*, 4(1-2):1–6. 5.2, 5.6

Nakamura, J. and Ohsawa, Y. (2009). A shift of mind - Introducing a concept creation model. *Information Sciences*, 179(11):1639–1646. 7.2

Nelissen, R. M. a., Dijker, a. J. M., and de Vries, N. K. (2007). Emotions and goals: Assessing relations between values and emotions. *Cognition & Emotion*, 21(4):902–911. 5.5

Newell, A. (1992). Unified theories of cognition and the role of Soar. In *Soar A cognitive architecture in perspective A tribute to Allen Newell Studies in cognitive systems Vol 10*, pages 25–79 ST – Unified theories of cognition and the. 2.1.1

Nguyen, A., Yosinski, J., and Clune, J. (2014). Deep Neural Networks are Easily Fooled: High Confidence Predictions for Unrecognizable Images. 6.2

Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H. G., and Ogata, T. (2014). Audio-visual speech recognition using deep learning. *Applied Intelligence*. 6.2

Orgel, L. E. (1973). *The origins of life: molecules and natural selection*. 2.4.2, 2.4.2, 2.4.3

Orgel, L. E. (1986). RNA Catalysis and the Origins of Life. *Journal of theoretical biology*, 123:127–149. 4.2.1

Paul, N. and Joyce, G. F. (2002). A self-replicating ligase ribozyme. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12733–40. 4.2.1

Petersen, S. and Posner, M. (2012). The Attention System of the Human Brain: 20 Years After. *Annual review of neuroscience*, 21(35):73–89. 5.3.3

Poincaré, H. (1908). L'invention mathématique. In *Science et méthode*. 5.1

Popper, K. (1978). Natural selection and the emergence of mind. *Dialectica*, 32(3-4):339–355. 5.1

Pressman, M. R. (2011). Sleep driving: sleepwalking variant or misuse of z-drugs? *Sleep medicine reviews*, 15(5):285–92. 5.2

Prusiner, S. B. (1998). Prions. *Proceedings of the National Academy of Sciences*, 95(23):13363–13383. 4.1.3.1

Pullman, A. and Pullman, B. (1963). On the mechanism of ultraviolet-induced mutations. *Biochimica et Biophysica Acta*, 75:269–271. 4.1.1, 4.2.1, 7.4

Purves, D., Cabeza, R., Huettel, S. A., LaBar, K. S., Platt, M. L., and Woldorff, M. G. (2013). *Principles of Cognitive Neuroscience. 2nd Edition*. Sinauer Associates, Inc. 1.2

Raghuveer, M., Naryana, L. L., and Dutt, B. S. M. (1993). Dehiscent fruit of Hydrocera triflora (Linn.) Wt. and Arn. (Balsaminaceae): its anatomy and dispersal. *Rheedea*, 3:12–14. 2.4.4

Robertson, M. P. and Joyce, G. F. (2012). The Origins of the RNA World. *Cold Spring Harbor Perspectives in Biology*, 4(5). 4.2.1

Röder, B., Stock, O., Bien, S., Neville, H., and Rösler, F. (2002). Speech processing activates visual cortex in congenitally blind humans. *European Journal of Neuroscience*, 16(5):930–936. 1

Rolls, E. T. and Treves, A. (2011). The neuronal encoding of information in the brain. *Progress in neurobiology*, 95(3):448–90. 5.3.1

Roser, M. and Gazzaniga, M. (2006). The interpreter in human psychology: The evolution of primate nervous systems. *Oxford: Elsevier*. 5.6

Sammut, C. (2012). When Do Robots Have to Think? *Advances in Cognitive Systems*, 1:73–81. 2.1.1

Sandberg, A. (2013). An Overview of Models of Technological Singularity. In *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*, pages 376–394. 2.1

Shaffer, M. L. (1981). Minimum Population Sizes for Species Conservation. *BioScience*, 31(2):131–134. 7.3

Shi, Z. (2011). *Advanced Artificial Intelligence*. World ScientificPublishing Co. Pte. Ltd. 1

Siegelmann, H. T. and Sontag, E. D. (1991). Turing computability with neural nets. *Applied Mathematics Letters*, 4(6):77–80. 6.2

Sivanandam, S. and Deepa, S. N. (2010). *Introduction to Genetic Algorithms*. Springer. 6.3

Sniegowski, P. D., Gerrish, P. J., Johnson, T., and Shaver, A. (2000). The evolution of mutation rates: separating causes from consequences. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 22(12):1057–66. 4.1.1, 7.2

Spearman, C. (1904). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, 15(2):201–292. 2.2

Stanley, K. O. and Miikkulainen, R. (2002). Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127. 6.3

Sternberg, R. (1985). *Beyond IQ: A triarchic theory of human intelligence*. 2.2

Suresh, L. P., Dash, S. S., and Panigrahi, B. K. (2015). *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*, volume 324 of *Advances in Intelligent Systems and Computing*. Springer India, New Delhi. 6.3

Surewicz, W. K. and Apostol, M. I. (2011). Prion protein and its conformational conversion: a structural perspective. *Topics in current chemistry*, 305:135–67. 4.1.3.1

Sutton, R. and Barto, A. (1998). *Introduction to reinforcement learning*. MIT Press. 7.2

Szathmáry, E. (2006). The origin of replicators and reproducers. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 361(1474):1761–76. 4.2.1

Tarabichi, M., Antoniou, a., Saiselet, M., Pita, J. M., Andry, G., Dumont, J. E., Detours, V., and Maenhaut, C. (2013). Systems biology of cancer: entropy, disorder, and selection-driven evolution to independence, invasion and "swarm intelligence". *Cancer metastasis reviews*, 32(3-4):403–21. 2.4.1

Tenaillon, O., Taddei, F., Radman, M., and Matic, I. (2001). Second-order selection in bacterial evolution: selection acting on mutation and recombination rates in the course of adaptation. *Research in Microbiology*, 152(1):11–16. 4.2.2, 7.2

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. *Journal of experimental psychology. Human perception and performance*, 20(4):799–806. 5.4.3

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135:77–99. 1

Theeuwes, J., Reimann, B., and Mortier, K. (2006). Visual search for featural singletons: No top-down modulation, only bottom-up priming. *Visual Cognition*, 14(4-8):466–489. 2

Theodoridis, S. and Koutroumbas, K. (2008). *Pattern Recognition, Fourth Edition*. 6.2

Thierens, D. (1999). Scalability problems of simple genetic algorithms. *Evolutionary computation*, 7(4):331–52. 6.3

Thierens, D. (2010). The Linkage Tree Genetic Algorithm. In *Parallel Problem Solving from Nature, PPSN XI*, pages 264–273. Springer Berlin Heidelberg, Berlin, Heidelberg. 6.3

Thurstone, L. (1938). *Primary Mental Abilities*. University of Chicago Press: Chicago. 2.2

Tononi, G. (2004). An information integration theory of consciousness. *BMC neuroscience*, 5:42. 2.2

Top500.org (2014). List of top 500 supercomputers. *http://www.top500.org/lists/2014/11/*. 6.3

Trakhtenbrot, A., Nathan, R., Perry, G., and Richardson, D. M. (2005). The importance of long-distance dispersal in biodiversity conservation. *Diversity and Distributions*, 11:173–181. 7.4

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, 3(6):449–459. 5.4.1

Treisman, A. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12:97–136. 5.4.1

Turing, A. (1950). Computing machinery and intelligence. *Mind*. 1, 2.1

Turner, N. J. (2009). Directed evolution drives the next generation of biocatalysts. *Nature chemical biology*, 5(8):567–73. 4.1.2, 7.4

Valencia, M., Pastor, M. a., Fernández-Seara, M. a., Artieda, J., Martinerie, J., and Chavez, M. (2009). Complex modular structure of large-scale brain networks. *Chaos*, 19(2):1–7. 5.3.1

van de Kamp, T., Vagovič, P., Baumbach, T., and Riedel, A. (2011). A biological screw in a beetle's leg. *Science (New York, N.Y.)*, 333(6038):52. 2.4.4

Verwey, W. B. and Eikelboom, T. (2003). Evidence for lasting sequence segmentation in the discrete sequence-production task. *Journal of motor behavior*, 35(2):171–181. 5.3.2

Vowk, B., Wait, A., and Schmidt, C. (2004). An evolutionary approach generates human competitive coreware programs. *Workshop and Tutorial Proceedings Ninth International Conference on the Simulation and Synthesis of Living Systems (Alife XI)*, pages 33–36. 6.3

Wentura, D., Müller, P., and Rothermund, K. (2014). Attentional capture by evaluative stimuli: Gain- and loss-connoting colors boost the additional-singleton effect. *Psychonomic Bulletin & Review2*, 21(3):701–707. 2

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, 95(12):6578–6583. 6.3

Yarus, M. (2011). Getting past the RNA world: the initial Darwinian ancestor. *Cold Spring Harbor perspectives in biology*, 3(4). 4.2.1

Zahn, R., Moll, J., Krueger, F., Huey, E. D., Garrido, G., and Grafman, J. (2007). Social concepts are represented in the superior anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 104(15):6430–6435. 5.3.1